



HAL
open science

Guide pratique de la gestion des données à leur valorisation

Françoise Acquier, Louise-Anne Charles, Laurence Bizien

► **To cite this version:**

Françoise Acquier, Louise-Anne Charles, Laurence Bizien. Guide pratique de la gestion des données à leur valorisation. 2022, pp.23. hal-04923785

HAL Id: hal-04923785

<https://hal.science/hal-04923785v1>

Submitted on 31 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

GUIDE PRATIQUE

DE LA GESTION DES DONNEES A LEUR VALORISATION

INTRODUCTION

Le but de ce document est de **faire le point sur les outils et aides à votre disposition** et de vous guider dans les méandres des données de la recherche **de manière pratique** !

Avant, un peu de contexte...

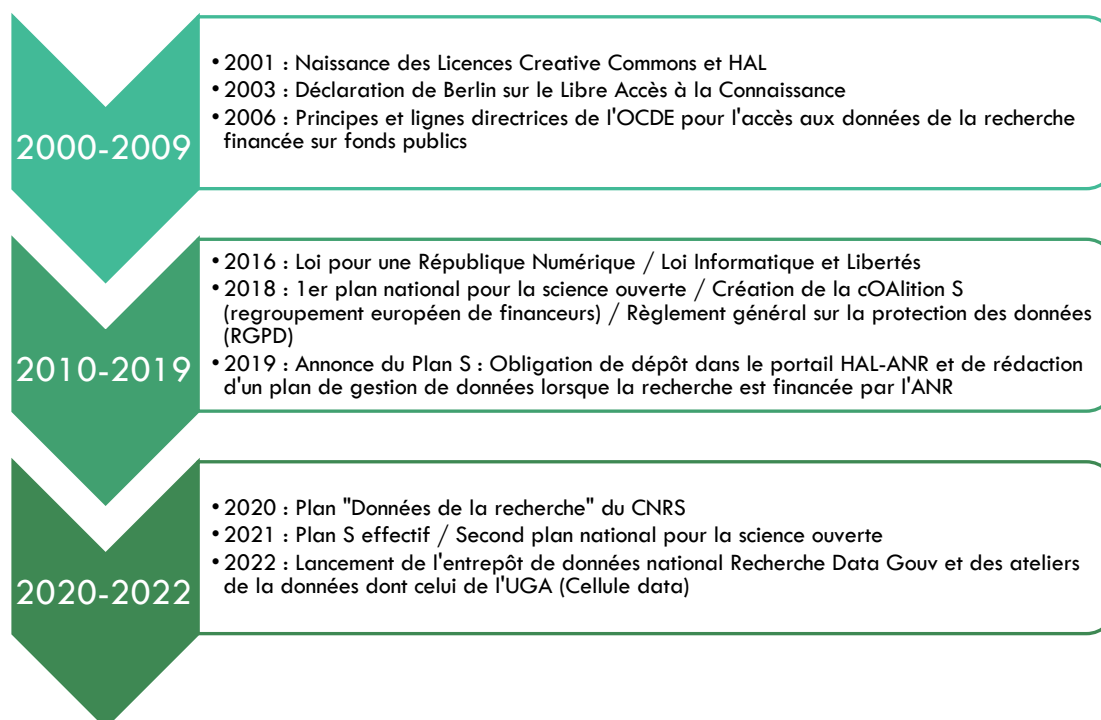
Les données de la recherche relèvent, depuis la **loi pour une République Numérique** de 2016, de données dites publiques quand elles sont financées à plus de 50% par de l'argent public. Elles sont alors des **archives publiques** à mettre à disposition, à l'instar de l'open data dans les administrations et les services publics.

Faisant l'objet d'une attention de plus en plus forte, des politiques volontaristes et incitatives se mettent en place au niveau de l'Etat et de l'Union Européenne.

En France, le premier Plan national pour la science ouverte a été lancé en 2018, dans le but de « généraliser l'accès ouvert aux publications, de structurer et ouvrir les données de la recherche. » Le second plan national (2021) a étendu son périmètre aux codes sources.

Pour soutenir cette démarche, une plateforme nationale fédérée des données de la recherche a été créée, **Recherche Data Gov**, avec la mise en place de plusieurs types d'accompagnement pour les chercheurs¹.

QUELQUES DATES CLES – UN SCHEMA



¹ Les liens vers des textes de références se situent en fin de guide

Cependant, ces objectifs doivent prendre en compte plusieurs aspects juridiques, comme ceux concernant les données personnelles, sensibles, les données protégées par le droit d'auteur. Cela a mené à la maxime « **aussi ouvert que possible, aussi fermé que nécessaire** ».

Cela peut également être un enjeu éthique et écologique, avec une reconnaissance de plus en plus forte du coût écologique du digital.

« Il faut alors **négoier la place de la frontière entre ouvert et fermé**, en inventant une série de zones intermédiaires, à temporalité variable [...] discuter des limites et temporalités de l'ouverture, corpus par corpus, voire dans bien des cas, matériau par matériau. »

- Monica Heintz pour le Collectif ANR Anthropen²

Lorsqu'il n'y a pas d'ouverture des données, le plan de gestion de données (PGD) permet d'**argumenter** ce choix. Ce document, **rendu obligatoire par certains financeurs** comme l'Agence Nationale de la Recherche, sensibilise aux pratiques qui entourent la vie numérique des fichiers. L'objectif est d'également éviter les pertes et les doublons de fichiers, de sensibiliser au stockage, à la sécurité et à la conservation. Le PGD sert de **véritable fil rouge dans les questions à se poser**, de la création de la donnée à sa documentation et in fine à son partage et son archivage.

Comme les termes de « feuille de route » et de « plan » peuvent le suggérer, **il s'agit d'une première mise en œuvre de ce vers quoi nous allons**, alors que de nombreux éléments se mettent encore en place.

Ce présent guide liste les **outils** à votre disposition et les **modèles de documents** que nous avons créés et que **vous pouvez maintenant utiliser**, en s'appuyant sur des retours d'expérience. Autant de petits cailloux qui nous espérons vous aideront lors de vos futurs projets de recherche !

