



HAL
open science

Gestion des données de recherche : bonnes pratiques

Bénédicte Macgregor

► **To cite this version:**

Bénédicte Macgregor. Gestion des données de recherche : bonnes pratiques. Master. méthodologie de l'enquête, Université Paris 8 - Vincennes Saint-Denis, France. 2023. halshs-04240283

HAL Id: halshs-04240283

<https://shs.hal.science/halshs-04240283>

Submitted on 13 Oct 2023

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.

Gérer des données de recherche : bonnes pratiques

Bénédicte MacGregor
Chargée de données scientifiques
LADYSS - CNRS





Plan

1. Les données de la recherche: Tentative de définitions
2. Pourquoi faut-il gérer les données ?
3. Organiser les données
4. Les formats de fichiers
5. Stocker les données
6. Documenter les données
7. Enjeux juridiques (RGPD...)
8. En savoir plus (Webographie)

1. Les données de la recherche: définitions (1/2)

Des enregistrements **factuels** (chiffres, textes, images, sons) qui sont utilisés comme **sources principales** pour la recherche scientifique et sont reconnus par la communauté scientifique comme **nécessaire pour valider des résultats de recherche**.

OCDE, Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics, 2007.

Définition plus large: TOUT (ce qui est collecté, analysé, interprété pendant le processus de recherche).





1. Les données de la recherche (2/2)

Concrètement, de quoi parle t-on en sciences sociales ?

- Des enquêtes quantitatives par questionnaires (par téléphone, en face à face, en ligne)
- Des entretiens qualitatifs (audio ou vidéo)
- Des données géographiques (collectées par GPS, objets connectés)
- Des données de réseaux (sites Web, forums, données relationnelles)
- ...

En SHS, on désigne également les matériaux de recherche avec des termes comme « sources », « corpus », « traces », « enquêtes ».

Différentes formes (cartes, CD, tableaux) et différents supports (nativement numériques, analogiques (livres, photos, objets), numérisés).

2. Pourquoi faut-il (bien) gérer les données ? (1/2)

Les données sont fragiles. Gérer les données pour les protéger des risques.

- L'obsolescence, le vieillissement du support, l'oubli, la destruction

Exemples :

- Le lecteur fonctionne mais le connecteur ne se branche pas sur le PC (obsolescence)
- J'ai 43 copies du même document sur des supports différents, à des stades d'avancement différents, j'ai fait des modifications sur une ancienne version qui ne contient peut-être pas les ajouts récents. (oubli)



2. Pourquoi faut-il (bien) gérer les données ? (2/2)

≡ **The Mercury News**

Tech accidentally wipes out info on Alaska's...

NEWS

Tech accidentally wipes out info on Alaska's \$38 billion fund



By **ASSOCIATED PRESS** |
March 20, 2007 at 1:37 a.m.

JUNEAU, Alaska (AP) – Perhaps you know that sinking feeling when a single keystroke accidentally destroys hours of work. Now imagine wiping out a disk drive containing information for an account worth \$38 billion.

That's what happened to a computer technician reformatting a disk drive at the Alaska Department of Revenue. While doing routine maintenance work, the technician accidentally deleted applicant information for an oil-funded account – one of Alaska residents' biggest perks – and mistakenly reformatted the backup drive, as well.

There was still hope, until the department discovered its third line of defense, backup tapes, were unreadable.

Un technicien écrase les données fiscales d'un fond de financement de 38 milliards de dollars en reformatant un disc dur. Les deux sauvegardes sont inutilisables. ([source](#))

Un exemple de fausse manipulation



3. Organiser les données

Bonnes pratiques

- Adopter des règles explicites et les appliquer

- Organiser sa documentation et ses sources

Exemple : Zotero ou un autre logiciel de gestion bibliographique

- Organiser les dossiers de son ordinateur

Exemple : Organisation hiérarchique

Exemple : Isoler les données brutes dans un dossier à part et ne pas y retoucher



3. Organiser les données

Plusieurs possibilités

- Par activités de recherche (collecter des données, publier, enseigner...)
- Par étapes de traitement des données (collecte, transcriptions, analyse...)
- Chronologie
- Géographie
- Par jeux de données/producteurs de données
- Par type (administration, données, publications...)

