

Outil d'aide à la décision « Trouver un entrepôt de données » : version statique

Ce document propose un récapitulatif des éléments de contenu présentés de façon dynamique dans l'outil d'aide à la décision en ligne « Trouver un entrepôt de données », disponible en ligne sur le site des bibliothèques de l'université de Bordeaux à l'adresse suivante : <http://busec2.u-bordeaux.fr/aide-choix-entrepot>

Table des matières

Textes d'accompagnement des questions	2
Question 1 accès : « Allez-vous publier des données dont l'accès doit être restreint ? »	2
Question 2 coûts : « Recherchez-vous un entrepôt avec des options de dépôt gratuit ? »	2
Question 3 licences : « Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données ? »	2
Question 4 relecture : « Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne ? »	3
Fiches individuelles de présentation des entrepôts de données	3
Fiche entrepôt : 4TU.ResearchData	3
Fiche entrepôt : Zenodo	4
Fiche entrepôt : Dryad	5
Fiche entrepôt : Open Science Framework (OSF)	5
Fiche entrepôt : Mendeley Data	6
Fiche entrepôt : Figshare	7
Fiche entrepôt : B2SHARE	8
Fiche entrepôt : Harvard Dataverse	9
Tableau récapitulatif des réponses aux questions pour chaque entrepôt de données	10

Textes d'accompagnement des questions

Question 1 accès : « Allez-vous publier des données dont l'accès doit être restreint ? »

Avec les restrictions d'accès, vous pouvez rendre publiques uniquement des informations (les métadonnées) concernant les données que vous déposez. L'accès aux fichiers de données est conditionné à votre autorisation. L'entrepôt de données gère tous les aspects en lien avec ces restrictions (échanges avec les demandeurs, communication des fichiers pour les demandes acceptées, etc.).

Les restrictions d'accès diffèrent de **l'embargo**. L'embargo est une **période limitée prédéfinie** pendant laquelle les fichiers de données ne sont pas accessibles. Les métadonnées restent visibles immédiatement. Certains entrepôts, Figshare par exemple, ne gèrent pas de restrictions d'accès mais permettent de paramétrer un embargo. Grâce à ce paramétrage, les fichiers de données deviennent accessibles automatiquement, à la fin de la période d'embargo définie.

Question 2 coûts : « Recherchez-vous un entrepôt avec des options de dépôt gratuit ? »

Le fait que le dépôt dans un entrepôt soit gratuit ou non dépend souvent de la taille du jeu de données que vous soumettez. Veillez à consulter les informations disponibles pour chaque entrepôt de données, afin de vérifier si la taille de votre jeu de données entraînerait des coûts. Le dépôt dans 4TU.ResearchData est ainsi gratuit pour des jeux de données jusqu'à 10 Go par an et par utilisateur, alors que Dryad applique des frais de publication (DPC - Data Publishing Charges) de 120 \$ pour toute soumission. Notez que vous pouvez être dispensé de ces frais, par exemple si vos données sont liées à un article publié dans une revue partenaire de Dryad. Figshare propose de son côté des services supplémentaires, notamment en termes de volumétrie, via 2 types d'offre payante : l'abonnement institutionnel et Figshare+ pour une publication ou un projet spécifique.

Question 3 licences : « Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données ? »

Les entrepôts varient en fonction de la gamme de licences d'utilisation des données que vous pouvez choisir. Ces licences informent tout utilisateur des types de réutilisation autorisés. Dans certains entrepôts, vous pouvez publier vos données uniquement sous la licence [Creative Commons Zero - CCo](#), (domaine public, aucun droit réservé). D'autres entrepôts proposent une sélection fermée de licences parmi lesquelles vous pouvez choisir. D'autres entrepôts enfin vous laissent définir librement les conditions sous lesquelles vous souhaitez mettre à disposition vos données.

