

Focus sur : les Données de la Recherche : Questions-Réponses

Données de la recherche (Research data) : « enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. » (OCDE, 2007).

Foire aux questions - FAQ

LES DONNÉES DE LA RECHERCHE

Qu'est-ce que l'ouverture des données (Open data) ?

L'**ouverture des données (Open data)** donne la liberté d'utiliser, de modifier et de redistribuer librement les données, plus particulièrement celles financées sur fonds publics. Elle a pour objectif la diffusion libre, gratuite et universelle, via internet, des données d'origine publique ou privée.

Quelles données préserver ?

La préservation des données est une étape délicate. Leur sauvegarde sur le long terme génère un coût non négligeable d'où l'importance d'évaluer en amont celles qui pourront avoir un **intérêt scientifique ou historique** dans l'avenir. *The Natural Environment Research Council* (UK NERC) a établi une liste de **critères de sélection des données** autour de l'intérêt scientifique des données, des conditions de reproductibilité...
NERC Data Value Checklist : <http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist/>

Où déposer mes données ?

Dans un **entrepôt de données (Data repository) certifié*** dans lequel vos données seront accessibles :

- institutionnel : [University of Edinburgh](#) (Royaume-Uni); [Open Data LMU](#) (Allemagne); ...
- thématique, disciplinaire : [GenBank](#) ; [SeaDataNet](#) ; [PANGAEA](#) ; ...
- pluridisciplinaire : [Zenodo](#) (public) ; [OpenAIRE+](#) ; [Dryad](#) (privé); ...

* L'**entrepôt de confiance (trusted repository)** répond aux critères de qualité exigés pour obtenir une certification (format des données, qualité des métadonnées, conditions d'accès et de réutilisation, identifiant pérenne, archivage à long terme, ...).

Comment protéger mes données ?

Avant de rendre public un jeu de données, il faut lui attribuer une **licence de diffusion** fixant les conditions de son utilisation : droits d'utilisation et de modification, droits de réutilisation commerciale et non commerciale, obligations éventuelles comme la mention de la source des données ou le partage à l'identique. Les principales licences utilisées pour la publication des jeux de données sont les suivantes :

Les licences *Creative Commons* :

- [licence CC-BY 4.0](#) (*Creative Commons Attribution 4.0 International*)
- [licence CC0](#) (*Creative Commons Public Domain Dedication*)

Les licences de l'*Open Knowledge Foundation* (OKF)

- [licence ODC-BY](#) (*Open Data Commons Attribution License*)
- [licence ODC-ODbL](#) (*Open Data Commons Open Database License*)
- [licence PDDL](#) (*Public domain dedication and license*)

La licence ouverte (LO) : <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

Et si mes données sont déjà déposées dans les réseaux sociaux académiques (ResearchGate, Academia, ...) ?

- Redéposer vos données dans un entrepôt de confiance (institutionnel ou autre...)
- Dans un deuxième temps, supprimer ou demander la suppression de ce jeu de données du réseau social, et faire le lien du réseau social vers l'entrepôt.

Rappel :

Tout ce que vous déposez sur *ResearchGate* ou *Academia* devient la propriété du réseau

LES JEUX DE DONNÉES

Qu'est-ce qu'un jeu de données ?

Un **jeu de données scientifiques (data set)** peut être défini comme « l'agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine « unité », rassemblées pour former un ensemble cohérent. » (Gaillard, 2014). **Quelques exemples de jeux de données :**

- cumuls de lame d'eau ; profils de vents d'altitude... ([Portail de données publiques de Météo-France](#))
- résultats des **élections législatives 1958-2012** par circonscription (Plateforme ouverte des données publiques françaises : <https://www.data.gouv.fr/fr/>)
- relevés de concentration du phytoplancton Dinophysis sur 20 ans ([Ifremer](#))...

Où rechercher des jeux de données ?

Dans **les annuaires et répertoires d'entrepôts :**

- **Multidisciplinaires :**
[re3data](#) répertorie 1400 entrepôts dont 71 en France (mars 2016)
[Datacite](#) Data Citation Index de Thomson Reuters (payant)
- **Spécialisés :** en sciences de la vie et biomédecine, [BioSharing](#) répertorie les bases de données et les standards de métadonnées utiles au partage des données produites par les sciences "omiques", en particulier.