Louise-Anne Charles apprentie documentaliste au laboratoire AAU-Cresson

&

Françoise Acquier documentaliste au laboratoire AAU-Cresson

Avec la relecture de Laurence Bizien documentaliste au laboratoire AAU-Crenau

- Version de Septembre 2022

² HEINTZ, Monica pour le COLLECTIF ANR ANTHROPEN, 2022. La science ouverte et les frontières des données anthropologiques. *INSH CNRS*. [en ligne]. [Consulté le 29 août 2022]. Disponible à l'adresse : <https://www.inshs.cnrs.fr/fr/cnrsinfo/la-science-ouverte-et-les-frontieres-des-donnees-anthropologiques>

TABLE DES MATIERES

Introduction	1
Quelques dates clés – un schéma	1
Un projet de recherche se constitue... que dois-je prendre en compte ?.....	4
Chronologie indicative	4
Au démarrage du projet de recherche.....	6
Choisir un lieu de stockage pour les fichiers.....	6
Créer l'arborescence des dossiers.....	7
Des exemples de nommage et de types d'abréviations.....	8
S'appuyer sur le plan de gestion de données (PGD).....	10
De la collecte à l'analyse	12
Format des fichiers.....	12
Sécurité des données.....	12
Documenter les données : de la captation aux métadonnées	14
Valoriser et publier des données	17
Des outils à votre disposition	17
Une ouverture contrôlée	18
Lier données et publication.....	19
Quoi faire de mes données ? « Fair » du tri !.....	20
Objectif conservation à long terme :	22
Aller plus loin.....	23

UN PROJET DE RECHERCHE SE CONSTITUE... QUE DOIS-JE PRENDRE EN COMPTE ?

- **L'obligation ou non de rédiger un plan de gestion de données** – cela dépend du financeur mais cela pourrait devenir obligatoire ...
- Les conséquences du RGPD (Règlement Général sur la Protection des Données)
 - Lors d'enquêtes qualitatives ou de prises de son ou d'images, devoir réaliser des **formulaire d'autorisation ou de consentement** et des **notes d'information** ;
 - Devoir faire une **déclaration de traitement de données** au Délégué de la Protection des Données (DPD) du CNRS ;
- Il est généralement recommandé d'avoir **une personne référente dans votre équipe projet** tout au long de la durée du projet.

CHRONOLOGIE INDICATIVE

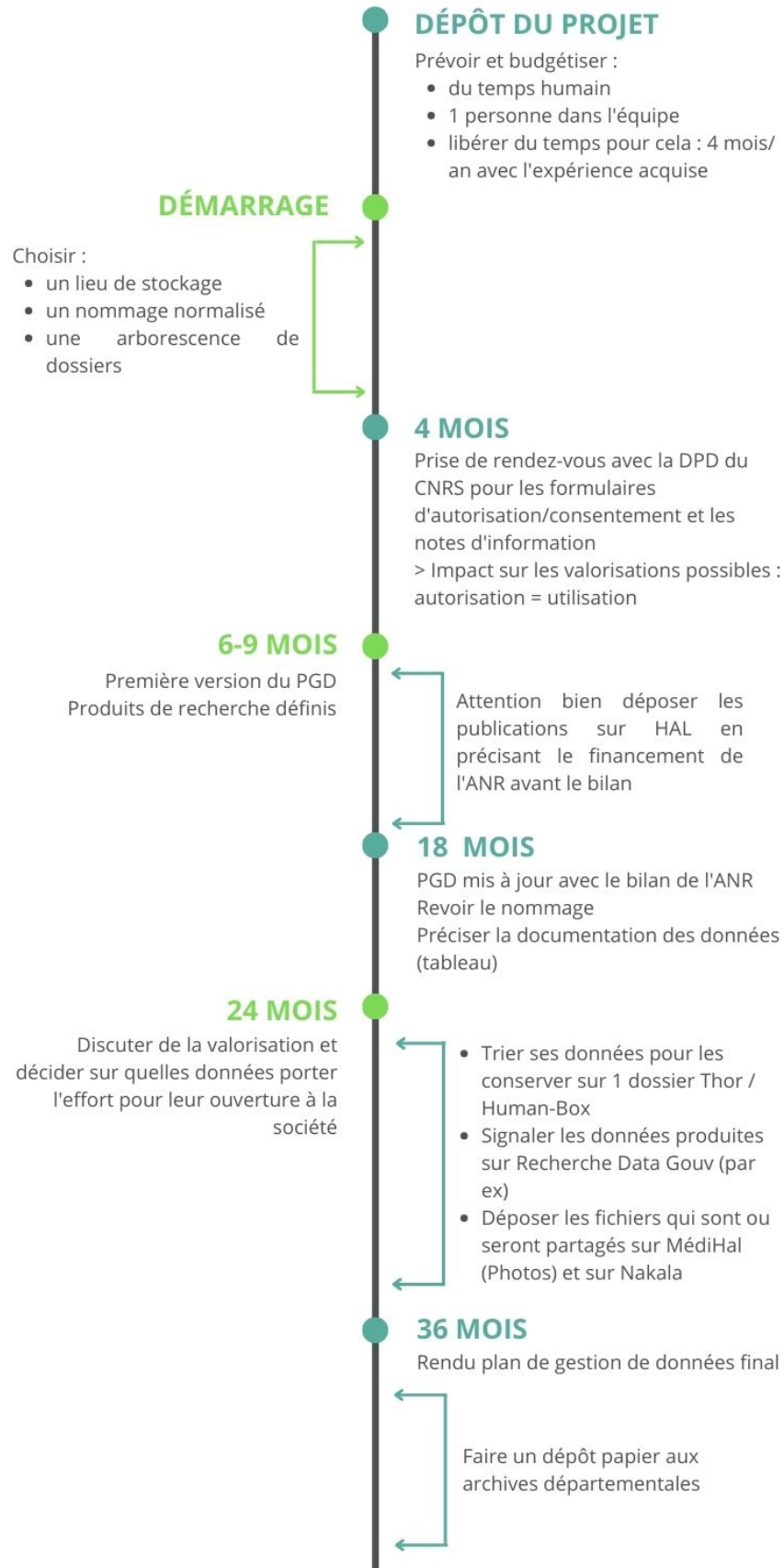
Jusqu'à présent, les expériences ont montré qu'il valait mieux attendre un peu (tout dépend de la durée du projet, à 6 mois pour un projet ANR par exemple) avant de remplir le plan de gestion de données et la déclaration de traitement de données, le temps que les protocoles se stabilisent car beaucoup d'éléments peuvent alors encore changer.

De plus, trois versions du PGD doivent être réalisées, une à 6 mois, une à mi-parcours (pour les projets de plus de 30 mois) et une en fin de projet. La version intermédiaire facilite le suivi de la gestion des données et la rédaction de la version finale.