Question 4 relecture : « Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne ? »

Dans certains entrepôts de données, la mise en ligne des données est immédiate lorsque le contributeur clique sur le bouton "Publier" ou son équivalent. Dans d'autres entrepôts, les dépôts sont mis en ligne après un processus de relecture plus ou moins approfondi. Dans Mendeley, les dépôts sont modérés pour s'assurer qu'ils répondent aux critères demandés, sans qu'une validation du contenu soit effectuée. 4TU.ResearchData et Dryad effectuent une relecture plus poussée, incluant les vérifications suivantes : complétude des métadonnées, documentation, liens, formats de fichier, respect de la propriété intellectuelle etc. - voir les pages "[After uploading your dataset](#)" de [4TU.ResearchData](#) et "[Submission process](#)" de [Dryad](#). Figshare+, l'offre payante de Figshare, comporte une relecture et une vérification des métadonnées et de la description des données, ainsi que des vérifications ponctuelles sur un échantillon des fichiers de données.

Fiches individuelles de présentation des entrepôts de données

Fiche entrepôt : 4TU.ResearchData



4TU.ResearchData, créé à l'initiative de trois universités techniques néerlandaises, est ouvert à tous les chercheurs, avec un accent sur les sciences exactes, l'ingénierie et le design - voir [Data collection policy](#) pour une liste des domaines couverts.

4TU.ResearchData fournit depuis 2010 une infrastructure et des services (assistance pour l'organisation et la documentation des données, modération des dépôts, outils pour la réutilisation des données) pour le partage et la conservation à long terme des données de recherche.

4TU.ResearchData est spécialisé dans le traitement du format NetCDF avec le protocole OPeNDAP, permettant l'utilisation de données provenant d'un serveur distant sans avoir besoin de télécharger les fichiers de données, voir [NetCDF and OPeNDAP](#).

URL : <https://data.4tu.nl>

Identifiant pérenne : DOI

Restrictions d'accès aux données : oui, les jeux de données peuvent être en accès ouvert, sous embargo ou restreint

Flexibilité pour les licences d'utilisation : choix parmi des licences prédéfinies : toutes les licences Creative Commons ; licences spécifiques pour le code informatique et les logiciels : MIT, BSD 3-Clause, Apache-2.0, EUPL-1.2, GPL, LGPL et AGPL

Fichiers de données volumineux : jusqu'à 10 Go

Coûts : 10 Go gratuits de données par an et par utilisateur pour les utilisateurs extérieurs ; au-delà de ce seuil, des frais de 4,5€ par Go s'appliquent

Certification CoreTrustSeal : oui

Stockage localisé en Union Européenne : oui, aux Pays-Bas

Préservation : minimum 15 ans

Propriétaire : consortium associant les universités technologiques de Delft et Eindhoven et l'université de Twente

Fiche entrepôt : Zenodo



Zenodo, financé par l'Union Européenne et hébergé par le CERN, est ouvert à tous les contributeurs, accepte tous les types de productions scientifiques (publications, données, images, fichiers audio et vidéo, livrables de projet, etc.), de toutes les disciplines.

Zenodo offre une interface de dépôt très simple et propose des services tels que : création et organisation de collections (communautés), synchronisation avec GitHub, intégration dans le système de *reporting* pour les projets européens.

URL : <https://zenodo.org>

Identifiant pérenne : DOI - Zenodo enregistre au moins deux DOI pour chaque jeu de données : un DOI représentant chaque version spécifique et un DOI représentant toutes les versions du jeu de données

Restrictions d'accès aux données : oui, les jeux de données peuvent être en accès ouvert, sous embargo, restreint ou fermé

Flexibilité pour les licences d'utilisation : choix parmi de très nombreuses licences prédéfinies pour les dépôts en accès ouvert ou sous embargo ; possibilité de définir sa propre licence en téléchargeant un fichier "license.txt"

Fichiers de données volumineux : jusqu'à 50 Go par jeu de données ; des volumes plus importants peuvent être négociés avec Zenodo

Coûts : le dépôt est gratuit, toutefois Zenodo accepte les dons

Certification CoreTrustSeal : non

Stockage localisé en Union Européenne : oui, en Suisse

Préservation : minimum 20 ans

Propriétaire : CERN

Fiche entrepôt : Dryad



Dryad est géré par une organisation à but non lucratif, créée à l'initiative de sociétés savantes et d'éditeurs pour diffuser les données associées à des articles. Certaines revues partenaires de Dryad ont intégré la soumission des données dans Dryad à leur processus de soumission des articles.

