juin
2021

GUIDE DE BONNES PRATIQUES EN COLLECTE ET GESTION DE DONNÉES MÉDICO-SOCIALES

MÉDECINS DU MONDE

Conception graphique :
Adriana Lyra

En cas de question sur ce guide, merci de contacter le pôle Recherche et Apprentissage à l'adresse suivante : PoleRechercheApprentissages@medecinsdumonde.net
Vous pouvez également consulter la **page intranet dédiée à la gestion et au traitement des données.**

INTRODUCTION ET CONTEXTE

Ce recueil de bonnes pratiques s'adresse aux acteurs de Médecins du Monde ayant à intervenir dans le processus de collecte, gestion ou analyse de données médico-sociales ou de données personnelles sensibles de manière plus générale. Son objectif est de contribuer à l'amélioration de la culture de la gestion des données au sein de l'association, en proposant des solutions concrètes pour répondre aux deux grands enjeux de la gestion de données que sont la qualité et la sécurité.

Ce document est volontairement général afin d'être pertinent pour le plus de cas possibles. Ainsi, dans certains cas la lecture de ce guide ne sera pas suffisante pour résoudre des problèmes opérationnels précis. Il ne faut alors pas hésiter à entrer en contact avec les personnes du siège à même d'apporter des solutions adaptées à chaque situation (pôle Recherche et Apprentissage, Data Protection Officer du Service Juridique, Observatoire France, Service Informatique, référent.e.s Santé, référent.e.s techniques, etc.).

Ce guide/document s'intègre dans le travail en cours sur le projet GDMS – Gestion des données médico-sociales[1]. Ce projet va, sur l'ensemble de nos terrains, nous aider à parler la même langue en ayant les mêmes références sur la gestion des données pour :

1. Assurer le suivi médical et social des personnes au sein des programmes MdM France ;
2. Faciliter et fiabiliser la collecte et la gestion des données pour le suivi/monitoring des projets ;
3. Améliorer la coordination des soins, ici et là-bas.

Ce projet a pour objectif d'harmoniser les processus de gestion et la qualité des données. Il s'inscrit dans le plan de transformation Horizon 2025[2], qui représente l'évolution nécessaire et voulue pour Médecins du Monde France, en se dotant notamment, d'outils performants, intelligents, en phase avec notre temps, pour obtenir des impacts plus larges, mais aussi un pilotage global et transversal.

La qualité des données est primordiale pour Médecins du Monde, car elle contribue à améliorer la qualité de nos interventions et de nos soins auprès des personnes concernées par nos projets, ainsi qu'à porter un plaidoyer fort et impactant pour l'accès aux droits et aux soins. Il est important que les données collectées soient de la meilleure qualité possible pour renforcer la crédibilité de l'association auprès de ses usager.e.s, bailleurs, partenaires, etc., ainsi que pour piloter nos projets (via le monitoring, l'évaluation et la capitalisation) de façon efficace. La qualité des données se mesure formellement sur la base des sept critères suivants :

- **La validité :** les données mesurent et décrivent ce qu'on souhaite mesurer ou décrire.
- **La fiabilité :** les données sont collectées de manière cohérente.
- **La précision :** les données sont suffisamment détaillées pour expliquer les phénomènes étudiés.
- **La complétude :** toutes les données prévues sont collectées.
- **La temporalité :** les données collectées sont représentatives du moment où elles sont collectées.
- **L'intégrité :** les données sont exactes, et protégées de toute manipulation visant à déformer la source originale.
- **L'unicité :** une même donnée/observation ne doit apparaître qu'une seule fois dans la base de données quantitatives.

[1] Pour plus d'informations sur le projet GDMS – données médico-sociales, **voir la page intranet dédiée**.

[2] Pour plus d'informations sur le plan de transformation Horizon 2025, **voir la page intranet dédiée**.

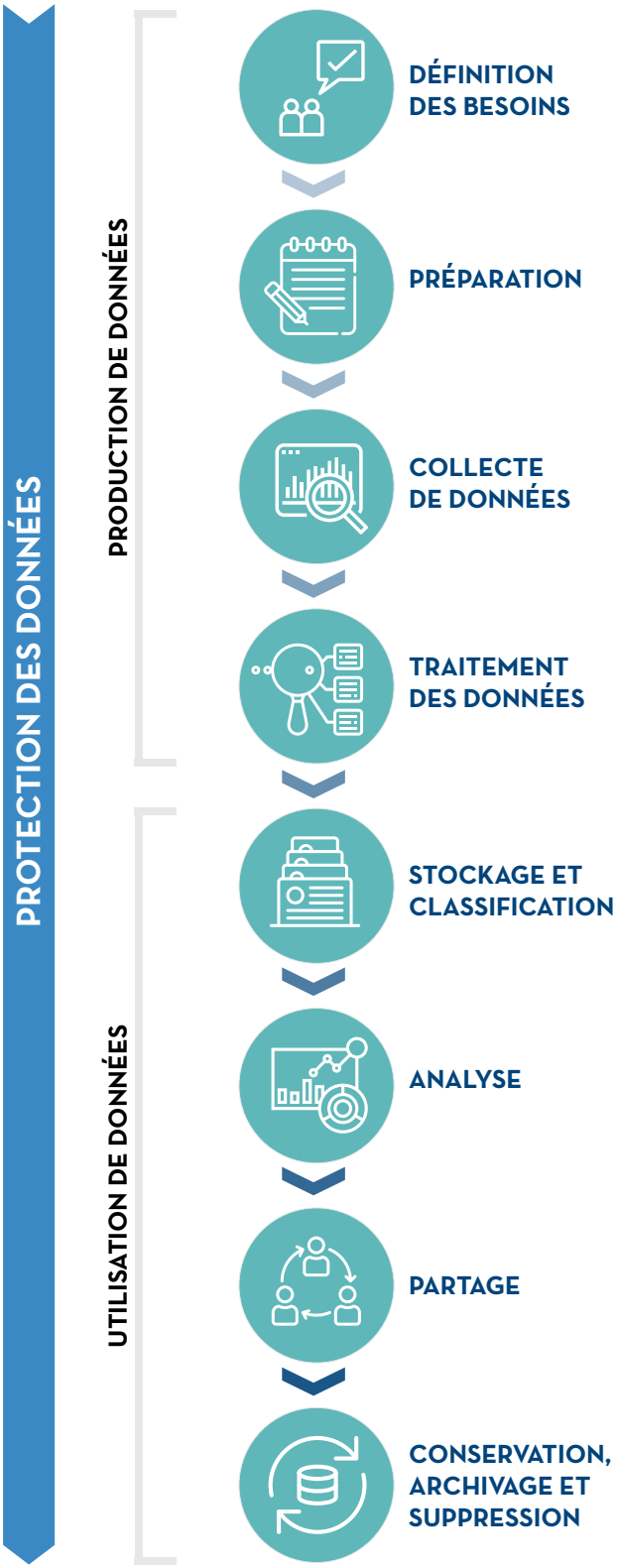
L'autre grand enjeu concerne la sécurité des données. C'est particulièrement important quand des données personnelles sensibles sont collectées et conservées. En effet, la diffusion de données sensibles peut dans certains cas causer d'importants préjudices aux personnes concernées, ce qui va à l'encontre du principe de protection des personnes que nous portons. De plus, il existe un cadre légal très strict encadrant le traitement de ces données (Règlement Général de Protection des Données – RGPD[3]) et ne pas le respecter exposerait Médecins du Monde à des dommages réputationnels ainsi qu'à de fortes sanctions administratives et financières. Les données médico-sociales tombent sous la définition des données sensibles et doivent donc être manipulées avec grande attention. La protection des données est la responsabilité partagée de toutes les personnes intervenant dans leur cycle de gestion. Quels que soient les moyens et les infrastructures mis en place pour sécuriser les données, ce sont les bonnes pratiques des utilisateur.trice.s qui sont leur ultime rempart de protection.

