

GUIDE N°3: LA GESTION DES DONNÉES

Ce guide vous propose quelques pistes pratiques pour aborder la question de la gestion des données, notamment en ce qui concerne les aspects éthiques (protection des répondants), l'organisation des fichiers, la sécurité des données ainsi que leur documentation.

Qu'est-ce que la gestion des données et pourquoi est-ce important?

La gestion des données est un terme générique qui englobe de nombreuses pratiques. Elle couvre tous les aspects de manipulation, d'organisation, de documentation et de valorisation des données ainsi que les pratiques nécessaires à leur réutilisation et leur partage. La gestion des données s'applique à tous les types de données, qu'elles soient quantitatives ou qualitatives (voir guide n°1 pour une définition des données).

Une bonne gestion des données permet entre autre:

- de protéger les répondants
- d'augmenter la qualité et le potentiel de ses données
- de comprendre ses données et ses fichiers sur le long terme
- de pouvoir partager ses données avec d'autres
- d'assurer la survie des données!

Plus que jamais, une bonne gestion des données est au cœur de la recherche scientifique. L'apparition de nouvelles technologies et des capacités de stockage grandissantes ont révolutionné la façon de faire de la recherche et notre rapport aux données.

De nos jours, une bonne gestion des données est devenue un gage de qualité, mais également une précondition pour le partage des données. Un mouvement en particulier, l'« Open data » a prit de l'ampleur sur la scène politique, préconisant l'accès libéralisé aux données financées par des fonds publics. Le partage des données se fait souvent au travers d'archives telles que celle de [FORS](#) (voir guide n°2)



La gestion des données, qu'est-ce que c'est?

Les principales pratiques de gestion des données concernent:

- la sécurité et le stockage des données
- l'organisation des fichiers
- la documentation des fichiers et des données
- le nettoyage et recodage des données
- la transcription des données
- l'anonymisation des données
- le consentement éclairé
- le copyright
- le partage des données

LE CYCLE DE VIE DES DONNÉES

Pour tout projet, nous recommandons l'élaboration d'un plan de gestion des données. Il s'agit d'une planification détaillée des pratiques à mettre en place dès le début du projet et à adapter tout au long des différents stades.



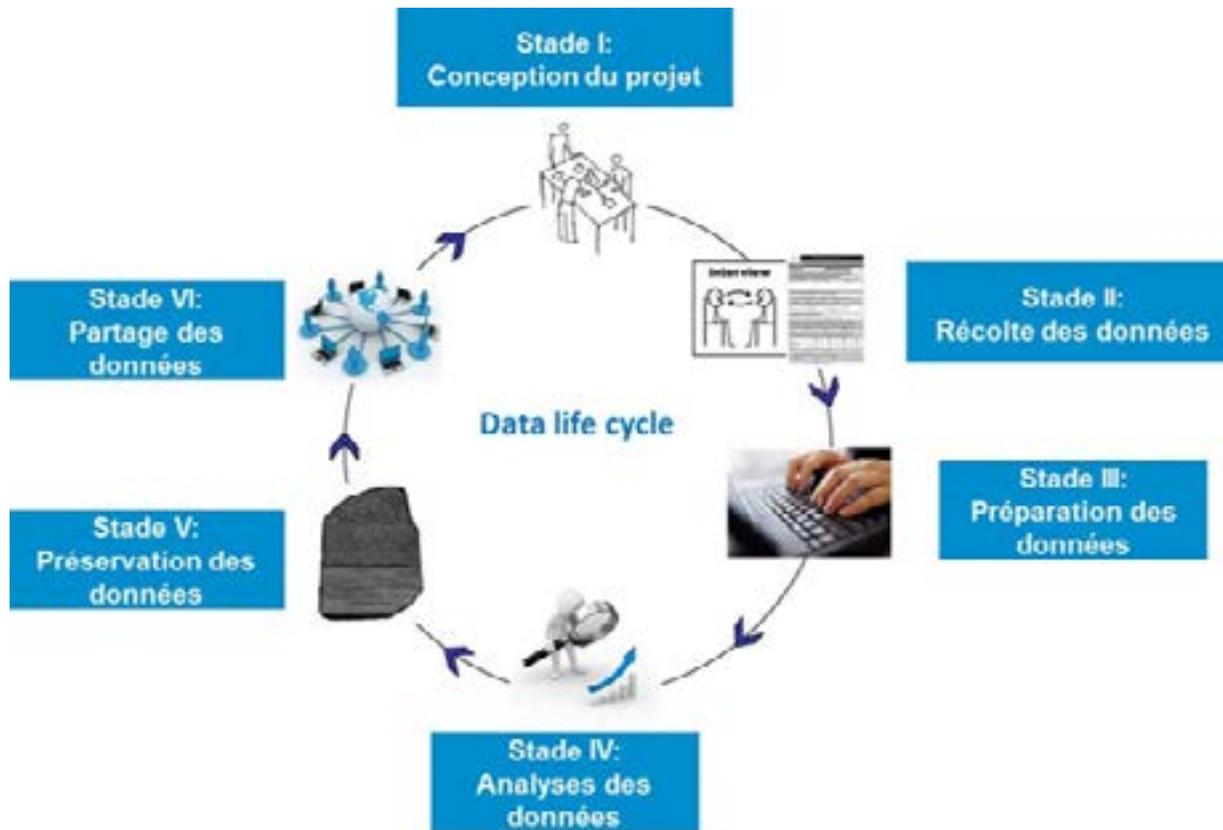
Quelles pratiques à quels stades?