Règles de nommage des fichiers

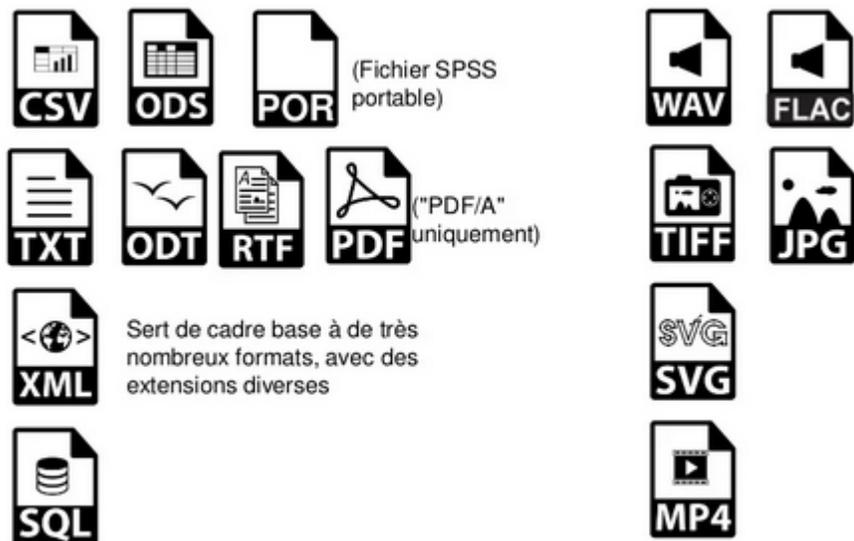
- Noms courts
- Pas de ponctuation
- Pas d'espaces
- Pas de caractères spéciaux
- Pas de caractères accentués
- Ajouter un chiffre devant les noms pour éviter le tri alphabétique (01_DONNEES ; 02_BIBLIO)

Exemples : 2012-03-07_SujetA_Audio.mp3 ; 2012-03-07_SujetA_Transcription-brut.docx

4. Formats des fichiers

Choisir des formats ouverts qui sont utilisables librement, gratuitement, qui sont bien documentés et non attachés à un logiciel spécifique.

Formats dont la pérennité est garantie



Source : Mathieu Saby

Voir : [Qu'est ce qu'un format de fichier ? CINES](#)

Voir : [les formats recommandés par UK Data Service](#)

4. Formats ouverts ? Formats fermés ?

On entend par format ouvert «un format de données dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre »

-> format utilisable sans demander d'autorisation

Ex : ODT, XLSX, ODS, XML ; DOC et XLS depuis 2006 (Open Specification Promise)

Format fermés (ou propriétaires) : format non ou mal documenté et/ou dont l'utilisation est payante ou soumise (au moins en théorie) à autorisation

Ex : PSD, JPEG, logiciels SPSS, STATA, SAS, NVIVO

4. Formats ouverts et fermés

En pratique, on peut souvent travailler avec un format fermé populaire et le convertir en format ouvert.

Mais il faut vérifier que la conversion n'altère pas les informations, et prendre des mesures de compensation si nécessaire.

Exemple : la conversion XLSX -> CSV perd les mises en forme



Où stocker les données ?

Il n'y a pas que Google dans la vie



5. Stocker les données en sécurité

Stockage et sauvegarde sécurisé selon vos usages

Questions: Aller vous partager vos données ? Est-ce que ce sont des données publiques ou confidentielles ?

Il existe plusieurs solutions de stockage

- Réseau sécurisé de votre institution
- Cloud commercial ou proposé par votre institution
- Supports externes (clés USB, disque dur externe, DVDROM)

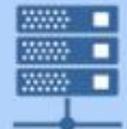
Bonnes pratiques: Sauvegardes régulières et fréquentes.