Ce document s'organise de façon à suivre les différentes grandes étapes de la gestion des données, de la définition des besoins à la suppression (voir schéma ci-dessous). Pour chaque étape, des bonnes pratiques visant à atteindre les objectifs de qualité et de sécurité sont énoncées. Ces pratiques sont catégorisées entre celles requises, qui doivent impérativement être respectées, et celles conseillées dont l'application est (fortement) recommandée.

A l'heure actuelle, les outils et pratiques peuvent différer entre les opérations France et International. Bien que la majorité des recommandations se trouvant dans ce guide s'appliquent à tous les terrains d'opérations, nous tenterons d'identifier clairement les parties dont le champ d'application se limite à l'international, ou inversement.

[3] Le RGPD est le texte encadrant le traitement des données personnelles sur le territoire de l'Union européenne. Pour plus d'information, visiter le site de la CNIL.

étapes de la gestion de données et d'informations



Il est à noter que **Médecins du Monde** travaille actuellement au développement d'un nouvel outil interne de gestion des données médico-sociales. Ses objectifs seront, entre autres, d'améliorer la qualité et la sécurité des données collectées, d'harmoniser les pratiques, et de simplifier le travail des acteurs sur le terrain. Ce guide est donc amené à évoluer pour prendre en compte les spécificités de ce nouvel outil une fois qu'il sera disponible[4].

Ressources utiles :

- Guide planification de projets
- Guide de gestion de données personnelles sensibles : Pour une éthique de terrain
- Guide collectes de données : méthodes quantitatives
- Guide collectes de données : méthodes qualitatives
- Note de clarification du concept de redevabilité à Médecins du Monde
- Guide pratique des études de satisfaction
- Fiche pratique de gestion des données VLG en situation de crise
- Charte éthique de la recherche

[4] Pour plus d'informations sur le projet GDMS – données médico-sociales, **voir la page intranet dédiée**.

SOMMAIRE

PHASE DE PLANIFICATION

PHASE DE PLANIFICATION

Pratiques requises

Définition des besoins

Ne doivent être collectées que les données ayant un objectif **précis** et **spécifié**.

Les raisons à cela sont d'ordre pratiques. En effet, collecter des données est un processus qui peut être long, difficile, et coûteux en ressources humaines et matérielles, mieux vaut donc se concentrer sur ce qui est réellement utile en évitant de se disperser en recueillant des informations inutiles. Elles sont également légales. En effet, le principe de finalité du RGPD (voir annexe) indique que les données ne peuvent être collectées et utilisées que dans un but « précis », interdisant ainsi la collecte de données si leur but n'a pas été spécifiquement défini. De même, le principe de proportionnalité et de pertinence (ou principe de minimisation) insiste sur la nécessité de ne collecter uniquement les données strictement nécessaires à l'atteinte d'un objectif.

Dans ce but, dans le cadre d'une recherche, il est utile de rédiger un **plan d'analyse** en amont de la collecte des données. Celui-ci a pour but de décrire toutes les analyses que l'on souhaite réaliser sur les données collectées. Cela peut donc permettre de mieux se rendre compte de quelles données sont réellement nécessaires pour répondre aux questions posées.

Dans le cadre du monitoring projet, il est essentiel d'élaborer le **plan de monitoring (résumé des indicateurs et le data flow)** en amont de la collecte des données. Celui-ci a pour but de décrire précisément les indicateurs, leur utilisation au cours du projet, mais aussi d'identifier les données nécessaires au calcul de ces indicateurs et l'organisation de la collecte et de la saisie de ces données. Cela permet d'identifier et de collecter uniquement les données nécessaires pour le projet.

En ce qui concerne les données personnelles identifiantes (nom, prénom, adresse, email, téléphone, numéro de sécurité sociale, etc.), il est recommandé de recueillir le moins d'information possible. Seul ce qui est strictement nécessaire doit être collecté, par exemple pour recontacter la personne en cas de suivi patient. Le premier objectif de la collecte de ces données est de contribuer à la qualité du service offert.

Ressources utiles :

- E-learning Monitoring, voir module 2.1. Elaborer des indicateurs spécifiques avec le résumé des indicateurs.
- Guide collectes de données : méthodes quantitatives. Page 38
- Guide collectes de données : méthodes qualitatives.
- Guide de gestion de données personnelles sensibles : Pour une éthique de terrain

Prendre en compte les principes d'éthique

Il est important de considérer les grands principes d'éthique tels que le **respect des personnes**[6], le **principe de bénéficence**[7] (ou générosité) ou le **principe de justice**[8] au moment de la planification d'une collecte de données. Il s'agira ainsi de s'assurer que les individu.e.s sont informé.e.s de leurs **droits** et en capacité de les faire valoir, que la collecte de données

[6] Respect de l'autonomie et de l'autodétermination des participant.e.s et à la protection de ceux/celles qui ne sont pas autonomes, notamment en leur offrant un abri contre les dangers ou les sévices

[7] Souvent mentionné en tant que « do no harm », il signifie que la décision de collecter les données doit se faire après s'être interrogé.e sur les risques encourus par les sujets humains contribuant à la collecte de données.

Les effets négatifs subis par les sujets peuvent être très variés, et d'amplitude variable. Des exemples sont (liste non-exhaustive) :

- Le désagrément de devoir passer du temps à répondre à un questionnaire,
- Le manque à gagner financier d'avoir passé du temps à répondre à un questionnaire,
- L'impact psychologique ou émotionnel de devoir répondre à des questions au sujets d'épisodes traumatiques,
- Le risque d'être discriminé voire de subir des violences dans le cas où des informations sensibles venaient à fuiter.

ne les expose pas à des **risques** inutiles, ou encore que l'on n'**exclut aucun groupe de population**. Si des risques liés à la collecte de données sont identifiés, un suivi de leur évolution devra être mis en place, ainsi qu'une stratégie pour les limiter un maximum. Pour cela, il est recommandé de remplir un tableau d'évaluation des risques liés aux données (voir modèle dans les ressources ci-dessous).

Selon les cas (le type de données collectées, leur finalité, etc.), il peut être obligatoire de faire valider la collecte par un comité d'éthique national. Il est important de se renseigner sur l'existence d'un comité d'éthique dans le pays en question, et sur ses modalités de fonctionnement. Cela est à prendre en compte au moment de planifier le calendrier et le budget. Dans le cadre d'une recherche opérationnelle, il est également indispensable de se référer à la charte éthique de la recherche de MdM.

Ressources utiles :

- Guide de gestion de données personnelles sensibles : Pour une éthique de terrain
- Charte éthique de la recherche à Médecins du Monde
- Modèle tableau évaluation des risques liés au données

Il est évident que le risque zéro n'est pas atteignable, mais la responsabilité de l'équipe en charge de la collecte de données est de minimiser les risques encourus par les sujets. De plus, il est important de se poser la question de savoir si les bénéfices de cette collecte de données excèdent les risques encourus. Si ce n'est pas le cas, la collecte de donnée devrait être remise en cause.

[8] Il s'agit de s'assurer que les données collectées sont représentatives de tous les groupes de la population afin de ne pas exclure qui que ce soit des bénéfices potentiels, et de répartir équitablement bénéfices et charges. On parle de « leave no one behind ».

[9] Voir annexe

Recueillir le consentement

Le consentement libre et éclairé des individus est obligatoire pour la collecte et le traitement de données personnelles sensibles.