Dans **les data journals** (Liste de *data journals* : <http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList>)

La recherche via Google n'est pas encore possible (octobre 2016).

Qu'est-ce qu'un DOI ?

Un **Digital Object Identifier (DOI)** est un identifiant pérenne qui permet l'identification unique d'un objet physique ou numérique et sa citation. Il fournit un lien stable vers des ressources en ligne, dont les données de la recherche.

Comment citer un jeu de données ?

Pour citer un jeu de données il est nécessaire d'utiliser les **standards de citation** qui recommandent 5 champs obligatoires : auteur, année de publication, titre, éditeur et numéro d'identification (DOI).

L'entrepôt du jeu de données peut préconiser un format de citation.

Le service [DOI Citation Formatter](#) permet de générer une citation à partir d'un DOI.

LE PLAN DE GESTION DES DONNÉES (PGD)

Qu'est-ce qu'un PGD ?

Le **plan de gestion des données (Data management plan - DMP)** est un document rédigé au début d'un projet de recherche et qui définit ce que les chercheurs feront de leurs données pendant et après le projet, explicitant notamment la mise à disposition des données. Le PGD doit être continuellement mis à jour tout au long du projet de recherche. Sa rédaction aide à la **mise en place de bonnes pratiques** à tous les niveaux du cycle de vie des données.

A quelles questions répond un PGD ?

- **Responsabilités dans le projet :**
En quoi consiste le projet ? Quels sont les partenaires impliqués ? Quelle est sa politique en matière de gestion des données ? Qui est responsable de la gestion des données ?
- **Collecte des données :**
Quelles données seront produites / utilisées au cours du projet (type, format, volume et accroissement...) ? Comment seront-elles produites, transformées ?
- **Documentation des données :**
Comment les données seront-elles identifiées, décrites ? Quels standards de métadonnées seront utilisés ? Comment seront générées les métadonnées ?
- **Sauvegarde des données :**
Comment, où, par qui, les données seront-elles stockées, sauvegardées, sécurisées ?
- **Archivage et préservation des données :**
Quel est le plan pour l'archivage et la préservation à long terme ?
- **Ethique :**
Des données sensibles seront-elles produites ou utilisées ? Comment sera assurée l'anonymisation des données ?
- **Propriété intellectuelle :**
Qui sera propriétaire des données produites ? Des données externes seront-elles utilisées ?
- **Accès et partage des données :**
Qui pourra accéder aux données ? Les données seront-elles publiées, partagées ? Avec qui ? Comment ? Dans quel délai ? Quelle licence sera utilisée ?
- **Ressources :**
Comment sera financée la gestion des données, notamment à long terme ?

De quels outils je dispose pour rédiger un PGD ?

Des outils de saisie en ligne :

- DMPOnline (<https://dmponline.dcc.ac.uk/>) : interface web développée par Digital Curation Centre (UK) pour la rédaction en ligne de plans de gestion de données. Intègre les directives H2020. Télécharger la trame de plan *Checklist for a Data Management Plan* : <http://www.dcc.ac.uk/resources/data-management-plans>.
- DMPTool (<https://dmptool.org/>) : interface web développée par l'Université de Californie (US) pour rédiger en ligne des plans de gestion de données selon différents modèles d'agences de financement américaines.

Des modèles, checklists, guides et exemples :

- *Checklist for a Data Management Plan*. DCC (utilisée par DMPOnline). <http://www.dcc.ac.uk/resources/data-management-plans>
- *Data Management Checklist* de Data Archive <http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist>
- *Research data management planning checklist*. Université de Sydney <https://library.sydney.edu.au/research/data-management/>
- Lignes directrices pour la gestion des données dans Horizon 2020. Annexe 1 : modèle de plan de gestion de données (PGD) http://openaccess.inist.fr/IMG/pdf/lignes_directrices_gestion_des_donnees_horizon_2020_version2_1_tr_f_r.pdf
- Réaliser un plan de gestion des données : guide de rédaction. A. Cartier, M. Moysan, N. Reymonet. http://www.univ-paris-diderot.fr/Documents/FCK/recherche/Realiser_un_DMP_V1.pdf
- Guides et exemples : <https://dmptool.org/guidance?method=get&scope1=all>
- *Guidelines for Effective Data Management*. Plans de l'IPCSR : <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html>

Avec qui interagir pour élaborer un PGD ?