Certaines pratiques s'appliquent **à tous les stades**. Il s'agit des aspects liés aux considérations éthiques, des pratiques qui se rapportent à la sécurité des données ou encore à leur documentation.

D'autres pratiques s'appliquent **à des moments particuliers**. Par exemple, la question du consentement éclairé est à aborder au début du projet.

Dans ce guide, nous proposons de réfléchir aux questions suivantes:

- Comment assurer la protection des répondants?
- Comment organiser les fichiers?
- Comment assurer la sécurité des données?
- Comment documenter mes données et mon projet?



LA GESTION DES DONNÉES EN PRATIQUE

La protection des répondants

La protection des répondants et les considérations éthiques sont au cœur de la recherche. Elles influencent de nombreuses pratiques relatives à différents stades d'une recherche et du traitement des données (la conception du projet, la récolte des données, la préparation des données, l'analyse des données, l'exploitation des données ou encore la conservation, l'archivage et le partage des données).

Selon la question de recherche, on peut être amené à récolter des données personnelles, dites sensibles, qui pourraient permettre l'identification de la personne interrogée voire, en fonction du contenu, lui porter préjudice.

Si vous êtes amenés à récolter des données sensibles, la protection de la personne interrogée et la sécurité des données doivent être une priorité absolue. Il vous faudra alors :

- Dûment informer, notamment dans le cadre d'entretiens qualitatifs, la personne interrogée par rapport au projet et obtenir son consentement pour participer à l'étude mais également par rapport à une éventuelle réutilisation (par des tiers) des données.
- Anonymiser les données, c'est-à-dire enlever ou modifier tous les éléments qui pourraient permettre à eux seuls, ou combinés, de reconnaître une personne. L'anonymisation est de mise pour toutes les données qui seront diffusées ou partagées.
- Adapter les règles de sécurité, de partage et d'accès aux données en fonction de leur sensibilité. Dans le cas de données sensibles, il est important que vous mettiez en place des mesures de protection (voir la section 'sécurité des données').



Les données personnelles et sensibles, qu'est-ce que c'est?

La Loi fédérale sur la protection des données (LPD) du 19 juin 1992 Art.3 définit les **données personnelles** comme « toutes les informations qui se rapportent à une personne identifiée ou identifiable ».