Pour être valide, le consentement des individus doit être :

- **Libre :** C'est-à-dire qu'il ne doit pas être influencé ou contraint. Par exemple, les personnes ne doivent pas risquer de subir des conséquences négatives (telles que l'exclusion d'un projet ou le non-accès à une forme d'aide) en cas de refus, ou se voir promettre un avantage en cas d'acceptation.
- **Spécifique :** Le consentement ne donne pas l'autorisation à Médecins du Monde d'utiliser les données collectées librement. Il est nécessaire d'indiquer les traitements (analyse, stockage, transfert, archivage) des données collectées, ainsi que leurs finalités, afin que la personne sache à quoi elle consent.
- **Eclairé :** La personne doit avoir accès à toutes les informations nécessaires pour prendre sa décision en toute connaissance de cause. Cela inclut l'identité des personnes en charge du traitement et de leurs responsabilités respectives, les finalités poursuivies, et les droits de la personne (droit de refuser de répondre ou de sortir de l'étude à tout moment, droit de consultation, à l'effacement, de modification[9]). Ces informations doivent être exprimées dans un langage compréhensible par la personne. En outre, il faut indiquer à la personne un moyen de contacter une personne responsable en cas de question.
- **Univoque :** Il faut que le consentement soit clairement et expressément donné par la personne. Il ne doit pas être ambigu. Par exemple, dans le cas d'un formulaire à remplir, la case du consentement ne doit pas être pré-cochée afin que la personne consente « activement ».

Si la personne interrogée est incapable de consentir elle-même (mineur.e non émancipé.e, personne vivant avec un handicap psychique ou mental, personne en situation d'enfermement) il faudra obtenir le consentement formel de son/sa tuteur.trice légal.e en plus du sien. Dans certains cas, le consentement d'une structure ayant autorité peut remplacer celui du tuteur légal.

Ressources utiles :

- Guide de gestion de données personnelles sensibles : Pour une éthique de terrain
- Charte éthique de la recherche à Médecins du Monde

Pratiques recommandées

Planifier en amont les mesures de protection de la vie privée

Afin d'anticiper les risques plutôt que de les subir, il est important de mettre en place des mesures de protection des données personnelles et sensibles en amont de leur collecte.

Ces mesures doivent être préventives plutôt que correctives, c'est-à-dire qu'il faut anticiper les failles de sécurité (formulaires papier perdus/volés, données interceptées par email, accédées sans autorisation sur des plateformes de partage, récupérées sur appareils perdus/volés, etc.) et tout faire pour les éviter. L'utilisation du tableau d'évaluation des risques est recommandée pour leur identification.

Les mesures peuvent être, par exemple, la conservation des formulaires dans des armoires fermées à clef, le chiffrage systématique des fichiers contenant des données sensibles, la séparation des données identifiantes directement des données d'analyse, la restriction de l'accès aux dossiers contenant des données sensibles à un minimum de personnes, etc.

Les mesures doivent également, dans la mesure du possible, être systématiques (automatisées par ordinateur, ou inscrites régulièrement dans l'agenda de membres de l'équipe) dans le but d'éviter les « oublis » de la part de l'équipe en charge de la gestion des données. Toutes les étapes de la gestion des données doivent être concernées par ces mesures, de la collecte à la suppression en passant par le stockage, le nettoyage, l'analyse, le partage ou l'archivage. Il est cependant possible que des problèmes non-anticipés apparaissent au cours du cycle de gestion des données. Il est donc recommandé de mettre en place un système fluide de communication permettant de signaler toute alerte concernant la protection des données vers les personnes responsables, afin que celles-ci puissent rapidement appliquer des mesures correctives. En cas de fuite avérée des données, il est de la responsabilité de Médecins du Monde d'en informer la CNIL et/ou les autorités locales compétentes sous 72h, ainsi que les personnes concernées si cette fuite engendre un risque pour leur sécurité ou leurs droits.

Toutes ces mesures seront idéalement compilées et conservées dans un plan de gestion de données, qui sert de feuille de route de la gestion des données. Dans le cas où les données collectées sont particulièrement sensibles et/ou que le risque est très élevé, il faut considérer d'établir un « Privacy impact assessment » (analyse d'impact relative à la protection des données) et de le soumettre à la Commission Nationale de l'Informatique et des Libertés (CNIL). Contacter la DPO de Médecins du Monde en cas de doute.

Ressources utiles :

- Modèle tableau évaluation des risques liés au données

Etablir un système de codage d'identifiants des usager.e.s pertinent

Lors de la première rencontre avec une personne dont nous souhaitons collecter les informations, il faut lui attribuer un **code** qui permettra de l'identifier dans une base de données anonymisée, et son suivi dans le temps. Ce code doit donc être **unique** (deux usager.e.s ne peuvent pas avoir le même code), **anonyme**, et idéalement être facile à retenir pour l'usager.e. En effet, le code sera toujours inclus dans les bases de données contenant des informations sensibles au sujet des usager.e.s, il ne faut donc pas que ce code contienne des informations permettant l'identification des personnes.

Ce code permet de faire le lien entre la **base de données d'analyse** (contenant toutes les informations utiles au suivi des usager.e.s, des activités, etc., mais ne permettant pas d'identifier directement un individu) **et le registre de correspondance** (contenant le nom des personnes, et les autres informations identifiantes telles que le numéro de téléphone ou l'adresse email). Ce registre de correspondance est particulièrement sensible et doit donc être protégé en conséquence. Seules une ou deux personnes doivent être autorisées à le consulter, les accès doivent être répertoriés, et le fichier doit impérativement être chiffré (voir page 23).

Choisir l'outil de collecte adapté

Avant de se lancer dans le choix et/ou la mise en place d'un outil de collecte, il convient de garder en tête que le projet GDMS (gestion des données médico-sociales)[10] est au travail en ce moment pour répondre à nos besoins de collecte et d'analyse de données, notamment certification HDS et que le futur outil interne de gestion des données médico-sociales deviendra la norme une fois disponible, et a donc vocation à remplacer tous les outils actuellement utilisés[11]. Il n'est donc pas recommandé de mettre en place de nouveaux outils ad-hoc à ce stade, sauf impératif. Un choix éventuel doit se faire parmi les solutions existantes actuellement

Pour schématiser, il existe deux types d'outils : les formulaires dits « papier » ou **les outils de collecte de données mobiles** (formulaires « numériques ») (tels que KoboToolBox, ODK, SurveyCTO, Sphinx, etc.). **En France**, seules le Dossier Patient Informatisé (DPI) et Sphinx peuvent être utilisés pour des collectes de données personnelles.

Les formulaires papiers présentent l'avantage d'être faciles à créer et ne requièrent aucune compétence technique de la part des utilisateurs.trice.s. Ils sont toutefois plus susceptibles d'entrainer des erreurs, et impliquent un temps conséquent de saisie des données dans une base de données informatique (Excel, Monitool). Il y a également le risque de perdre les données en cas de perte des formulaires.

La collecte de données mobile nécessite un peu plus de connaissances techniques de la part de l'équipe pour développer et administrer l'outil dans la durée (attention à la perte de compétences en cas de turnover dans l'équipe), ainsi que des utilisateur.trice.s (savoir utiliser une tablette ou un smartphone est suffisant). Toutefois, ils permettent la création d'outils de collecte plus efficaces et assurant un contrôle plus strict de la qualité des données. Ils permettent également d'éviter de s'encombrer avec de grandes quantités de papier, de réduire les risques de perte, de limiter les risques d'erreur dans la saisie, d'exporter les données directement sur Excel, voire de fournir un environnement de stockage sécurisé et de permettre de réaliser directement les analyses souhaitées.

