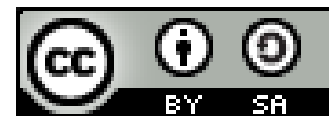


Panorama des entrepôts de données : comment choisir ?

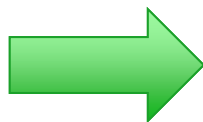
Luc DECKER
IRD (service IST – MCST)
data@ird.fr



Comment choisir parmi les nombreux entrepôts disponibles ?



Analyse des besoins
Bonnes pratiques



Caractéristiques requises

Reconnaitre les différents types d'entrepôts de données

- ☐ Généraliste
- ☐ Multi-disciplinaire
- ☐ Thématique / Spécialisé
- ☐ Dédié aux données scientifiques
- ☐ Institutionnel (exemple: mis en place par une université)
- ☐ Lié à une infrastructure de recherche
- ☐ Lié à un projet de recherche
- ☐ Indépendant / Associatif
- ☐ Géré par un éditeur scientifique
- ☐ Ouvert à tous
- ☐ Gratuit
- ☐ Commercial

Où déposer ?

Quel entrepôt de données choisir ?

- ☐ Parfois imposé (ou recommandé) par...
 - le financeur du projet de recherche
 - un journal où l'article associé aux données est publié
- ☐ Existe-t-il un **entrepôt thématique**, leader dans la discipline concernée ?

→ **Meilleur choix**

1^{er} objectif :
visibilité des données
FAIR : Facile à trouver

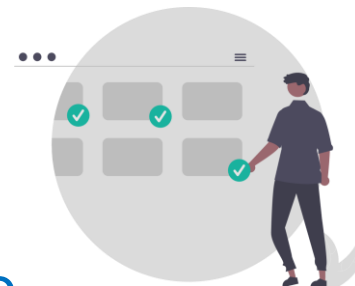
- ☐ Attention...



- ☐ « Faux entrepôts »
- ☐ Captation des données, cession des droits
- ☐ Services commerciaux : pérennité, conditions d'utilisation

→ **Critères de confiance**

Entrepôts : des **critères de confiance**, en cohésion avec les **principes FAIR**



- ✓ Pérennité du service, financement stable
- ✓ Identifiants pérennes (exemple : DOI)
- ✓ Formats ouverts
- ✓ Attribution de citations aux jeux de données
- ✓ Attribution de licences aux jeux de données
- ✓ Gestion des mises à jour, historique des versions
- ✓ Gestion des droits d'accès et accès restreints
- ✓ Service d'assistance ou/et de modération

Exemple de certification : **Core Trust Seal**

<https://www.coretrustseal.org>



Entrepôts : autres caractéristiques

- Durée de conservation minimum garantie
- Volume maximum de stockage autorisé
- Service associé d'archivage à long terme
- Métadonnées spécialisées dans la discipline
- Traitements de données en ligne, à la demande : conversions de format, calculs, tracé de graphiques, ...
- Services supplémentaires : anonymisation, ...



Où trouver un entrepôt de données ?

Consulter un annuaire d'entrepôts

[re3data](#) répertorie actuellement
2752 entrepôts



Exercice (en groupes)

Etudier les caractéristiques de quelques services de partage/préservation de données

Questions

- 1 ? Type de service (s'agit-il d'un entrepôt ?)
- 2 ? Conditions pour déposer / consulter / télécharger des données
- 3 ? Service support pour aider les déposants
- 4 ? Modération / vérification des dépôts, qualité
- 5 ? Coûts d'utilisation
- 6 ? Pérennité garantie
- 7 ? Identifiants pérennes
- 8 ? Citations
- 9 ? Licences
- 10 ? Certification
- 11 ? Recommandé (ou non) : pourquoi et pour quel usage
- ? Autres remarques

Exercice - Etude/comparaison des caractéristiques de quelques services de partage/préservation de données

Zenodo https://zenodo.org	Figshare https://figshare.com
Mendeley Data https://data.mendeley.com	Dryad https://datadryad.org
DataverseNL https://dataverse.nl	Vivli https://vivli.org
GenBank https://www.ncbi.nlm.nih.gov/genbank	ICPSR https://www.icpsr.umich.edu
PANGAEA https://www.pangaea.de	Dropbox https://www.dropbox.com
Archaeology Data Service https://archaeologydataservice.ac.uk	INDEPTH Data Repository https://www.indepth-ishare.org
ResearchGate https://www.researchgate.net	