Quant aux données sensibles, elles consistent en des « données personnelles sur :

1. les opinions ou activités religieuses, philosophiques, politiques ou syndicales,
2. la santé, la sphère intime ou l'appartenance à une race,
3. des mesures d'aide sociale,
4. des poursuites ou sanctions pénales et administratives. »

Le principe du consentement

Le consentement est une pratique qui vise à protéger les participants à une étude. Le consentement est souvent exigé pour la récolte de données qualitatives. Pourtant, il peut également être requis pour les données quantitatives, par exemple lors d'une enquête par questionnaire. Il peut être obtenu par voie orale, de préférence enregistré; par écrit, notamment lors d'un entretien par la signature d'un formulaire de consentement; ou par voie électronique, par exemple sous forme d'une case à cocher lors de l'administration d'une enquête en ligne.

Un formulaire de consentement type doit contenir différentes informations. Il doit notamment indiquer:

- Le titre du projet, ainsi que le(s) nom(s) du ou des responsables du projet ainsi que du pourvoyeur de fonds
- Le(s) objectif(s) du projet
- Les raisons pour lesquelles les personnes ont été sélectionnées et la nature de leur participation
- Les avantages et les inconvénients résultant de la participation ainsi que les éventuels risques encourus
- Une information concernant le droit de se retirer du projet, ceci sans préjudice
- Des informations concernant le principe de confidentialité et la protection des données
- Des informations concernant l'éventuel partage des données et les publications prévues



Si le partage des données avec d'autres personnes pendant ou au terme de la recherche est envisagé, l'obtention d'un consentement est indispensable. En cas de doute quant à un éventuel futur usage, mieux vaut inclure la possibilité dans la procédure de consentement: il est très compliqué de recontacter les répondants par après et les données risquent de ne plus pouvoir être partagées avec d'autres.

L'organisation des fichiers

Bien organiser et nommer ses dossiers, fichiers et documents de travail est essentiel pour s'y retrouver. Le nom des documents peut indiquer une brève description du contenu (ex : questionnaire), une date (jj.mm.aaaa), des initiales, les versions (1ère, 2ème,... finale), l'auteur,...

Il est vivement conseillé de :

- Définir une arborescence de dossiers et sous-dossiers en début de projet
- Définir une règle pour nommer les documents en tout début de projet et la noter sous forme de fichier ReadMe (afin de ne pas l'oublier !)
- Appliquer les bonnes pratiques, c'est-à-dire classer les documents dans les dossiers et les nommer de manière systématique en suivant les règles prédéfinies.
- Modifier les règles au besoin et ne pas oublier de mettre à jour le fichier ReadMe.

A vous maintenant de définir vos conventions, de noter les règles et de les appliquer avec rigueur. Cela deviendra vite une habitude et au final, vous gagnerez du temps et éviterez bien des problèmes.



Créer un fichier ReadMe

Un fichier ReadMe est généralement un fichier enregistré sous forme de texte (.txt) qui décrit sommairement certains aspects d'un projet.

Il s'agit d'un fichier contenant des informations importantes concernant des fichiers d'un même répertoire.

Il peut comprendre des instructions d'utilisation, des conventions pour nommer des documents, une liste des noms et des utilités des autres fichiers, des informations sur la personne les ayant créés et la date.

Quelques exemples de conventions pour nommer les documents

- Documents produits par vous-même: Description_Date_Version
- Articles de journaux scientifiques: Auteur_Titre_Année_Journal
- Rapports: Auteur_Description_Date

La sécurité des données

La sécurité est le b.a.-ba de la gestion des données. Sans règles de sécurité, la vie des données est fortement compromise, mais également la protection des répondants dans le cas de données personnelles dites sensibles. Les questions de sécurité comprennent à la fois des questions de stockage et de back-up, mais également des questions de transmission des données et d'accès aux données.

Stockage et back-up

Mal stockées, les données peuvent être perdues, endommagées, voire détruites. Tout d'abord, assurez-vous qu'elles soient sécurisées, c'est-à-dire que des personnes non autorisées ne puissent pas y accéder.