Avant d'entamer une collecte de données, il faut prendre le temps de peser le pour et le contre de chaque type d'outil afin de faire le meilleur choix possible. Les paramètres à prendre en compte sont

[10] Pour plus d'informations sur le projet GDMS – données médico-sociales, **voir la page intranet dédiée.**

[11] Ce travail a déjà démarré avec un certain nombre de vos collègues, et peut-être vous-même.

la quantité et le type de données, leur finalité, la pérennité de l'outil utilisé[12], la qualité et la sécurité des données, le coût, le temps, les compétences de l'équipe, etc.

Dans le cadre d'une collecte de données de routine dans des établissements sanitaires, il faut absolument prioriser l'utilisation du **Système National d'Information Sanitaire** (SNIS) s'il existe, plutôt que de développer et mettre en place un outil de collectes de données parallèle.

Être attentif à la bonne conception de l'outil de collecte

Un mauvais outil produira invariablement des mauvaises données[13]. La bonne conception du formulaire, ainsi que le choix de questions pertinentes, sont déterminants à la collecte de données de qualité. Les pratiques varient selon l'outil utilisé et le type de données collectées, mais les pratiques suivantes sont de bons réflexes à adopter dans le cas d'un formulaire servant à la collecte de données individuelles :

- Pour les données individuelles, toujours commencer par demander le consentement éclairé de l'individu
- Si des données permettant l'identification directe sont collectées, les isoler afin de pouvoir aisément les séparer du reste des données (Formulaire papier : les renseigner sur une feuille à part pouvant être détachée, et liable au reste via le code d'identification. Formulaire numérique : inclure dans les noms de variables une indication sur leurs caractères identifiants)
- Limiter si possible l'utilisation des questions ouvertes, qui requièrent plus de temps et de ressources pour être exploitées
- Rédiger les questions entièrement et explicitement. Les collecteur.trice.s de données sont censé.e.s les lire au mot près, ils/elles ne doivent pas les interpréter ou les reformuler
- Enoncer les questions dans un langage intelligible pour les individus, et en évitant de créer des biais.
- Numéroter les questions
- Coder les modalités de réponse de façon consistante. Par exemple, toujours utiliser la même échelle de satisfaction (e.g. 1. Très satisfait, 2. Satisfait, 3. Partiellement satisfait, 4. Insatisfait, 5. Très insatisfait) dans un formulaire.
- Si le formulaire est numérique, utiliser des contraintes pour les réponses (sans en abuser). Par exemple, empêcher la saisie d'un âge supérieur à 120 ou négatif.
- Utiliser des liens logiques et sauts de questions. Par exemple, ne pas poser des questions sur les enfants si la personne a déclaré ne pas avoir d'enfant précédemment.
- Si le formulaire est numérique, donner des noms de variables explicites.

Vous pouvez trouver plus de conseils pour le développement de questionnaires quantitatifs de qualité sur cette page (externe, en anglais).

Tester l'outil de collecte avant son utilisation

Avant de déployer un outil sur le terrain pour collecter les données, il est recommandé de le **tester autant que possible**. Cela évite de rencontrer des problèmes durant la collecte, qui nécessiterait des corrections dans l'urgence.

Il est ainsi conseillé que le formulaire soit testé dans un premier temps par la personne en charge de sa conception. L'objectif de ce premier test est de vérifier que toutes les questions sont bien là, que les contraintes et liens logiques fonctionnent (le cas échéant), qu'il n'y a aucun « bug », etc.

[12] Sur une collecte de données dans le cadre du suivi d'un service (consultation de n'importe quel type), il faut s'assurer que les outils choisis seront durables et continueront de pouvoir être appliqués après départ de MdM

[13] Référence au GIGO (garbage in, garbage out)

Ensuite, un membre de l'équipe du projet devrait à son tour tester l'outil. L'apport d'un regard neuf permet souvent de détecter des erreurs invisibles pour la personne ayant programmé l'outil. Ce test permet d'identifier les erreurs manifestes, les enchainements de questions illogiques, les contraintes abusives, etc.

Dans le cadre particulier d'une enquête, il est également nécessaire de tester le formulaire en conditions réelles. Il s'agit de demander à une personne qui fera effectivement partie de l'équipe de collecte de le tester sur quelques personnes présentant les mêmes caractéristiques que la population cible. L'objectif est de déceler si l'outil est facilement utilisable par les collecteurs de données, et si les questions sont bien comprises et acceptées par les répondants. S'il s'agit d'une collecte de données qualitatives, ces tests permettront de déterminer si les grilles d'entretiens/d'observations sont claires pour les collecteurs de données, et si les informations qu'elles produisent sont à la fois fidèles à la source d'origine et exploitables à l'analyse.

Le feedback collecté durant ces différentes phases de test doit être utilisé pour ajuster l'outil avant son déploiement général.

Un cas particulier est celui de la collecte de données de routine utilisant le SNIS. Il est grandement recommandé de toujours utiliser ce système, même si l'outil semble inefficace, afin d'obtenir des données harmonisées avec les autres organisations sur place. L'outil ne doit donc pas être ajusté.

Ressources utiles :

- Guide collectes de données : méthodes quantitatives. Page 28.
- Guide pratique des études de satisfaction

Prendre le temps de former les collecteurs.rice.s de données

La qualité des données collectées dépend en grande partie des personnes en charge de leur collecte. Qu'il s'agisse de membres de l'équipe MdM en charge de l'implémentation des activités du projet en question, ou de personnel externe contractés pour la collecte, la formation ne doit surtout pas être négligée.

La formation sert à :

- Familiariser les collecteur.trice.s de données avec le sujet du projet
- Familiariser les collecteur.trice.s de données avec l'outil et le formulaire. A la fin d'une formation, les collecteur.trice.s de données doivent être très à l'aise avec le formulaire de collecte. Ils/elles ne doivent douter du sens d'aucun mot ou question.
- Sensibiliser les collecteur.trice.s de données aux principes d'éthique (droits de la personne, consentement éclairé...)
- Enseigner les bonnes attitudes à adopter face aux individus (respect de la personne, neutralité, attitudes face aux sujets sensibles...)

De plus, il est nécessaire de faire signer un **accord de confidentialité** aux collecteur.trice.s de données avant le lancement d'une collecte de données personnelles et/ou sensibles.

Dans le cadre d'une collecte de données continue (pour du monitoring par exemple), il est important que la formation et l'accompagnement des collecteur.trice.s de données s'inscrive dans la durée afin de garder la qualité des données constante tout au long du projet.

Ressources utiles :

- Guide collectes de données : méthodes quantitatives. Page 46.
- Guide collecte de données : méthodes qualitatives.

Liste des documents à produire durant la phase de planification

- **Formulaire de collecte :** Outil permettant la saisie primaire des données. Nécessaire pour collecter les données quantitatives.
- **Grilles d'entretiens et d'observations (si collecte qualitative) :** Outil permettant de saisir les données qualitatives.
- **Le dictionnaire de données :** De préférence sur Excel, c'est une base de données contenant le nom de chaque variable collectée, leur type, la liste des catégories de réponses, et les contraintes associées. Utile pour le nettoyage et l'analyse des données.
- **Tableau d'évaluation des risques liés au données :** Compilation des risques de sécurité et de confidentialité résultant de la collecte de données. Utile pour pouvoir mieux les anticiper et donc les prévenir.
- **Système de monitoring (si monitoring) Cela inclue :** le cadre logique, les rôles et responsabilités, le résumé des indicateurs, la cartographie des sources de données et le dataflow chart. Voir le e-learning dédié au monitoring et le manuel de monitoring pour plus de détails.
- **Plan de gestion des données :** Document répertoriant les processus à respecter tout au long du cycle de gestion des données. Il contient les procédures à respecter pour le traitement, le stockage, le partage, l'archivage et la suppression des données. Il sert à s'assurer que les bonnes pratiques en matière de gestion de données seront respectées tout au long du projet.
- **Plan d'analyse :** Document détaillant les analyses à réaliser sur les données collectées. Comme le data flow, il sert à s'assurer que suffisamment de données sont collectées pour répondre aux questions posées, et que chaque donnée collectée a une utilité. Sert aussi à anticiper les besoins en données afin de collecter tout ce dont on a besoin.