Les données non associées à des publications sont également acceptées dans Dryad. Tous les jeux de données soumis sont modérés et validés avant d'être mis en ligne et publiés en accès ouvert sous licence CCo.

URL : <https://datadryad.org>

Identifiant pérenne : DOI

Restrictions d'accès aux données : non, seul un accès ouvert est possible. Un accès sous embargo peut être sollicité auprès du support Dryad lors de la soumission des données. Un jeu de données peut rester privé pendant le processus de publication de l'article auquel il est lié : une URL privée de téléchargement des fichiers est alors fournie.

Flexibilité pour les licences d'utilisation : seule la licence Creative Commons Zero CCo est disponible

Fichiers de données volumineux : jusqu'à 300 Go par jeu de données ; des volumes plus importants peuvent être négociés avec Dryad

Coûts : 120 \$, auxquels s'ajoutent des frais de 50 \$ par tranche de 10 Go pour les jeux de données excédant 50 Go. Ces frais de publication ne sont pas facturés aux contributeurs provenant de certains pays ou d'institutions membres de Dryad, ni aux auteurs d'articles publiés dans des revues partenaires.

Certification CoreTrustSeal : indirectement : Dryad est archivé dans l'entrepôt Merritt de l'U3C (University of California Curation Center), certifié CoreTrustSeal

Stockage localisé en Union Européenne : non, aux Etats-Unis

Préservation : indéfinie

Propriétaire : Dryad

Fiche entrepôt : Open Science Framework (OSF)



Open Science Framework (OSF) est une plateforme créée par le Center for Open Science (COS), une organisation à but non lucratif fondée à l'université de Virginie aux Etats-Unis. OSF permet à la fois le partage des données de recherche et le travail collaboratif, grâce à des fonctionnalités telles que la gestion des droits et des contributions, le versionnement,

l'organisation en projet et en composants imbriqués, ou la connexion avec des services tiers (Google, GitHub, Zotero, etc.).

OSF héberge également des services de prépublications (PsyArXiv et SocArXiv par exemple) et d'enregistrement d'études.

URL : <https://osf.io>

Identifiant pérenne : DOI

Restrictions d'accès aux données : non, seul un accès ouvert est possible. Un projet OSF peut rester privé à ses membres : une URL privée est alors fournie et peut être transmise aux personnes extérieures au projet qui doivent accéder aux données.

Flexibilité pour les licences d'utilisation : choix parmi des licences prédéfinies ; possibilité de définir sa propre licence en téléchargeant un fichier "license.txt"

Fichiers de données volumineux : jusqu'à 5 Go pour les projets et composants privés, jusqu'à 50 Go pour les projets et composants publics ; des services de stockage tiers (Dropbox, GitLab, etc.) peuvent être connectés à un projet - voir l'article [OSF Storage Caps](#) dans la FAQ OSF

Coûts : gratuit

Certification CoreTrustSeal : non

Stockage localisé en Union Européenne : oui, en Allemagne

Préservation : le financement est assuré pour 50 ans

Propriétaire : Center for Open Science

Fiche entrepôt : Mendeley Data



Mendeley Data, fourni par l'éditeur commercial Elsevier, est ouvert à tous les contributeurs et accepte tous les types de données, de toutes les disciplines. Le processus de dépôt est simple et rapide.

Mendeley Data bénéficie des services de l'infrastructure nationale néerlandaise d'archivage Data Archiving and Network Services (DANS).