Ensuite, choisissez où déposer vos données. Il peut s'agir d'un serveur institutionnel, d'un ordinateur personnel, d'un serveur en ligne, d'un périphérique externe (DVD, clef USB, mémoire externe, etc.). Chaque outil a des avantages et inconvénients. Privilégiez si possible les serveurs institutionnels car les périphériques externes peuvent être endommagés ou perdus, tout comme les ordinateurs. Les plateformes de stockage en ligne (par ex. google drive, dropbox) sont souvent hébergées à l'étranger et donc régies par d'autres lois de protection des données. Elles ne sont donc pas recommandées.

Aucun outil de stockage n'est totalement sûr et il est primordial de faire régulièrement des sauvegardes et d'avoir sur différents supports plusieurs copies des documents importants.



Choix des plateformes de stockage et de partage des données

Privilégiez les solutions suisses telles que SWITCHdrive et SWITCHfilesender.

Ces infrastructures académiques permettent non seulement le stockage de vos fichiers dans un environnement sécurisé mais aussi le partage de fichiers volumineux et leur synchronisation.

Accès aux données

Si les données récoltées sont personnelles et sensibles, assurez-vous que seules les personnes autorisées peuvent y avoir accès, c'est-à-dire vous-même et éventuellement d'autres membres de l'équipe.

En plus de stocker les données dans un lieu sécurisé, comme un serveur institutionnel, vous pouvez protéger l'accès aux fichiers par un mot de passe ou alors en les encryptant s'ils sont très sensibles.

Les mesures de protection dépendront de la sensibilité des données et des éventuels besoins d'accès aux données par d'autres personnes que vous-même.

Transmission des fichiers

Le partage des fichiers avec d'autres peut se faire de différentes manières, par exemple par e-mail, par la remise en main propre d'une clef USB, en les déposant sur un répertoire commun ou encore par transmission en ligne.

C'est à vous de faire preuve de bon sens, surtout si vous devez transmettre des données sensibles. Comme pour le point précédent, faites très attention avec les clouds et autres serveurs en ligne.

La documentation

Le mot 'documentation' est une notion générale, qui s'applique à divers aspects d'un projet de recherche.

Il est essentiel de 'mettre de côté' des informations utiles, afin de pouvoir les retrouver plus tard pour décrire son travail ou permettre à d'autres de comprendre ce qui a été fait.

La grande difficulté est de définir à l'avance ce dont on aura besoin plus tard et avec quel degré de précision. C'est donc à vous d'anticiper et de constituer un réflexe de documentation au fur et à mesure de l'avancement de votre projet.

La documentation peut se faire au niveau du projet et des données. La documentation peut se faire sous forme de notes, classées, sauvegardées et nommées de façon à les retrouver facilement dans vos fichiers ou alors s'intégrer directement au jeu de données, par exemple les descriptions de variables que vous aurez constitués.

Quelques exemples d'aspects à documenter

Documentation au niveau du projet:

- La genèse du projet: raison du choix du sujet, évolution des idées, grands changements opérés, etc.
- Le développement de l'instrument de recueil des données : difficultés rencontrées, modifications effectuées, éventuels problèmes avec l'instrument utilisé, etc.
- L'échantillonnage: sélection et contact des répondants, difficultés rencontrées, taux de réponses, etc.
- La protection des données: mesures mises en place, règles d'accès aux données.

Au niveau des données:

- Le contexte : comment, quand et où s'est passée la récolte de données (ex: entretiens qualitatifs)
- La préparation et le nettoyage des données: préparation de la base de données, qualité des données, codes utilisés pour les valeurs manquantes, procédures d'anonymisation, règles utilisées pour les transcriptions de données qualitatives.
- Les procédures, les méthodes et les outils d'analyses
- Toutes informations utiles à d'autres pour comprendre vos données et éventuellement les réutiliser

Pour conclure, rappelez-vous qu'une stratégie de gestion des données évolue avec votre projet. Pour être efficace il faut non seulement planifier vos bonnes pratiques au début d'un projet, mais également les réévaluer en cours de route et au besoin les adapter. A vous de jouer !



www.forscenter.ch, dataservice@fors.unil.ch, + 41 21 692 37 17