PHASE DE COLLECTE DE DONNÉES

PHASE DE COLLECTE DES DONNÉES

Pratiques requises

Protéger les outils de collecte

Protéger les données passe par la protection des outils de collecte. Ces derniers pouvant héberger temporairement ou de façon continue des données, leur perte, le vol, ou le piratage constituent des risques majeurs de fuite de données. Pour s'en prémunir, de bons réflexes sont à adopter, tels que :

- Garder les outils de collecte (formulaires papiers ou outil numérique) **sous surveillance** à tout instant. Lorsqu'ils ne sont pas utilisés, les conserver dans une pièce/un meuble **fermé à clef**.
- Sécuriser les outils de collecte de données numériques avec **un mot de passe solide** (voir ressource ci-dessous). Les paramétrer pour qu'ils se verrouillent automatiquement après quelques minutes d'inactivité.
- **Ne pas conserver de mots de passe à l'écrit** (post-it, cahier, etc.) ou sur un emplacement visible de l'ordinateur (sur le bureau), où ils pourraient être vus par quelqu'un d'autre
- **Ne pas installer d'applications inutiles** ou venant de sources non fiables sur les outils numériques.
- **Désactiver le Wi-Fi et le Bluetooth** des appareils numériques si ce n'est pas nécessaire. Eviter d'utiliser les WI-FI publics tant que possible.
- **Ne jamais utiliser de clef USB/disque dur externe/CD non sécurisés** sur les appareils numériques. Ne branchez pas vos téléphones/tablettes à des ordinateurs inconnus. Ils pourraient contenir des virus et infecter les appareils numériques.
- S'assurer que le système d'exploitation et l'antivirus des appareils numériques sont constamment à jour.
- Sauvegarder les données collectées sur la carte SIM/carte mémoire plutôt que sur l'appareil directement. Si possible, conservez ces cartes séparées de l'appareil lorsque vous ne l'utilisez pas afin de limiter le risque de vol.
- Activer le dispositif de localisation GPS des appareils numériques afin de pouvoir les retrouver en cas de perte ou de vol .

Ressources utiles :

- Guide création d'un mot de passe sécurisé.
- Best IT practices guideline

Mener les entretiens individuels dans des conditions adaptées

Dans le cadre d'une collecte de données individuelles auprès des personnes, il est important de s'assurer qu'aucune tierce personne ne peut voir ou entendre les informations divulguées au cours de l'entretien. Des informations sensibles pourraient en effet fuiter de cette manière. De plus, si une personne n'est pas à l'aise pour discuter des sujets sensibles discutés durant l'entretien, elle risque de donner des informations inexactes.

La confidentialité parfaite n'étant pas toujours atteignable, essayer de privilégier un lieu d'entretien choisi par la personne, afin qu'elle s'y sente à l'aise. De même, l'identité du collecteur.trice de données peut également affecter la qualité des données. En effet, une personne ne sera pas forcément à l'aise pour répondre à certaines questions si elles sont posées par quelqu'un du sexe opposé, d'une ethnie différente, beaucoup plus jeune/âgé, etc.

[14] Ne doit en aucun cas être utilisé pour géolocaliser le personnel de Médecins du Monde.

Pratiques recommandées

Tester la qualité des données au fur et à mesure de la collecte

La qualité des données étant un enjeu primordial, plusieurs étapes au cours du cycle de gestion des données servent à l'assurer. C'est ainsi le rôle de la bonne programmation de l'outil de collecte et de la formation des collecteur.trice.s de données durant la phase de planification. Durant la collecte, il est recommandé de régulièrement tester la qualité des données afin de pouvoir corriger les problèmes décelés (outil de collecte défaillant, mauvaise compréhension des questions, mauvaise qualité de travail d'un ou plusieurs collecteur.trice.s de données, etc.).

Pour cela, il faut que les données soient saisies régulièrement afin de pouvoir les scruter. Ces analyses rapides peuvent permettre de détecter des erreurs dans les données (valeurs aberrantes[15], illogismes[16], etc.). Les erreurs potentielles détectées peuvent alors être remontées aux collecteur.trice.s de données pour correction. Ces analyses peuvent également aider à identifier un collecteur de données dont le travail est de mauvaise qualité, par exemple si une forte proportion de ses formulaires contient des valeurs manquantes. Il faudra alors réfléchir à reformer le/la collecteur.trice de données en question.

Il est également possible pour le/la responsable de la collecte de réaliser des visites aléatoires des terrains de collecte pour vérifier que les outils de collecte sont bien utilisés par les collecteur.trice.s, que les personnes comprennent les questions correctement, etc.

Il est aussi recommandé de régulièrement se demander si les 6 critères de qualité des données sont respectés. Prenons l'exemple d'une collecte de données auprès d'utilisateur.rice.s d'une structure de santé.

- **La validité :** Les données mesurent ce que l'on attend, et sont dans le format attendu. Par exemple, si l'on cherche à étudier les comportements des travailleurs.euses du sexe entre 18 et 30 ans, seules les personnes correspondant à cette définition doivent être considérées. De plus, les dates de naissance sont dans le format « Date » plutôt que « Texte ».
- **La fiabilité :** Les données collectées ne sont pas affectées par le contexte de la collecte. Par exemple, une femme peut avoir plus de mal à donner des informations sur sa santé sexuelle à un homme qu'à une autre femme. Il faut donc s'assurer que les personnes sont dans des conditions favorable lors de la collecte des données.
- **La précision :** Les données sont suffisamment détaillées. Par exemple, plutôt que de simplement indiquer que des préservatifs ont été distribués, on préfèrera souvent connaître le nombre exact de préservatifs distribués à chaque utilisateur.trice. Également, il s'agit de s'assurer que les données représentent fidèlement la réalité. Pour cela, traquer les valeurs aberrantes est très utile.
- **La temporalité :** Les données remontent-elles assez rapidement et fréquemment pour éclairer les prises de décision ? S'assurer de toujours bien saisir la date à laquelle chaque donnée est collectée.
- **L'intégrité :** Les données ne doivent pas être manipulées, volontairement ou pas. Il ne faut jamais supposer des informations que des personnes ne communiquent pas. Par exemple, il ne faut pas supposer qu'une femme a été victime de violences si elle ne le dit pas, ou à l'inverse supposer qu'une déclaration de violences par un.e patient.e n'est pas valide.

[15] Des valeurs aberrantes sont des données si éloignées des valeurs attendues qu'on suppose qu'une erreur de saisie a pu avoir lieu. Par exemple, un individu âgé de 70 ans dans une collecte de données chez des jeunes usagers de drogue, ou un individu avec plus de 15 enfants. Il n'est pas toujours possible de s'apercevoir des erreurs, mais des valeurs ou des écarts surprenants, aberrants ou incohérents doivent attirer l'attention. »

[16] Incohérences entre plusieurs variables. Par exemple, un homme enceinte, ou un enfant plus âgé que ses parents.

- **L'unicité :** Il ne faut jamais saisir la même information plusieurs fois, et ne pas entrer les mêmes données d'une personne à plusieurs reprises. Si possible, préférer la saisie dans une base de données unique et centralisée. S'assurer de l'absence de doublons (cependant, la même personne peut dans certains cas être observées plusieurs fois, par exemple si elle utilise un service plus d'une fois).