URL : <https://data.mendeley.com>

Identifiant pérenne : DOI

Restrictions d'accès aux données : non, seul un accès ouvert est possible. Un embargo d'un an maximum peut être paramétré. Tant que le dépôt n'est pas publié et reste privé, une URL privée de téléchargement des fichiers peut être générée.

Flexibilité pour les licences d'utilisation : choix parmi des licences prédéfinies : licences Creative Commons et licences spécifiques à l'informatique (*code, software et hardware*)

Fichiers de données volumineux : jusqu'à 10 Go par jeu de données ; cette limite peut être portée à 100 Go pour les membres d'une institution ayant souscrit un abonnement au service

Coûts : gratuit

Certification CoreTrustSeal : oui

Stockage localisé en Union Européenne : oui, en Irlande

Préservation : archivage à long terme assuré par DANS

Propriétaire : Elsevier

Fiche entrepôt : Figshare



Figshare, fourni par l'éditeur commercial Digital Science, est ouvert à tous les contributeurs, accepte tous les types de productions scientifiques (publications, données, images, fichiers audio et vidéo, ressources en ligne, etc.), de toutes les disciplines.

Figshare offre des fonctionnalités complémentaires de travail collaboratif et d'éditorialisation des données telles que : collections, référencement de données déposées dans d'autres entrepôts, référencement de données sans fichiers associés.

L'offre payante Figshare+ permet d'accéder à des fonctionnalités complémentaires, notamment en termes de volumétrie de données déposées, de choix de licences d'utilisation et de support de la part des experts de Figshare.

URL : <https://figshare.com>

Identifiant pérenne : DOI

Restrictions d'accès aux données : oui, les jeux de données peuvent être en accès ouvert, sous embargo ou fermé (dépôt confidentiel). L'embargo peut s'appliquer seulement aux fichiers ou au dépôt lui-même ; dans ce dernier cas, les métadonnées ne sont alors pas visibles pendant la période d'embargo. Tant que le dépôt n'est pas publié et reste privé, une URL de partage privée peut être générée.

Flexibilité pour les licences d'utilisation : choix parmi des licences prédéfinies : licences Creative Commons CCo et CC-BY ; licences MIT, GNU et Apache pour le code informatique et les logiciels. **Une flexibilité complète est disponible via l'offre payante Figshare+.**

Fichiers de données volumineux : jusqu'à 5 Go par fichier ; espace de stockage privé jusqu'à 20 Go ; cette limite peut être augmentée pour les membres d'une institution ayant souscrit un abonnement au service ou dans le cadre de l'offre Figshare+.

Coûts : gratuit ; des services complémentaires (relecture des dépôts par exemple) sont disponibles via un abonnement institutionnel ou l'offre Figshare+.

Certification CoreTrustSeal : non

Stockage localisé en Union Européenne : non précisé, Figshare est hébergé par Amazon Web Services, mais la localisation des serveurs n'est pas spécifiée

Préservation : archivage à long terme assuré par Chronopolis (University of California at San Diego)

Propriétaire : Digital Science

Fiche entrepôt : B2SHARE



B2SHARE est ouvert à tous les contributeurs et accepte tous les types de données, de toutes les disciplines. B2SHARE est l'un des services proposés par l'infrastructure européenne EUDAT, financée par l'Union européenne ; il est connecté aux autres services de EUDAT : B2STORE, B2DROP, B2NOTE, etc.

B2SHARE propose des options étendues pour la description des données. Le formulaire standard de dépôt permet notamment de saisir plusieurs descriptions, de différents types (résumé, méthode, table des matières, etc.). Les communautés, créées à la demande, permettent d'aller plus loin en paramétrant un formulaire de dépôt spécifique, avec des métadonnées complémentaires personnalisées.