Du temps doit aussi être pris en compte pour assurer la documentation des données, leur tri et leur ouverture.

FRISE CHRONOLOGIQUE INDICATIVE

Comment et quand prendre en compte la gestion
des données de la recherche dans un projet ?



AU DEMARRAGE DU PROJET DE RECHERCHE

CHOISIR UN LIEU DE STOCKAGE POUR LES FICHIERS

Intérêt : Centraliser les documents pour éviter l'éparpillement et la création de doublons inutiles et donc augmenter la visibilité, créer des méthodes de sauvegarde, et simplifier sa vie en fin de projet.



FICHE D'IDENTITE : SHAREDOSCS

Description :

Outil développé par **Huma-Num** qui permet de stocker, gérer et partager des fichiers, **en ligne et gratuit**. Accès : <https://humanid.huma-num.fr/>

Accès par création d'un compte HumanID et par envoi d'un formulaire en ligne qui décrit le projet.

Pour quels projets ?

- Les thèses (durée de la thèse 3 ans prolongeable)
- Les projets de recherche qui impliquent plusieurs laboratoires (durée du projet)

Avantages :

- ⊕ Stockage sécurisé conforme aux attentes du RGPD ;
- ⊕ A la fois espace personnel et espace collaboratif, avec une gestion des accès qui permet également de partager avec des personnes qui ne sont pas sur Sharedocs (via des liens externes) ;
- ⊕ Téléchargement en lot avec un panier ;
- ⊕ Historique des versions d'un même fichier ;
- ⊕ Prévisualisation des formats standards...

Attention : le travail collaboratif sur un même document ne fonctionne pas bien ; outil non adapté pour faire des montages à partir des fichiers déposés sur l'outil (privilégiez de déposer des versions du montage dont la finalisée).

Prendre en main Sharedocs :

Contact : pour ouvrir un espace Sharedocs pour votre projet : cogrid@huma-num.fr

Références :

Léa Celle et Françoise Acquier, en collaboration avec Christophe Séraudie. 2021. Partager les données de son projet en cours avec ShareDocs d'Huma-Num : l'exemple de l'ANR Sensibilia. *Le Cresson Veille et Recherche*. En ligne sur : <https://lcv.hypotheses.org/16781>

Télécharger ou déposer un nombre important de fichiers avec Webdav : Tuto sur Thor / Documents communs / Docs informatique



FICHE D'IDENTITE : THOR

Description :

Serveur sécurisé de l'équipe CRESSON. Accès : <https://thor.grenoble.archi.fr/>

Pour quels projets ?

- Les thèses
- Les projets de recherche qui impliquent seulement des membres de l'équipe CRESSON
- Ses données de travail

Avantages :

- ⊕ Stockage sécurisé ;
- ⊕ Espace personnel et espace collaboratif ;
- ⊕ Sauvegarde grâce à une synchronisation par QSync ;
- ⊕ Au sein du laboratoire : accès aux dossiers dans le finder (dépôt et téléchargement plus aisé) ...

Prendre en main Thor :

Contact : argoud.d@grenoble.archi.fr

Références : Plus d'info sur Thor / Documents communs / Docs informatique

Au minimum :

- Disque dur (chiffré*) et sauvegarde sur le serveur du laboratoire ;
- Éviter les clouds commerciaux ;
- Faire attention à son ordinateur personnel (ex : éviter de laisser une session ouverte sans surveillance).

* Chiffrer un disque dur, un répertoire ou des fichiers permet d'empêcher la lecture des fichiers tant que la personne ne possède pas la clé de déchiffrement (qu'il ne faut alors pas oublier !). Exemple d'outil : 7-ZIP. Plus d'informations et tutoriels : <https://www.cnil.fr/fr/comment-chiffrer-ses-documents-et-ses-repertoires>

CREER L'ARBORESCENCE DES DOSSIERS

Créer une arborescence qui **fait sens à tous les membres du projet** : cela permet dès le départ de bien organiser les fichiers et de ne pas devoir faire un tri par la suite (plus chronophage).

Il faut par contre **éviter une trop grande profondeur** (trop de sous-dossiers peut entraîner la perte de lisibilité des fichiers entreposés)

A titre d'exemple – une partie de l'arborescence du projet ANR PROSECO :

- AAP_ANR
- Ateliers_Lecture
 - AL1
 - AL2
 - AL3
 - AL4
- Carnet_Hypotheses
 - Billets
 - Formation_Hypotheses
 - Image_Carnet
- Enquetes_1_2_Acteurs_Protocoles_Corpus
 - Corpus
 - E1
 - E2
 - Schemas_Methodo
- Equipe
- Gestion_Donnees
 - DPD
 - Nomenclature_&_arborescence
 - Note_Info_Formulaire_Consentement
 - PGD
 - ZOTERO
- Logos
- Reunions
 - PROSECO_Reunion_20211122
 - PROSECO_Reunion_20220516
 - Etc.
- Terrains
 - Archivage
 - Cartes
 - Charte_graphique
 - Methode_selection_Terrains_a_travailler
 - Terrains_Selectionnes

DES EXEMPLES DE NOMMAGE ET DE TYPES D'ABREVIATIONS

- A utiliser par tous les membres du projet
- Attention aux initiales (elles ne doivent pas correspondre à des personnes différentes !)
- Caractères à éviter : . / +, les accents

20210208|IMA|Affiche_atelier_lecture|LC_FA|VF.jpg
Date Type Titre du document Initiales Version

Figure 1 : Schéma de nommage du projet ANR Sensibilia

Schéma du nommage :

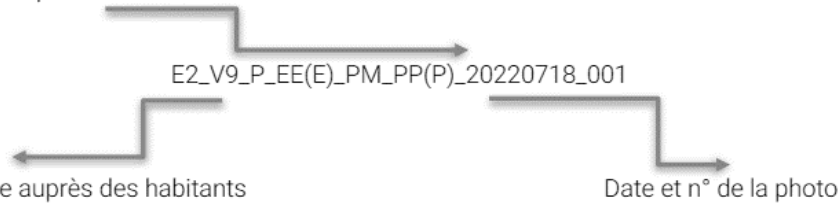
XX_QQ(Q)_Y_EE(E)_TT_PP(P)_DDDDDDDD_NNN

Exemple appliqué à une situation d'enquête auprès des habitants :

EE(E) et PP(P) Enquêté et enquêteur

P = Photo

PM = Photo de passation mettant en scène l'entretien



E2 = Enquête auprès des habitants

V9 = Ecoquartier populaire Villeneuve

Date et n° de la photo

Figure 2 : Schéma de nommage du projet ANR PROSECO

Exemple d'abréviations :