Ressources utiles :

- Guide nettoyage de données sur Excel

PHASE DE NETTOYAGE ET D'ANALYSE DES DONNÉES

PHASE DE NETTOYAGE ET D'ANALYSE DES DONNÉES

Pratiques requises

Créer un masque de saisie si les données sont retranscrites sur Excel

Dans le cas des formulaires papiers devant être saisi sur Excel (dans le but de pouvoir être analysées, agrégées en vue de renseigner le monitool, etc.), un **masque de saisie** doit être préparé en amont. Cela permet souvent de **gagner du temps** pendant la saisie, mais également de **limiter les erreurs** et de gagner en **cohérence** entre les observations. Les principes suivants sont également à respecter :

- La première ligne contient les noms de variables, qui doivent être explicites. Entrer un seul tableau de données par feuille Excel
- Entrer une seule information par cellule
- Ne pas fusionner de cellules

Respecter ces principes permet de faciliter l'analyse des données, et le portage des données d'Excel vers d'autres outils.

Dans le cadre d'une enquête, la **double saisie** est également recommandée[17] . Cela consiste à demander à deux personnes de saisir les mêmes données, pour ensuite comparer leurs entrées. En cas de différence, on pourra retourner consulter le formulaire papier pour retrouver la valeur correcte. Cela permet de limiter le risque d'erreur de saisie.

Ressources utiles :

- Guide création de masque de saisie sur Excel

Isoler les données directement identifiantes

Les données permettant d'identifier un individu directement (nom, numéro de téléphone, adresse mail, numéro de sécurité sociale, etc.) doivent être conservés dans une base de données séparée (le **registre de correspondance**) du reste (**base de données d'analyse anonymisée**). La base de données d'analyse ne doit être liable au registre de correspondance que via le code unique attribué à chaque personne.

Aucune donnée personnelle identifiante ne doit subsister dans la base de données d'analyse. Le registre de correspondance ne doit être accessible que par un nombre réduit d'utilisateurs.

Nettoyer les données

Le nettoyage est la dernière étape de l'assurance qualité des données. Il consiste en la recherche d'erreurs manifestes ou suspectées dans les données. Bien que la démarche soit similaire, la différence avec l'étape du test de qualité durant la collecte est qu'il est souvent trop tard pour contacter les collecteurs de données en cas d'erreur suspectée. Il peut donc être nécessaire de faire des choix pour résoudre les valeurs aberrantes ou extrêmes qui risquent de fausser l'analyse. Les choix peuvent être de conserver les données telle quelle, de les remplacer par une valeur plus plausible (à justifier !), ou de supprimer la donnée.

[17] Pas toujours possible car demande beaucoup de temps et de ressources

Durant le nettoyage, il est très important de :

- **Ne jamais écraser la base de données brutes (données primaires)**. Sauvegarder la base nettoyée sous un nom différent, et garder la base originelle intacte
- **Documenter** les changements effectués sur les données.

Ressources utiles :

- Guide nettoyage de données sur Excel

Analyser les données

L'analyse consiste à transformer les données en **informations utiles à la prise de décisions**, et qui permettront de **rendre des comptes** à toutes les parties prenantes du projet. Les informations ainsi produites pourront également être présentées aux populations étudiées, afin de mettre en avant l'importance des collectes de données et conserver leur engagement.

Ne réaliser que les analyses prévues en amont dans le plan d'analyse (cf. p. XX). Les analyses doivent répondre à des besoins spécifiés à l'avance, tels que suivre l'évolution des indicateurs de monitoring, alimenter le reporting, le plaidoyer, la veille, le suivi des patient.e.s/activités, etc. Il s'agit de ne pas perdre de temps à réaliser des analyses sans valeur-ajoutée. Toutes les données collectées doivent être analysées, et donc les données ne doivent pas être collectées si aucune analyse n'est prévue.

Concrètement, une bonne analyse inclut les étapes suivantes :

- **Formulation d'hypothèses :** Celles-ci doivent être définies au moment de la définition du plan d'analyse, et décrire ce que l'on va chercher à démontrer à l'aide de ces données
- **Analyse descriptive systématique des indicateurs :** Calculer des statistiques telles que la moyenne, la médiane, la variance, le maximum/minimum, etc. des différentes variables présentes dans un jeu de données est une bonne base pour la compréhension des informations collectées.
- **Extraction des informations pertinentes et partage avec l'équipe :** La personne en charge de l'analyse doit présenter les informations pertinentes de façon qu'elles soient compréhensibles et accessibles au reste de l'équipe, afin qu'elles puissent être utilisées à la prise de décisions. Pour cela, il est recommandé d'utiliser un maximum de **représentations visuelles** (graphiques, tableaux, cartes) des données, afin de rendre les données plus engageantes pour les personnes qui y auront accès. Il faut s'assurer qu'aucune donnée identifiante n'est visible dans les visualisations.
- **Interprétation des données :** Voir des chiffres est une chose, comprendre ce qu'ils signifient par rapport à une situation réelle en est une autre. Il est recommandé d'impliquer toutes les parties prenantes à un projet dans l'interprétation des données qui en découlent, afin d'éviter les conclusions erronées.

L'analyse est une étape primordiale du cycle de vie des données, car elle permet de de transformer les données en information permettant de les utiliser et les valoriser. Un guide entier pourrait être consacré à l'analyse de données, mais ce n'est pas l'objet de ce document. Pour d'avantage d'informations à ce sujet, il est possible de se référer aux chapitres 3.3.D et 3.3.E du guide de planification de projets.

Ressources utiles :

- Guide planification de projets.
- E-learning Monitoring, voir module 2.3. Utiliser les résultats du monitoring.

STOCKAGE ET MAINTENANCE DES DONNÉES

STOCKAGE ET MAINTENANCE DES DONNÉES

Notez bien : cette phase n'intervient pas après la phase de nettoyage et d'analyse chronologiquement. Il faut penser au stockage et à la protection tout au long du cycle de vie des données - de l'amont de leur collecte à leur suppression.

Pratiques requises

Stocker les données sensibles de façon sécurisée

Les données sensible « papier », même une fois anonymisées, doivent absolument être conservées **sous clef**, et leur accès doit être limité à un nombre restreint de personnes.

En ce qui concerne les données « informatisées » (fichier Excel, Word, enregistrement audio, image numérisée, etc.), il vaut mieux éviter de les conserver uniquement en « local », c'est-à-dire sauvegardées directement sur un ordinateur ou disque dur. En effet, les données risqueraient d'être perdues en cas de panne, de piratage ou d'obsolescence du matériel. De plus, le stockage limite la possibilité de travailler collaborativement sur les données – d'autant qu'il est fortement déconseillé de partager des données sensibles par clef USB ou par email.

Médecins du Monde travaille activement au développement d'un outil de collecte et de gestion des données médicosociales[18], qui aura vocation à remplacer tous les outils actuellement en place et harmoniser nos pratiques. Il permettra notamment de stocker toutes les données de façon sécurisée (l'outil sera certifié Hébergeur de Données de Santé), et permettra de travailler collaborativement facilement et sans risque.

Pour les missions à l'international et en attendant le déploiement de cette solution, la seule plateforme de stockage à la disposition des missions est SharePoint (en France, on recommande fortement l'utilisation de l'outil Sphinx et/ou du Dossier Patient Informatisé en attendant le nouvel outil). Cette plateforme n'offrant pas un niveau de sécurité optimal pour les données y étant sauvegardées, **il est de la responsabilité de tou.te.s d'adopter les bonnes pratiques décrites dans ce guide pour réduire le risque de voir les données sensibles des usager.e.s des projets fuiter**. Cela implique de limiter à un minimum le nombre d'informations personnelles identifiantes ou sensibles collectées, de chiffrer tous les fichiers contenant des données sensibles, de séparer les données identifiantes des données sensibles, et de limiter l'accès aux dossiers contenant des données sensibles aux seuls membres de l'équipe projet autorisés.