URL : <https://b2share.eudat.eu>

Identifiant pérenne : DOI et handle

Restrictions d'accès aux données : oui, les jeux de données peuvent être en accès ouvert, sous embargo ou restreint

Flexibilité pour les licences d'utilisation : choix parmi des licences prédéfinies grâce à l'outil d'aide à la décision intégré "Select License" ; saisie libre possible dans le champ "License" ; champ "URL" pour indiquer le lien vers la description de la licence que vous souhaitez appliquer

Fichiers de données volumineux : jusqu'à 10 Go par fichier et 20 Go par jeu de données (dépôt, *record*) ; des volumes plus importants peuvent être négociés avec EUDAT

Coûts : gratuit

Certification CoreTrustSeal : non, le centre de données Datacenter CSC Kajaani est certifié ISO/IEC 27001:2005 pour son système de gestion de la sécurité informatique

Stockage localisé en Union Européenne : oui, en Finlande (Finnish IT Center for Science, Datacenter CSC Kajaani)

Préservation : indéfinie ; repose sur le réseau des 25 membres de l'infrastructure EUDAT (organismes de recherche, centres de données et de calcul nationaux, fournisseurs de technologies)

Propriétaire : EUDAT (European Data Infrastructure)

Fiche entrepôt : Harvard Dataverse



Harvard Dataverse est ouvert à tous les contributeurs et accepte tous les types de données, de toutes les disciplines, avec un accent sur les sciences humaines et sociales.

Harvard Dataverse offre des fonctionnalités étendues de description des données et de personnalisation. L'utilisateur peut créer un jeu de données ou un Dataverse. Créer un Dataverse permet de paramétrer une interface de dépôt et de consultation, en optant par exemple pour un modèle de métadonnées spécialisé (Geospatial, Social Sciences and Humanities, Astronomy and Astrophysics, Life Sciences ou Journal Metadata).

Des organismes de recherche comme le Cirad - <https://dataverse.cirad.fr> - et l'INRAE - <https://data.inrae.fr> - ont choisi le logiciel Dataverse pour créer leur entrepôt de données.

URL : <https://dataverse.harvard.edu>

Identifiant pérenne : DOI

Restrictions d'accès aux données : oui, les jeux de données peuvent être en accès ouvert, sous embargo ou restreint. Les restrictions peuvent être paramétrées pour chaque fichier individuellement ; la fonctionnalité de demande d'accès par les utilisateurs peut être désactivée. Tant que le dépôt n'est pas publié et reste privé, une URL de partage privée peut être générée.

Flexibilité pour les licences d'utilisation : choix entre la licence Creative Commons Zero CCo ou une licence propre définie par le contributeur

Fichiers de données volumineux : jusqu'à 2,5 Go par fichier et jusqu'à 1 To par utilisateur

Coûts : gratuit

Certification CoreTrustSeal : non

Stockage localisé en Union Européenne : non, aux Etats-Unis

Préservation : indéfinie ; les données publiées en libre accès dans Harvard Dataverse sont répliquées au sein de Data-PASS (Data Preservation Alliance for the Social Sciences)

Propriétaire : Harvard University, Institute for Quantitative Social Sciences

Tableau récapitulatif des réponses aux questions pour chaque entrepôt de données

	Question 1 accès « Allez-vous publier des données dont l'accès doit être restreint ? »	Question 2 coûts « Recherchez-vous un entrepôt avec des options de dépôt gratuit ? »	Question 3 licences « Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données ? »	Question 4 relecture « Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne ? »
4TU.ResearchData	Oui	Oui	Pas nécessairement	Oui
Zenodo	Oui	Oui	Oui	Pas nécessairement
Dryad	Pas nécessairement	Pas nécessairement	Pas nécessairement	Oui
Open Science Framework (OSF)	Pas nécessairement	Oui	Oui	Pas nécessairement
Mendeley	Pas nécessairement	Oui	Pas nécessairement	Oui
Figshare	Oui	Oui	Oui	Oui
B2SHARE	Oui	Oui	Oui	Pas nécessairement
Harvard Dataverse	Oui	Oui	Oui	Pas nécessairement