La règle de la sauvegarde 3-2-1

3 copies sur 2 supports différents, avec au moins 1 une copie à distance

Exemple: L'original sur votre ordinateur, une copie sur le réseau de l'institution et une autre copie sur un disque dur)

5. Stocker les données en sécurité

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes	★☆☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★★★★ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★★★★ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★☆☆ Coût assez important mais pas forcément répercuté sur l'utilisateur	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★★★★ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données

Source : Doranum



5. Stocker les données

Huma-num propose plusieurs outils pour le Ladyss

Un espace ShareDocs : <https://humanid.huma-num.fr/>

Un espace de stockage et de travail collaboratif.

Il est possible de stocker des données personnelles.

Quelles données conservées ?

Dans quelle version ?

Dans quel format ?

Pour combien de temps ?



6. Documenter les données

Pourquoi ?

Pour pouvoir travailler pendant le projet

Pour pouvoir les réutiliser plus tard

Pour les diffuser de façon raisonnée (pour savoir quelles données diffusées, à qui ?, quand ? Et sous quel format)

Pour les préserver à plus long terme

Pour qui ?

Pour vous-même dans 4 ou 5 ans

Pour les membres du projet (partenaires internes et externes)

Pour d'autres chercheurs

...

Quel niveau de détail ?



6. Documenter les données

Documents concernant le projet

Résumé

Protocole

Documents administratifs

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure>

Recueil de données

Méthodes

Grilles d'entretien

Questionnaires

Formulaires de consentement

Traitement des données

Fichier « readme »

Codebook (instructions de codage des réponses)

Dictionnaire de données



7. Enjeux juridiques

Cadre général: RGPD

Qu'est-ce que des données à caractère personnel

Ce sont les informations relatives aux personnes physiques qui permettent de les identifier directement ou indirectement. Par exemple : le nom d'une personne, l'adresse IP de son ordinateur, sa voix.

Par exemple: une combinaison d'informations qui permettront de reconnaître cette personne comme genre, profession et commune. ([définition de la CNIL](#))

Qu'est-ce que des données sensibles

C'est une catégorie de données personnelles. Ce sont des informations qui révèlent

- Les opinions politiques
- L'origine raciale et ethnique
- Les convictions religieuses
- L'appartenance syndicale
- Les données génétiques
- Les données biométriques
- Les données de santé
- L'orientation sexuelle ou la vie sexuelle des personnes ([définition de la CNIL](#))

7. Enjeux juridiques et éthiques

Cadre général: RGPD

La commission européenne a souhaité mieux encadrer et protéger le traitement et la circulation des données personnelles et des données sensibles avec le RGPD (règlement européen sur la protection des données) face à l'utilisation abusive des données par les plateformes.

Ce règlement s'applique à tous les pays européens depuis le 25 mai 2018.

Ce droit protège les personnes physiques qui sont propriétaires de leurs données personnelles. Il protège les personnes vulnérables (mineurs, personnes âgées, demandeurs d'asile ou personnes souffrant d'un handicap...).

Ce règlement s'applique à la recherche scientifique. Il s'applique à l'ensemble du cycle de données de la collecte jusqu'à l'archivage.



7. Enjeux juridiques et éthiques

RGPD : Comment ce règlement s'applique à un contexte de recherche ?

Les droits existants sont maintenus et de nouveaux droits sont créés

- Droits d'accès, de rectification, d'opposition, de suppression

Nouveaux droits

- Droit à la portabilité des données ([définition CNIL](#))
- Droit à l'effacement des données ([démarches auprès de la CNIL](#))
- Droit à la limitation des traitements
- En cas de demande d'accès : délai d' 1 mois après la réception de la demande

Responsabilisation des acteurs

Responsables de traitement, co-responsables de traitement et sous traitants (s'il y a lieu) doivent apporter à tout moment la preuve de la conformité au RGPD.