En outre, certains pays interdisent que des données personnelles concernant leurs citoyen.ne.s soient stockées en dehors de leurs frontières. Lorsque c'est le cas, il n'est pas possible d'utiliser de plateformes de stockage en ligne en ligne tels que SharePoint. Il faut alors privilégier des solutions de stockage en local.

Chiffrer les fichiers contenant des données sensibles

Chiffrer des fichiers les rend **illisibles** pour quiconque ne possédant pas le mot de passe et/ou le fichier clef associés. Les logiciel gratuits AxCrypt et 7zip, utilisés par Médecins du Monde, emploient la méthode de chiffrage AES-256, qui offre un niveau de sécurité optimal. Le chiffrage des fichiers contenant des données sensibles est donc une manière très efficace d'éviter les fuites en cas de perte ou de vol de matériel informatique où ils sont sauvegardés, ou en cas de piratage des fichiers. Pour pouvoir ouvrir et utiliser un fichier chiffré, il suffit de le déchiffrer en local (sur un ordinateur). Une fois utilisé, il doit impérativement être chiffré à nouveau.

Les fichiers contenant des données sensibles doivent impérativement être chiffrés avant d'être sauvegardées sur des plateformes en ligne ou envoyées par email (en général l'envoi par email est fortement déconseillé).

Attention : Une fois chiffrés, les fichiers ne peuvent en aucun cas être déverrouillés si le mot de passe est oublié et/ou le fichier clef est perdu. Il est donc primordial de consciencieusement conserver les détails de déchiffrage. Il est recommandé que les mots de passe soient toujours connus par au moins 2 personnes afin de limiter le risque d'oubli (mais ne pas diffuser le mot de passe outre mesure).

Ressources utiles :

- Guide pas-à-pas de chiffrage des fichiers avec AxCrypt

Conserver séparément les données directement identifiantes

Comme indiqué précédemment, les données d'analyse doivent être conservées séparément des données personnelles directement identifiantes (registre de correspondance). De préférence, les deux fichiers doivent être conservés dans des dossiers séparés, et l'accès au registre de correspondance doit être restreint à un minimum de personnes.

Toutefois, l'identification des données doit rester possible afin de permettre aux personnes de requérir la consultation, la modification ou la suppression de leurs données personnelles.

Transférer les données par des moyens sécurisés

Le partage des données est une phase risquée du processus de gestion des données, où les données sensibles sont les plus susceptibles de fuiter. C'est pourquoi il est nécessaire d'apporter une grande attention aux moyens utilisés pour partager les données entre utilisateur.trice.s, entre sites, etc.

D'une manière générale toutes les données, et d'autant plus si elles sont sensibles, ne doivent être accessible qu'à un minimum de personnes au sein de Médecins du Monde, et ne sont en principe pas partagées en dehors de l'association. Toutefois il arrive que ce genre de partage soit nécessaire, dans ce cas il faut s'assurer de ne partager que ce qui est absolument utile. Selon les besoins de la personne avec qui les données sont partagées, il est possible de les anonymiser (enlever toutes les données personnelles identifiantes, c'est-à-dire n'envoyer que la base de données d'analyse) ou de les agréger (par exemple, si un partenaire souhaite connaitre le nombre de femmes participant à un projet, calculer et envoyer ce nombre plutôt que d'envoyer la liste des noms des femmes en question).

Un bon moyen de partage est l'utilisation de plateformes de stockage en ligne (type SharePoint). Il suffit ensuite de donner accès à la personne souhaitée pour qu'elle puisse accéder aux données (puis de supprimer cet accès une fois qu'il n'est plus nécessaire).

Envoyer les données par email n'est en principe pas une solution sécurisée et est donc fortement déconseillé. Si l'envoi de fichiers sensibles par mail est nécessaire, il faut absolument prendre soin de chiffrer à la fois le mail, et les fichiers individuellement. Les fichiers clefs associés doivent être envoyés dans un email séparé, chiffré également. Il est préférable de communiquer les mots de passe de vive voix, lors d'un appel. Il est par contre interdit d'envoyer des données via des réseaux sociaux ou des applications de messagerie instantanées (Facebook, WhatsApp, Messenger, etc.).

Bien que **fortement déconseillé**, il reste possible de partager les données physiquement, via une clef USB, un CD, etc. Il est toutefois nécessaire de transmettre les données en mains propres, en utilisant un périphérique appartenant à Médecins du Monde, sécurisé, et dédié à cet effet. Ne jamais confier un périphérique contenant des données à une personne tiers, et ne jamais le poster.

Ressources utiles :

- Fiche gestion des données VLG

Archiver et supprimer les données une fois leur utilisation terminée

La durée de conservation des données personnelles est encadrée légalement. Parce que MdM est une ONG de droit français, la législation française qui doit être respectée **quel que soit le terrain d'intervention**. Une fois leur durée d'utilisation terminée, **les données doivent être supprimées ou archivées**. C'est-à-dire qu'elles doivent être séparées des bases données actives, et conservées dans un dossier dédié dont l'accès est restreint.

Les dossiers médicaux ou patients doivent être conservés pour une période bien déterminée. Dans le cas des personnes adultes ou des enfants âgés de 8 ans ou plus, les dossiers doivent être conservés pendant 20 ans. Pour les enfants âgés de moins de 8 ans, le dossier doit être conservé jusqu'à la date de leur 28ème anniversaire. Ils doivent en outre être conservés dans des conditions permettant d'assurer leur confidentialité et leur intégrité, qu'il s'agisse de documents papiers ou de fichiers informatiques.

Dans le cas de données individuelles qui ne sont pas des dossiers médicaux, **la règle générale est de ne pas les conserver une fois leur utilisation** (dans le cadre d'un monitoring, d'une recherche, etc.) terminée. Toutefois, dans le cas où ces données pourraient être à nouveau utile dans le futur, elles peuvent être archivées et donc conservées plus longtemps pour leur valeur statistique une fois qu'une **anonymisation complète**[19] a été effectuée. Quoi qu'il en soit, la durée de conservation des données personnelles ainsi que leurs traitements (partage, anonymisation) **doivent être indiqués aux personnes** avant la collecte, au moment du consentement.

La première étape consiste à déterminer quelles données doivent être archivées, comment et combien de temps. Il est important de formaliser le cycle de vie des documents et d'évaluer l'intérêt de leur conservation à titre intermédiaire ou définitif. Il est donc utile de construire un référentiel ou un tableau de gestion des archives afin de toujours savoir quelles données sont actuellement archivées, depuis quand, pour combien de temps, et quel sera leur sort final. Ce tableau doit être régulièrement mis à jour, et les nouveaux utilisateurs des données doivent en prendre connaissance dès leur prise de poste.

A noter, il n'est pas suffisant de placer des documents dans la corbeille de l'ordinateur pour les supprimer. Il faut immédiatement vider la corbeille, voire utiliser des logiciels de nettoyage de fichiers (type CCleaner) dans le cas de données sensibles. De plus, il est nécessaire de s'assurer que toutes les copies des données ont été correctement supprimées (version locale, en ligne, backup sur disque amovible, etc.).

[19] Une anonymisation complète consiste en faire en sorte qu'il soit impossible d'identifier une personne en se basant sur les données détenues. Il ne suffit pas de supprimer les informations directement identifiantes (nom, adresse, numéro de sécurité sociale, etc.), il faut également penser au risque d'identification indirecte. Par exemple, savoir d'une personne son genre, son nombre d'enfants, le village dans lequel elle vit, et sa profession peut suffire pour identifier une personne – si elle est la seule dans cette situation précise.