ANNE : Annexe	GUI : Guide	PODCAST : Podcast
AP : Avant-projet	ILL : Illustration	PUB : Publication
BU : Budget	IMA : Image	PRES : Présentation
CdC : Cahier des charges	LECT : Fiche de lecture	PROJ : Projet
CR : Compte-Rendu	LIS : Liste de mot	RAP : Rapport
CON : Convention	LIVRE : Livre en entier (PDF, Epub)	REG : Règle de nommage, écriture
ENQ : Enquête (enregistrement, retranscription, guide d'entretien)	LOGO : logo	SYNT : Synthèse
ETU : Etude	MIX : Mixage d'entretien	TAB : classeur xls
EX : Exemple	NOTE : Note	TEXTE : texte comme « chat »
FACT : Facture	NOTFR : Note de frais	VID : Vidéo inédite
	PHOTO : Photographie	SON : son
	PLAN : Plan	

Figure 3 : Exemples d'abréviations, source :

https://www.unil.ch/uniris/files/live/sites/uniris/files/documents/unites/UNIL_POL_RM_Regles_nommage_annexe_abreviations_VF.pdf

Un [support](#) réalisé par la Cellule Data de l'UGA, et disponible sur leur site, guide également dans les enjeux de nommage et d'arborescence :

« Décrire et expliquer ses données », Formation doctorale : gestion des données de la recherche – 2022.

La Cellule Data est le lieu ressource pour vos questions ! Davantage d'informations sur les actualités et les bonnes pratiques sont disponibles sur leur site : <https://scienceouverte.univ-grenoble-alpes.fr/>

S'APPUYER SUR LE PLAN DE GESTION DE DONNEES (PGD)

Le plan de gestion de données est un **document qui centralise toutes les informations relatives aux données d'un projet de recherche** (collecte, stockage, partage...). Il est **évolutif** : dans la V1, vous n'êtes pas obligé de répondre à toutes les questions. En effet, d'une part vous n'avez pas encore toutes les réponses et d'autre part la stratégie du projet en matière de conservation, publication et partage de certains jeux de données n'est pas complètement définie. Cette 1^{re} version pose certaines questions pour vous aider à anticiper des problèmes d'ordre juridique, éthique ou technique.

Pour être sûr de bien documenter votre PGD, mettez-vous à la place d'un chercheur qui arriverait en cours de projet et devrait se repérer dans vos fichiers de données

Le document se veut être une **aide concrète à la gestion des données**, durant et après le projet, dans le sens où il permet de se poser des questions et de réfléchir aux besoins à venir (menant par exemple à des demandes de financement supplémentaires).

Il reste un document **normalisé** et de nombreux modèles ont été créés par des financeurs ou des institutions, de nombreux conseils, recommandations et exemples sont disponibles pour vous aider. Il est possible de les retrouver sur l'outil DMP OPIDoR³, développé par l'INIST, et de rédiger directement le PGD sur celui-ci.

Les avantages de DMP OPIDoR :

- ⊕ Accès à des modèles et des recommandations ;
- ⊕ Importer directement les informations relatives au projet, lorsqu'il s'agit d'un projet ANR, et auto-complétion de certaines rubriques ;
- ⊕ Diffuser/publier le PGD ;
- ⊕ Ecriture collaborative...

Le PGD se structure en « **produits de recherche** ». Un produit de recherche est un ensemble de données qui est :

- Soit **homogène par leur format** (par exemple un ensemble de photographies, d'enregistrements sonores...) ;
- Soit **intellectuellement homogène** (un chantier dans le projet Sensibilia ou les baluchons sensoriels dans le projet PROSECO, qui mêlent tous deux photos, entretiens, enregistrements sonores...).

Comme leur traitement sera similaire ou fait de manière conjointe (stockage, collecte, analyse...), il est possible de les regrouper. C'est **à vous de définir** ce qui est un produit de recherche et ce qui ne l'est pas.

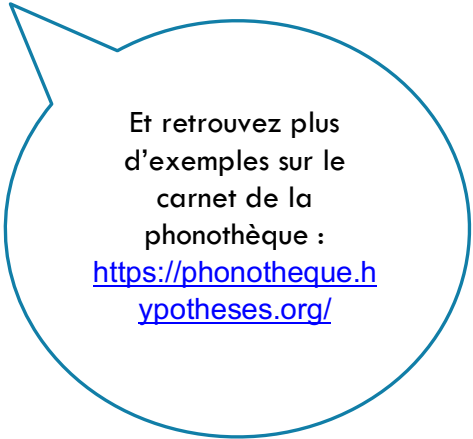
Un **modèle de PGD commenté est mis à disposition** des membres du laboratoire, composé à la fois de commentaires pour comprendre et d'exemples. Il peut être **utilisé pour rédiger son PGD**. Chemin : Thor / Documents communs / Doc Documentation / Plan_Gestion_Donnees

³ Lien Web : <https://dmp.opidor.fr/> accès institutionnel Janus ou individuel. Pour prendre en main l'outil, un guide a été réalisé par l'Université Paris Cité sur le modèle ANR. Son lien est dans la partie « Aller plus loin ».

Faire relire son PGD et être accompagné sur les questions de gestion des données : Cellule Data UGA sos-data[at]univ-grenoble-alpes.fr

Exemples de PGD :

- Première version du PGD du projet ANR PROSECO, consultable sur Zenodo : https://zenodo.org/record/6463020#.YIIA0bg6_UI
- Exemple de Plan de Gestion de Données rétrospectifs pour des fonds des archives de la recherche : <https://phonotheque.hypotheses.org/33813>



Et retrouvez plus
d'exemples sur le
carnet de la
phonothèque :
<https://phonotheque.hypotheses.org/>

DE LA COLLECTE A L'ANALYSE

FORMAT DES FICHIERS

Privilégier des formats dits ouverts pour les fichiers dans la mesure du possible, c'est :

- Permettre un accès au fichier, peu importe le système d'exploitation et le logiciel utilisés ;
- Garantir une meilleure pérennité pour le fichier dans le temps.

Liste non-exhaustive de formats recommandés par le Cinès (Centre Informatique National de l'Enseignement Supérieur) via son site <https://facile.cines.fr/>, pour :

- Les sons : Wav, MP3 ;
- Les vidéos : MP4 ;
- Les photographies : JPEG ;
- Les textes : ODT et TXT, PDF ;
- Les tableurs : CSV.