7. Enjeux juridiques et éthiques

Les différents acteurs et leurs rôles

Qui est le/la Délégué-e à la Protection des Données ?

Le Délégué à la Protection des données informe, conseille les responsables de traitement. Il contrôle le respect de la réglementation.

Quel est le rôle des étudiant.e.s ?

Ils/Elles réalisent les démarches pour la conformité à la réglementation en lien avec leur directeur/directrice de mémoire.

7. Enjeux juridiques et éthiques

Vos démarches



Identifiez-vous

Espace Personnel Paris8

Votre compte numérique

Aide

Mon onglet

ACCUEIL

E-p8 Aide

E-p8

L'E-p8 vous permet d'accéder à travers les réseaux à l'ensemble des ressources et services numériques mis à votre disposition, que vous soyez étudiant, enseignant, chercheur et personnel administratif à l'Université Paris 8.

L'E-p8 vous permet, à partir d'un navigateur situé dans n'importe quel lieu (salles multimédia, bornes, bureau, BU, domicile...), de vous connecter aux services mis à votre disposition, grâce à votre nom d'utilisateur et votre mot de passe.

CATALOGUE DES SERVICES NUMÉRIQUES

Etudiant Enseignant BIATOSS

SCOLARITE

- Calendrier universitaire
- Démarches administratives
- Formulaires en téléchargement
- Notes et résultats
- Notes et résultats des années précédentes
- Réinscription administrative
- Inscription pédagogique
- Tests de langues
- Mobilité internationale
- Vie étudiante

OUTILS

- Messagerie RoudCube
- Messagerie et agenda
- Planification de rendez-vous
- Transfert de fichiers
- Assistance technique

COURS

- Plateforme pédagogique

BU EN LIGNE

- Portail de la BU
- Rechercher dans Octo+
- Lire la presse en ligne
- Ebooks de la BU
- Octaviana, bibliothèque numérique P8
- Emprunter / prolonger un document

VIE UNIVERSITAIRE

- Commissions et conseils
- Suivi des usages numériques

INSERTION PROFESSIONNELLE

- Recherche Stages, emplois : Réso8
- Fonds documentaire d'information et d'orientation

IED

- Site internet de l'IED
- Plateforme pédagogique de l'IED (Moodle)

s://cas.univ-paris8.fr/cas/login?service=https://e-p8.univ-paris8.fr/uPortal/f/u6411s16/p/notes_2009-2010.u6411n152/max/render.uP?pC.n

À retenir

Se fixer des règles simples et explicites

Utiliser les outils, les logiciels en connaissant leurs avantages et leurs limites

Documenter ses choix pour ne pas perdre l'information (qui, quoi, pourquoi, comment, pour qui...)

Stocker les données de façon sécurisée

Protéger les données et les personnes propriétaires des données



To Do List Chalkboard de Mufidh Kassabias, Licence by-nd/2.0/



Pour en savoir plus

Guides

ANDRÉ-POYAUD Isabelle, ASTOR Sandrine, BAUDE Olivier, BOUDJAABA Fabrice, BUJAN Gaëlle, COLLIGON Béatrice, DUBOIS Frédéric, KESSOUS Emmanuel, MAUREL Lionel et ROGER Muriel, *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. Guide pour la recherche*, [s.l.] : INSHS-CNRS, juin 2019, 38 p. URL : https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf.

DELPLANQUE Catherine, LAMRINI Nawale, LECLÈRE Fabrice et MAUREL Lionel, *Règlement Général pour la Protection des Données (RGPD): fiches pratiques à destination des chercheurs*, Université Paris Lumières, Université Paris Nanterre, Université Paris 8, 2019. URL : <https://recherche.parisnanterre.fr/guide-rgpd-2019-894839.kjsp>.

Sites web

Intranet CNRS: https://intranet.cnrs.fr/protection_donnees/donnees/Pages/default.aspx

CNIL: <https://www.cnil.fr/fr>

Lois

Règlement européen sur la protection des données: <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>