Ressources utiles :

- Guide pratique : Les durées de conservation. (CNIL)
- Référentiel : Les durées de conservation. Recherches dans le domaine de la santé. (CNIL)
- Référentiel : Les durées de conservation. Traitements dans le domaine de la santé (hors recherche). (CNIL)

Double-sauvegarder (backup) les données aussi souvent que possible

Il est recommandé de conserver une copie des données collectées. C'est particulièrement critique dans le cas où les données sont sauvegardées en local (sur un ordinateur, disque dur, etc.), car une panne du matériel signifierait la perte définitive des données si celles-ci n'ont pas été backup.

Les données sont de préférence backup sur des appareils externes sécurisés (type clef USB ou disque dur externe) conservés sous clef, ou à défaut sur une plateforme autorisée.

Il est recommandé de backup les données aussi souvent que possible, pour réduire la quantité d'information perdue en cas de perte des fichiers.

Maintenir un contrôle des autorisations d'utilisation

Créer différents niveaux d'autorisations pour les utilisateur.trice.s de vos données informatisées (ou contrôler l'accès aux données papier). Cela peut être fait quelle que soit la solution retenue pour stocker vos données. En général, selon la solution de stockage des données, il est recommandé de différencier :

- Les comptes de consultation, permettant seulement à l'usager de consulter des fichiers
- Les comptes de modification, permettant à l'usager de consulter, modifier des données, ou en saisir de nouvelles
- Les comptes administrateurs, donnant à l'usager les mêmes autorisations qu'un compte de modification, et lui permettant de changer les rôles des autres utilisateurs.

Il est possible de donner des rôles différents au même usager selon le fichier en question. Une fois qu'un membre de l'équipe quitte le projet ou n'a plus besoin des données, son accès doit lui être retiré.

Ce contrôle est primordial, car le trop large accès aux données augmente le risque de fuite (accidentelle, par piratage, par partage externe non-autorisé, etc.). Ces différents niveaux d'autorisation doivent donc être définis avant même la mise en place d'une collecte de données.

ANNEXE

ANNEXE

Définitions

Base de données brutes : Une base de données brutes est une base de données qui n'a été soumise à aucun traitement ou manipulation.

Base de données d'analyse anonymisée : La base d'analyse anonymisée contient toutes les données collectées sur les personnes, sauf celles permettant leur identification directe. Ces dernières sont conservées dans le registre de correspondance. Les deux bases de données peuvent être liées via le code d'identification des personnes. La base d'analyse contient toutes les informations utiles à l'analyse pour le monitoring, la recherche, etc.

Culture de la donnée : Ensemble de connaissances et pratiques partagées par tous les membres d'une organisation vis-à-vis des données. Son but est de sensibiliser chacun à l'intérêt de collecter et d'analyser des données, à l'importance d'employer de bonnes pratiques, et de partager des outils communs permettant à chacun de d'utiliser au mieux les données disponibles.

Donnée : Une donnée est définie comme ce qui est admis, connu ou reconnu, et qui sert de base à un raisonnement. Il s'agit donc d'un élément brut, non interprété et non contextualisé. Les données pouvant être de nature très différente : qualitative, quantitative, structurées, non structurées, et de sources différentes.

Données personnelles : Les données personnelles sont celles permettant l'identification, directe ou indirecte, de personnes physiques. Par exemple le nom, l'adresse email, le numéro de téléphone, l'adresse postale, le numéro de sécurité sociale, ou la photo d'identité sont des données personnelles. Ces données peuvent être explicites, ou devinable par le recoupage d'informations. De plus, il faut se méfier des verbatims portant des indices qui permettraient la reconnaissance des personnes.

Données sensibles : Les données sensibles sont une catégorie spéciale des données personnelles. Ce sont des données à caractère personnel dont le traitement est comporte un risque de préjudice pour les personnes physiques. Les préjudices possibles sont, par exemple, la discrimination, l'embarras, ou le vol d'identité. Sont considérées comme sensibles les données concernant :

- L'origine ethnique ou raciale ;
- Les convictions religieuses ou philosophiques ;
- Le traitement des données génétiques ;
- Les opinions politiques ;
- L'appartenance syndicale ;
- La santé ;
- La vie sexuelle ou l'orientation sexuelle ;
- Les données biométriques qui permettent l'identification d'une personne ;
- Les données sur les infractions ou condamnations

Observation : Une observation est une instance d'une base de données. Dans une base de données sur des individus, chaque colonne représente une variable (nom, prénom, âge, profession, sexe, etc.), et chaque ligne représente une observation (individu 1, individu 2, individu 3, etc.)

Registre de correspondance : Base de données contenant toutes les données permettant l'identification directe d'une personne (nom, prénom, adresse, numéro de téléphone, numéro de sécurité sociale, etc.), ainsi que le code d'identification. Doit être conservée séparément des autres données, potentiellement sensibles, concernant les individus. Ainsi, si les données sensibles venaient à fuiter, les personnes concernées ne pourraient pas être identifiées. Ce registre doit être de manière très sécurisée, et accessible seulement à quelques personnes. Il sera utilisé pour réidentifier les personnes si besoin (par exemple dans le cas où une personne demande à consulter/modifier ses données).

Variable : Une variable est une donnée brute observée ou mesurée sur les différents individus d'une population, et qui est susceptible de changer d'un individu à l'autre, ou pour le même individu au cours du temps. Les variables permettent de calculer les indicateurs.

Principes du RGPD

Les grands principes du RGPD à prendre en considération sont les suivants :

- Le principe de finalité : ne peuvent être collectées et utilisées que les données personnelles ayant un but précis, légal et légitime
- Le principe de proportionnalité et pertinence (minimisation) : ne peuvent être collectées que des données pertinentes et nécessaires vis-à-vis de la finalité annoncée
- Le principe de conservation limitée : les données personnelles ne peuvent être conservées que pendant une période limitée, définie à l'avance
- Le principe de sécurité et de confidentialité : le responsable des données est garant de leur sécurité et confidentialité
- Les droits des personnes : les droits des personnes vis-à-vis de leurs données, tel que les droits de consultation, de modification ou de suppression, doivent être respectés.

Médecins du Monde étant une organisation de droit français, les principes du RGPD sont à respecter sur tous ses terrains d'interventions, en Europe ou ailleurs, s'il n'entre pas en conflit avec la règlementation locale. Pour plus d'information sur la protection des données ou en cas de question, voir la page du Service Juridique - Data Protection Officer sur l'Intranet de Médecins du Monde, ou contactez la Data Protection Officer directement.

Droits des personnes

Droit de consultation

Les personnes ont le droit de requérir qu'une copie de leurs données personnelles détenues par Médecin du Monde leur soit transmise pour consultation.

Droit de rectification

Les personnes peuvent demander la rectification de données à leur sujet si celles-ci sont inexactes ou incomplètes. Pour cela, la personne doit fournir une preuve de son identité, ainsi que la preuve que les données étaient effectivement inexactes.

Droit d'effacement

Les personnes, sous certaines conditions, demander la suppression de leurs données personnelles. La suppression peut être demandée si les données ne sont plus utiles au projet, si la personne se rétracte sur son consentement, si le traitement ne correspond pas à ce qui avait été indiqué lors du consentement, si elle était mineure au moment de donner son consentement.

Toutefois, Médecins du Monde peut légalement refuser d'accepter de supprimer des données si cela à l'encontre d'une obligation légale (par exemple, de conserver un dossier patient au moins 20 ans), si les données ont un intérêt de santé publique, ou si elles sont utilisées à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou à des fins statistiques.

Notez bien : Il est très important de documenter toutes les demandes de consultation, de rectification et d'effacement des données. En cas de doute sur la procédure à suivre, contacter la DPO.


MÉDECINS
DU MONDE