Attention aux formats utilisés automatiquement par votre téléphone ! Généralement, vous pouvez modifier le format de la prise de vue ou de son dans les paramètres.

L'outil Sharedocs permet également de convertir les fichiers vers d'autres formats (hanTools).

SECURITE DES DONNEES

LA DIFFERENCE ENTRE DONNEES PERSONNELLES ET DONNEES SENSIBLES

Données personnelles : donnée qui permet d'identifier directement ou indirectement une personne physique (un numéro de téléphone, une date de naissance par exemple).

Données sensibles : orientation et vie sexuelle, opinions politiques, philosophiques ou religieuses, origines, données de santé, appartenance syndicale

- **Notes d'information** : à distribuer aux personnes concernées lors du recueil des données pour expliquer le but du projet, l'utilisation et leurs valorisations des données, pourquoi et combien de temps ;
- **Formulaires d'autorisation** (lorsqu'il y a collecte de données personnelles, prise de son ou image) ;
- Ajout de la **demande de consentement** au formulaire d'autorisation (lorsqu'il s'agit de données sensibles).

Des modèles de chacun de ses documents ont été réalisés et se trouvent sur THOR Documents Communs / Docs Documentation / Données personnelles. De nombreux modèles en fonction des situations rencontrées sont proposés. **Le service de la Déléguée à la Protection des Données personnelles aide à la l'adaptation des formulaires** en fonction des situations de recherche.

Les formulaires ne doivent pas être rédigés trop tôt, il faut attendre que les protocoles d'enquête soient stabilisés et que vos intentions d'ouverture des données aient pris forme.

Contact : dpd.demandes@cnrs.fr

Note : Les mesures appliquées pour satisfaire au règlement général sur la protection des données (RGPD) doivent être expliquées dans le PGD.

QUELLE DIFFERENCE ENTRE ANONYMISATION ET PSEUDONYMISATION ?

Les données anonymisées : “de **manière irréversible**, qui ne permettent plus la réidentification d’une personne, ne sont pas soumises à la réglementation sur la protection des données personnelles”.

Les données pseudonymisées : “sont les données à caractère personnel **qui ne peuvent plus directement être attribuées à la personne concernée**. Mais le recours à des informations supplémentaires, par exemple une table de correspondance, permet de réidentifier cette dernière. Dans ce cas, la réglementation sur la protection des données personnelles s’applique”.

Anonymiser des entretiens est difficile sans en perdre l’intérêt. Pour les enquêtes qualitatives, on parlera plutôt de pseudonymisation en remplaçant les noms avec un code et en ayant recours à une table de correspondance. L’objectif est de conserver le contenu des fichiers audio d’une part et d’autre part leur retranscription à des fins d’analyse.

Pour plus d’informations : La source pour cette partie est le *Guide pour la recherche en SHS : les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte*, disponible à l’adresse suivante :

https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/Guide_rgpd_2021.pdf

DOCUMENTER LES DONNEES : DE LA CAPTATION AUX METADONNEES

Le minimum des informations/métadonnées attendues :

- Titre ;
- Auteur(s) ;
- Date ;
- Type de document ;
- (Langue).

Selon les besoins de la recherche et le type de fichier, le type et le nombre des métadonnées évoluent, ainsi que la méthode pour procéder. La documentation peut également se faire sous **la forme d'un tableur**.

EXEMPLE POUR LES PHOTOGRAPHIES :

- Vérifier que les informations sur la **date et l'heure sont à jour** sur les appareils photos pour que ce soit automatiquement dans les données EXIF (données embarquées) des photos ;
- Décider **d'un nommage**, plus ou moins précis selon l'utilisation de métadonnées IPTC ;
- Possibilité d'utiliser **les métadonnées IPTC** soit à partir des paramètres ou à partir de logiciel qui permet de remplir ces métadonnées et de renommer par lot d'images. Exemple de métadonnée IPTC : auteur(s), titre, ville, pays, copyrights, droits et conditions d'utilisation
- Possibilité également **d'ajouter des mots-clés**, de préférence à partir d'un thésaurus ou d'une liste de mots prédéfinie.

Liste non exhaustive de logiciels qui permettent de gérer des photographies : XnView, TROPY - Research Photo Management, Bridge ...

Contact et source : carbonnelle.a@grenoble.archi.fr

EXEMPLE POUR LES ENREGISTREMENTS SONORES :

Sur le terrain :

- Les enregistreurs – conseils généraux :
 - S'assurer que la date du jour est incrémentée automatiquement et si possible les coordonnées GPS ;
 - Les piles chargées et en avoir pour les changer ;
 - Prévoir un enregistreur de secours : éviter d'utiliser votre téléphone portable, le format est parfois non récupérable et de qualité moyenne ;
 - Prévoir une bonnette. En cas de vent sans bonnette, le son sera inaudible ;
 - Carnet de note : météo, localisation, parcours de l'enregistrement...
 - Contacter juL, David, Cédric pour des conseils plus précis.

- Enregistrer l'environnement sonore :
 - Le format doit être le wav ;
 - Prévoir un pied (prise de son statique) ou une perche ou tête artificielle (prise de son mobile).
- Enregistrer un entretien oral :
 - Le format peut être le mp3, surtout si l'entretien est très long et uniquement destiné à être retranscrit ;
 - Faire signer le consentement en double pour en conserver un ex avec tous les choix retenus par l'enquêté.

De retour de terrain :

Exemple projet ANR PROSECO :

- En fin de terrain, télécharger les enregistrements, les photos, les formulaires d'autorisation sur votre ordinateur chiffré et sur ShareDocs, création d'un dossier par enquêté ou par type de document
- Tableur du terrain :
 - Pour les habitants (pour varier l'échantillon des personnes enquêtées) : enquêteur, prénom, âge, date, logement, sexe, commentaire
 - Pour les professionnels inclure aussi profession, structure.
 - Les conditions de passation sont également ajoutées : durée, conditions, si des documents ont été transmis, s'il y a une carte du parcours.

Exemple de métadonnées (fragments d'environnement sonore) :

Un modèle de tableur et de fiche est disponible dans Thor/Cressound/Procédures.