L'élaboration d'un PGD implique plusieurs acteurs : **les chargés de projet, les chercheurs producteurs des données, les professionnels de l'IST, le service informatique...**

ÉTHIQUE DU PARTAGE

L'éthique du partage et moi ?

Les avantages au partage :

- Intégrité scientifique
- Contribution au progrès scientifique
- Accroissement de l'impact scientifique et de la notoriété du chercheur
- Préservation des données pour des usages futurs
- Contribution aux avancées de la recherche, de l'enseignement et de la formation

Les freins au partage :

- Questionnement juridique sur le statut des données
- Utilisation abusive des données
- Incompatibilité des données
- Manque d'infrastructure technique
- Manque de sources de financement
- Crainte de perte de primauté scientifique
- Accès restreint aux archives de données

Les leviers au partage :

- Reconnaissance voire rétribution pour le dépôt de données
- Prise en compte des données dans l'évaluation et la carrière scientifique
- Mise en œuvre d'infrastructures pérennes, durables et fiables

WEBOGRAPHIE

- **SITE D'INFORMATION SUR LES DONNEES DE LA RECHERCHE**
- **Compte rendu des journées FRéDoc 2013**
Hameau T., 2013
<http://www.donneesdelarecherche.fr/spip.php?article402>
- **Gestion des données de la recherche**
CIRAD-CoopIST, mise à jour juin 2016
<http://coop-ist.cirad.fr/gestion-de-l-information/gestion-des-donnees-de-la-recherche>
- **Le libre accès aux publications et aux données de recherche.**
Portail français Horizon 2020. MENESR, 2014.
<http://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html>

- **TUTORIELS, MODULES D'AUTOFORMATION :**
- **Dekkers M., Loutas N., De Keyzer M., Goedertier S. (2013).**
Licences pour les données et les métadonnées.
Module de formation 2.5. PWC
<http://fr.slideshare.net/OpenDataSupport/licences-pour-les-donnees-et-les-mtadonnees>
- **INIST-CNRS. (2015).** Tutoriels « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 ». Série de tutoriels multimédias.
<http://www.inist.fr/?-Tutoriels-multimedias-H2020->
- **INIST-CNRS. (2014).** Une introduction à la gestion et au partage des données de la recherche.
Module de sensibilisation en ligne
http://www.inist.fr/donnees/co/Donnees_recherche_web.html

BIBLIOGRAPHIE

- **Dzale E., L'Hostis D. (2015).** *Open Science. Gestion et partage des données de la recherche.* Journée de Formation Agropolis, 2015/04/01, Montpellier (France). Diaporama : 211 slides.
<http://prodirna.inra.fr/record/280536>
- **Gaillard R. (2014).** *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire pour le diplôme de conservateur d'Etat des bibliothèques (DCB). ENSSIB. 104 p.
<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-queelles-politiques-pour-les-donnees-de-recherche>
- **European Commission. (2016).** *Guidelines on open access to scientific publications and research data in Horizon 2020.* 10 p.
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- **European Commission. (2016).** *Guidelines on FAIR data management in Horizon 2020.* 12 p.
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- **INIST-CNRS (trad.). (2014).** *Lignes directrices pour le libre accès aux publications scientifiques et aux données de recherche dans Horizon 2020.* 14 p.
http://openaccess.inist.fr/IMG/pdf/14086_lignes_directrices_la_horizon_2020_tr_fr_versionavril2015.pdf
- **INIST-CNRS (trad.). (2014).** *Lignes directrices pour la gestion des données dans Horizon 2020.* 6 p.
http://openaccess.inist.fr/IMG/pdf/14081_lignes_directrices_pg_horizon_2020_tr_fr_versionavril2015-2.pdf
- **OCDE. (2007).** *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics.* 29 p. Version. 1.0 11 December 2013
<http://www.oecd.org/fr/sti/sci-tech/38500823.pdf>
- **Silvy C. (2015).** *De l'Open Access à l'Open Data : Enjeux et perspectives.* Séminaire CBGP, 2015/01/06, Montpellier (France). Diaporama : 100 slides
<http://www.ist.agropolis.fr/les-formations/tutoriels/item/de-l-open-access-a-l-open-data-enjeux-et-perspectives>