- Titre du fragment
- Nom fichier son
- Responsable de la recherche
- Preneur de son
- Jour, heure de l'enregistrement
- Résumé
- Mots-clés
- Durée de l'enregistrement
- Latitude
- Longitude
- Qualité du son, matériel utilisé
- Format
- Type d'enregistrement

EXEMPLES DE THESAURUS A UTILISER POUR DES MOTS-CLES :

ArchiRès : "Thésaurus de mots-clés hiérarchisés bilingue, avec des définitions, dans les domaines de l'architecture, l'urbanisme, du paysage ", utilisé par le centre de documentation du CRESSON et les bibliothèques du réseau ArchiRès, évolutif. C'est un des vocabulaires référentiels Humanum (Isidore, Nakala).

Disponible à l'adresse : http://archives.afisa.net/userfiles/file/vie_reseau/thesaurus_alpha_2020_etudiant.pdf

IDRef : plateforme qui réunit les notices d'autorité de l'enseignement supérieur et de la recherche (Sudoc, Calames, Star) dont des sujets issus de RAMEAU. Disponible à l'adresse : <https://www.idref.fr/> (bien cocher "nom commun"). Une liste indicative de noms communs, provenant de cette plateforme, déjà utilisés dans divers projets (Nakala, Canal-U), est également disponible pour vous : Thor / Documentation / Thésaurus Cresson

VALORISER ET PUBLIER DES DONNEES

DES OUTILS A VOTRE DISPOSITION

MEDIHAL

- **Archive ouverte interdisciplinaire** dédiée aux données visuelles et sonores, particulièrement adaptée aux images fixes. Celles-ci apparaîtront sur le CV HAL et sur les notices de publication liées. Accès : <https://medihal.archives-ouvertes.fr/>
- Il est possible de **géo-référencer les images**, et celles qui auront des métadonnées GPS seront automatiquement positionnées sur la carte. Médihal est essentiellement utilisée pour publier des images d'archives : fonds photographiques de chercheur, photos de fouilles archéologiques, cartes numérisées...

NAKALA

- **Entrepôt de données SHS** développé par Huma-Num. Accès par création d'un HumanID et formulaire qui décrit le projet de recherche. Une fois le compte ouvert, Nakala peut être utilisé pour tout autre projet. Accès : <https://nakala.fr/>
- Le projet Les Ondes de l'Eau (AAU-CRESSON / Regards des lieux) a déjà publié des données sur cet entrepôt. Par la suite **un guide a été réalisé pour accompagner vos futurs dépôts, ainsi qu'un tableur pour accompagner la réalisation des métadonnées** : Thor / Documents communs / Docs documentation.

Référence : Louise-Anne Charles et Françoise Acquier, "Retour d'expériences de publication de données : l'entrepôt de données SHS Nakala," in *Lab & doc*, 22/07/2022
<https://labeledoc.hypotheses.org/10424>

RECHERCHE DATA GOUV

- Entrepôt de données qui vise à **fédérer l'ensemble des données de la recherche en France** qu'elles soient déposées directement sur cette interface ou via des entrepôts dédiés comme Nakala d'Huma-Num. Accès : <https://recherche.data.gouv.fr/fr>

CRESSOUND / CARTOPHONIES

- Projets développés par le CRESSON qui partagent les enregistrements sonores effectués dans le cadre d'un projet de recherche (environnement sonore)
- La carte interactive intitulée « Cartophonies » contribue à construire une connaissance des milieux habités, du vécu des espaces et des ambiances contemporaines, celles du passé proche comme celles du futur. Cartophonies permet de collecter des fragments sonores situés et datés, commentés et classés, enrichis par des schémas ou photos issus de différentes recherches ou travaux portant sur la ville et sur l'architecture existantes. En écoutant ces fragments, on peut explorer la diversité des expériences auditives

comme des productions et actions phoniques ordinaires ou particulières. Accès : <https://www.cartophonies.fr/>

Pour contribuer : <https://www.cartophonies.fr/participe/> (GPS / parcours obligatoire, photos et schémas possibles pour illustrer le contexte)

- Pour être ajouté au catalogue de la bibliothèque du Cresson et à Cressound, contactez Françoise ou juL.

(Liste non exhaustive)

UNE OUVERTURE CONTROLEE

Lors du dépôt dans un entrepôt de données, il est possible de cadrer l'ouverture de la publication des données par deux moyens :

- La mise en place d'un embargo ou d'un accès restreint : l'utilisateur n'a accès qu'aux métadonnées et non à la donnée elle-même sans en faire la demande. Cet accès restreint est soit permanent soit lié à une date (et devenir accès ouvert après publication d'articles par exemple) ;
- L'utilisation des licences : l'utilisateur a accès à la donnée mais son utilisation est plus ou moins cadrée par la licence.

Nakala : Embargo ou accès ouvert, toujours avec un fichier
Recherche Data Gouv : Embargo, accès restreint, accès ouvert, notice sans fichier

Les Creative Commons sont un exemple de licence qui peuvent être utilisés. Ci-dessous un schéma représente les différents degrés de licence possibles :

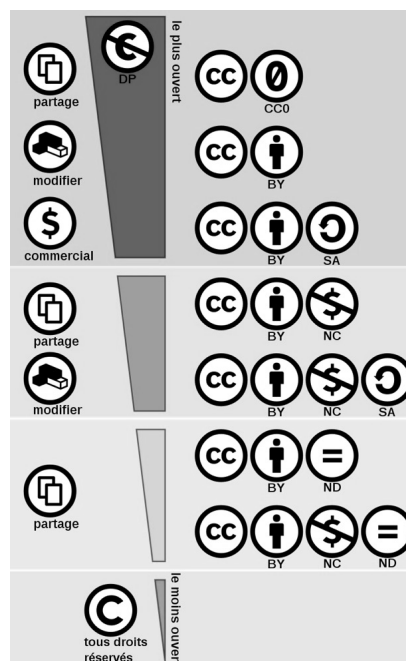


Figure 4 : Les familles de licences Creative Commons : commercial / non commercial ; modifiable / non modifiable. Les licences vont de la moins restrictive (DP en haut) à la plus restrictive (© en bas), Simon Villeneuve, Wikipédia, 2017, CC-BY- SA. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Licence_Creative_Commons#/media/Fichier:Creative_commons_license_spectrum_fr.svg

Attention, en France, **il faut utiliser “tous droits réservés”** et non le copyright.

Le choix des licences peut également **évoluer avec le temps**, être moins restrictif lorsqu’il y a eu publication.

LIER DONNEES ET PUBLICATION

Il est recommandé de **déposer les données avant publication** afin de pouvoir utiliser les identifiants (pérennes et uniques) attribués aux données par l’entrepôt choisi du type DOI.

Si vous avez déposé un jeu de données dans un entrepôt (entrepôt Nakala ou généraliste comme Zenodo), vous pouvez utiliser dans HAL le champ « Données associées » pour ajouter le DOI attribué à vos données ou bien son url.

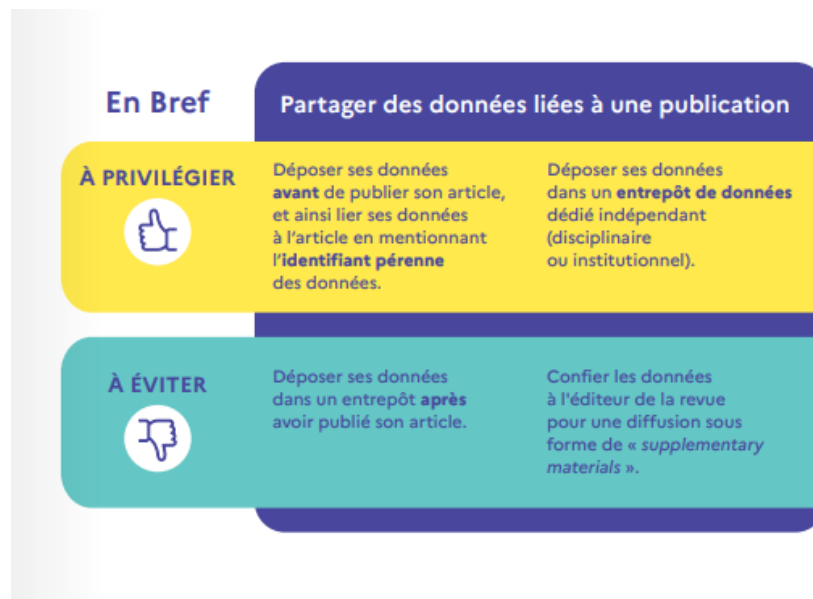


Figure 5 : COLLEGE « DONNEES » DU COMITE POUR LA SCIENCE OUVERTE, 2022. « En bref : Partager des données liées à une publication », *Partager les données liées aux publications scientifiques* [en ligne]. Guide. [Consulté le 29 août 2022]. Disponible à l'adresse : <https://www.inist.fr/nos-actualites/en-image-comment-partager-les-donnees-liees-aux-publications-scientifiques/>

QUOI FAIRE DE MES DONNEES ? « FAIR » DU TRI !

Même si elles ne sont pas ouvertes, les données doivent tout de même répondre aux principes FAIR : Findable, Accessible, Interoperable, Reusable. C'est-à-dire : nous devons pouvoir les trouver, les comprendre, y accéder, quel que soit le système utilisé, et pouvoir les réutiliser, sur le long terme. C'est une des raisons qui expliquent l'importance de la documentation et du format des données.

La solution proposée est de se servir du serveur sécurisé Thor pour conserver des données après la fin du projet de recherche (comme cela a pu être le cas auparavant avec les enregistrements sonores) en les contextualisant.

Ce ne sera pas possible de conserver l'ensemble des données produites dans le cadre d'un projet.

“Il y aura nécessairement une allocation des moyens dans la mesure où **tout ne pourra pas être bien stocké, bien décrit et bien communiqué** : il faut réfléchir à la manière de la réaliser. On gagnerait dès lors à dire que **la science ouverte implique des choix**. [...] Une bonne pratique devrait alors inclure une destruction programmée d'une partie des données, ce qui permet en contrepartie **un traitement approprié des données conservées**.”

- Usage et gouvernance des données⁴ (2019)

En effet, plusieurs aspects sont à prendre en compte :

- Scientifique (est-ce que c'est intéressant de les garder notamment dans le long terme ? Est-ce qu'elles sont liées à des publications ?) ;
- Écologique ;
- Financier ;
- Et les moyens à disposition.

Il est alors important d'identifier les données qui doivent être supprimées, celles qui doivent être conservées sur le serveur sécurisé THOR et celles conservées par vous. La durée de conservation (DUA **Durée d'utilité administrative**) est à définir à partir du [référentiel](#), réalisé par le groupe de travail AURORE des archivistes de la recherche.

Le support « Trier et conserver ses données »⁵ réalisé par la Cellule Data Grenoble Alpes met en avant les **questions** suivantes pour décider : est-ce que ces données existent déjà ? Sont-elles uniques ? Sont-elles déjà conservées par un autre organisme ?

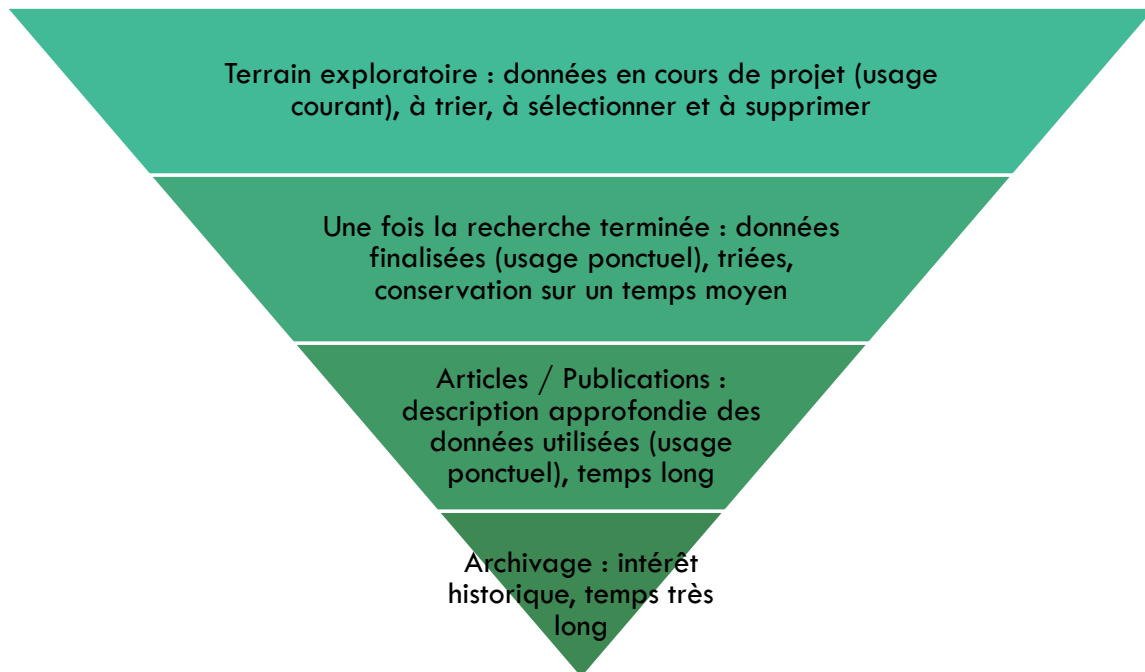
⁴ CHAMBEFORT, Hélène, HUEBER, Juliette, LEMERCIER, Claire, MAUSSANG, Kenneth et VANET, Anne, 2019. *Usage et gouvernance des données*. [en ligne]. Rapport. Comité pour la science ouverte. [Consulté le 11 août 2022]. Disponible à l'adresse : <https://www.ouvrirlascience.fr/usage-et-gouvernance-des-donnees/>

⁵ CELLULE DATA GRENOBLE ALPES, 2022. *Trier et conserver ses données*. Support réalisé dans le cadre de la formation doctorale : gestion des données de la recherche – 2022. [en ligne] [Consulté le 29 août 2022]. Disponible à l'adresse : https://scienceouverte.univ-grenoble-alpes.fr/wp-content/uploads/2022/06/Jour_2_-_3_Trier_donnees.pdf

Est-ce qu'il y a un intérêt historique ? Pourquoi garder ? Qui est-ce que ça peut intéresser ?
Quel public potentiel cela pourrait intéresser ?

Ce support propose également un tri **par finalité** pour les données uniques que vous créez, selon un schéma que nous reproduisons ici en le modifiant :

Schéma : tri par finalité des données



Ce tri sera bien plus simple à effectuer si les étapes précédentes ont été effectuées : le stockage des données centralisé, le nommage normalisé, PGD rempli et la documentation faite pour les données importantes.

Cependant, si vous êtes face à un vrac numérique (ça arrive), pas de panique ! Il est toujours temps de refaire une arborescence de dossiers, un schéma de nommage etc... Ainsi il est possible de renommer des lots de fichier (voir David !) et d'utiliser des logiciels comme [Archifiltre](#), qui permet d'analyser les dossiers en repérant les doublons.

OBJECTIF CONSERVATION A LONG TERME :

Pour la **conservation à long terme**, il a déjà été fait mention de Thor, mais il est également possible **d'utiliser Human-Box**, développé par Huma-Num.



FICHE D'IDENTITÉ : HUMAN-BOX

Description :

- Serveur de fichiers sécurisé, développé par Huma-Num avec des fonctionnalités supplémentaires qui permettent de garantir une préservation entre 5 à 10 ans ;
- Il est indépendant des autres services d'Huma-Num

Pour quelles données ?

- Ce sont pour les données de projet dites « froides » donc des fichiers peu modifiés et peu consultés. Le volume peut atteindre plusieurs Téraoctets.

Prendre en main Human-Box :

Contact : assistance@huma-num.fr

Référence : <https://documentation.huma-num.fr/humanum-box/>

Concernant les données que vous avez identifiées comme ayant une valeur historique, une stratégie d'archivage peut être engagée :

L'archivage **des données de la recherche**, notamment celles numériques, est une question en chantier pour les AD (Archives Départementales de l'Isère) et l'UGA et qui devrait évoluer d'ici quelques années. A ce jour les archives recommandent le versement d'un dossier représentatif de la recherche au format imprimé. Cependant les AD privilégient des données qui ont un **caractère local** et/ou qui sont capables d'être **représentatives des travaux du laboratoire**.

Aujourd'hui seul le CINES (Centre Informatique National de l'Enseignement Supérieur) a les capacités de conserver les fichiers numériques. Par exemple HAL collabore avec le CINES, ce qui garantit l'archivage des publications. Concernant les données de recherche Huma-Num peut accompagner des projets d'archivage avec cet opérateur.

Contact : Archiviste à l'UGA : Luce Forget, luce.forget@univ-grenoble-alpes.fr

ALLER PLUS LOIN

La loi pour une République Numérique, des liens vers :

- L'article 30 LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique https://www.legifrance.gouv.fr/loda/article_lc/JORFARTI000033202841/
- La note de Couperin : <https://scienceouverte.couperin.org/donnees-recherche-bonnes-pratiques/>

Sélection de plans :

- Deuxième Plan national pour la science ouverte Généraliser la science ouverte en France 2021-2024 <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>
- Plan « Données de la recherche » du CNRS de 2020 <https://www.science-ouverte.cnrs.fr/actualite/plan-donnees-de-la-recherche-du-cnrs/>
- La rédaction d'un plan de gestion de données est devenue obligatoire pour les programmes financés par l'ANR depuis 2019, et fortement recommandée par le Ministère Français de l'Enseignement Supérieur, de la Recherche et de l'Innovation (Plan national pour la Science Ouverte) comme par l'Europe (H2020 et European Research Council)

Références utilisées :

HEINTZ, Monica pour le COLLECTIF ANR ANTHROPEN, 2022. La science ouverte et les frontières des données anthropologiques. *INSH CNRS*. [en ligne]. [Consulté le 29 août 2022]. Disponible à l'adresse : <https://www.inshs.cnrs.fr/fr/cnrsinfo/la-science-ouverte-et-les-frontieres-des-donnees-anthropologiques>

CHAMBEFORT, Hélène, HUEBER, Juliette, LEMERCIER, Claire, MAUSSANG, Kenneth et VANET, Anne, 2019. *Usage et gouvernance des données*. [en ligne]. Rapport. Comité pour la science ouverte. [Consulté le 11 août 2022]. Disponible à l'adresse : <https://www.ouvrirlascience.fr/usage-et-gouvernance-des-donnees/>

D'autres outils :

- Guide pour prendre en main DMP OPIDoR :

UNIVERSITÉ PARIS CITÉ, 2022. Guide d'aide à la rédaction d'un plan de gestion des données à Université Paris Cité – Modèle ANR. [en ligne] Disponible à l'adresse : <https://cloud.parisdescartes.fr/index.php/s/2GaENEJp9xdBtsP#pdfviewer>

DMP OPIDoR a également réalisé une liste de ressources sur son site : <https://dmp.opidor.fr/static/resources> dont le site complet et abordable DORANum : <https://doranum.fr/>

- Sécuriser ses mots de passe :

KeePassXC, gestionnaire de mots de passe gratuit : <http://keepassxc.org/download/>. Un tuto est disponible sur Thor / Documents communs / Docs informatique.

Pour savoir si votre mail et/ou votre mot de passe ont été corrompus : <https://haveibeenpwned.com/>