CHECK-LIST DES PRINCIPAUX ÉLÉMENTS À PRENDRE EN COMPTE AVANT DE SÉLECTIONNER UN NOUVEL OUTIL NUMÉRIQUE DE COLLECTE DE DONNÉES DE MANIÈRE RESPONSABLE

La crise sanitaire a poussé de très nombreuses organisations à **modifier leurs pratiques de collecte de données**. Elles ont notamment eu **recours à de nouveaux outils** qu'elles n'utilisaient pas auparavant ou ont été contraintes de **recourir à des outils qu'elles n'utilisaient qu'à petite échelle**. Par exemple, certaines organisations se sont mises à collecter via des logiciels de type CATI (*Computer Assisted Telephone Interview*) des informations sensibles telles que des alertes de violences sexuelles et basées sur le genre (SGBV) qu'elles ne collectaient auparavant que sur papier. D'autres ont, quant à elles, généralisé l'utilisation de smartphones à toutes leurs activités de collecte qui se limitaient jusqu'alors au suivi des infrastructures.

L'introduction de nouveaux outils ou la généralisation de ceux-ci pose nécessairement des **questions de protection de données à partir du moment où les données collectées sont personnelles et/ou sensibles**. Or, il s'avère bien souvent que les organisations n'ont pas les ressources (temps, humaines, etc.) voire les compétences pour **analyser des outils sous l'angle de la question de la protection des données**.

Cette liste de contrôle vise donc à fournir quelques éléments de vulgarisation et de sensibilisation pour faciliter la prise de décision lors du choix d'un nouvel outil de collecte de données dans une optique de protection des données.

La question de la protection des données demeure complexe et cette check-list n'a que pour objectif de rendre un peu plus "accessible" ces questions. **Elle ne peut dans aucun cas remplacer une analyse légale détaillée, ni un DPIA complet** (Data Protection Impact Assessment) que les organisations ont dans de nombreux cas l'obligation légale et éthique de conduire avant d'introduire un nouvel outil de traitement des données et/ou avant de lancer une nouvelle activité de collecte.

Cette check-list n'intègre pas non plus les questions d'usage des solutions, aussi diverses qu'essentielles : Quelle base légale pour le traitement des données (ex. consentement éclairé) ? Comment évaluer au préalable le niveau de sensibilité des données collectées ? Comment assurer une collecte minimale et proportionnelle aux besoins d'information ? Quelles procédures internes pour la collecte (conception ou adaptation) existe-t-il ? Comment former des équipes aux principes de la gestion responsable des données ? Comment configurer et sécuriser des terminaux mobiles en tant que tel (chiffrement, code PIN, etc.) ? et ainsi de suite.

- Pour plus d'informations sur ces questions, merci de vous référer aux diverses ressources existantes et spécialisées propres au secteur de la solidarité internationale, telles que le "Manuel sur la protection des données dans l'action humanitaire" du CICR, les Guidance Notes du Centre for Humanitarian Data, les retours d'expériences d'OXFAM, etc.

Enfin **cette check-list se concentre sur des outils génériques destinés** par exemple **aux enquêtes ménages**. Elle ne traite pas des questions complexes d'application de "suivi des contacts" (ou *contact tracing* en anglais) et autres applications médicales dédiées.

CHECK-LIST

Pour garder la check-list la plus simple possible le **mot générique "outil" sera utilisé ci-dessous pour désigner une solution numérique de collecte de données**. Ce terme recouvre néanmoins des situations très différentes qui vont de la solution informatique open source installable soi-même sur son propre serveur - tel que ODK Central - au système propriétaire proposé en SaaS (*Software as a service*) par un prestataire indépendant - tel que SurveyCTO proposé par Dobility - en passant par des solutions mise à disposition gratuitement par des acteurs du secteur comme KoBoToolbox par OCHA. Chacune de ces situations recouvrent néanmoins des réalités informatiques et légales différentes - lesquelles demandent une analyse spécifique - qui ne seront que très brièvement évoquées ci-dessous pour garder une check-list accessible.

1 L'outil offre-t-il de prime abord des garanties/engagements sur son approche et sur les mesures qu'il met en œuvre pour garantir la protection, confidentialité et sécurisation des données qu'il collecte et stocke ?

- Est-il explicitement mentionné que l'outil peut être utilisé pour des données personnelles et/ou sensibles ?
- Une politique de confidentialité (ou des conditions générales) claire et compréhensible [et non pas rédigée dans un vocabulaire légal complexe] est-elle disponible ? Et/ou une page résumant les mesures mises en œuvre en termes de sécurité est-elle disponible et mise à jour régulièrement ?
- La propriété des données collectées et stockées est-elle bien garantie dans les documents contractuels ? Autrement dit l'éventuel prestataire s'interdit-il bien une réutilisation des données pour toute autre fin (ciblage publicitaire, analyse de comportement etc.) ou des doutes existent quant à cet aspect ?
- L'outil n'est-il pas connu pour avoir déjà subi des fuites de données ou pour avoir eu recours à des pratiques non compatibles avec le secteur ? Est-il connu pour figurer sur la liste noire d'autres organisations ?

2 L'outil offre-t-il (notamment en cas de mise à disposition en mode SaaS ou gratuites) les garanties légales nécessaires ?

Les garanties légales doivent - généralement - être **à la fois compatibles avec les législations des sièges des organisations** (RGPD pour une organisation basée en Europe par exemple) **et avec celles des pays de collecte de données** (Data Protection Act pour des données collectées au Kenya par exemple) ; certaines législations peuvent néanmoins êtres incompatibles entre elles.

- Les conditions générales précisent-elles des éléments "de base" garantissant leur conformité avec la législation tels que (i) l'existence d'un contact DPO ou poste similaire (ii) la garantie d'une durée maximale d'information en cas de fuite des données etc. ?

- La localisation des serveurs est-elle conforme aux exigences légales ? Par exemple obligation d'un serveur basé en Europe ou dans un pays validé comme adéquat dans le cadre du RGPD.
- Les dispositions contractuelles sont-elles conformes à la législation ? Bien souvent des DPA (*Data Processing Agreement*) doivent être signées spécifiquement pour être en conformité.
- Ou, l'outil peut-il être hébergé sur les propres serveurs de l'organisation ?

3 L'outil permet-il de gérer facilement les utilisateurs devant contribuer ou simplement accéder aux données ?

- Est-il possible de créer et d'utiliser des comptes individuels pour accéder à l'outil [plutôt que des comptes génériques demandant le partage de mots de passe entre plusieurs utilisateurs] ?
- Un administrateur peut-il disposer facilement d'une vue d'ensemble de tous les utilisateurs et de leurs droits respectifs ?
- Peut-il facilement gérer ceux-ci (les supprimer ou les désactiver, etc.) ?
- Pour des questions de facilité, l'outil permet-il d'utiliser le même système de contrôle des utilisateurs et d'authentification que celui déjà utilisé dans l'organisation (système de type SSO - *Single sign-on* - pour AD/LDAP) ?
- L'outil permet-il de configurer des accès suffisamment granulaires aux données ? Des droits différents existent-ils (lecture, modification, suppression etc.) par utilisateur ? Ces droits sont-ils configurables pour chaque collecte de données ou type de collecte de données (par activité de collecte et/ou zone géographique et/ou équipe et/ou champ) ?

4 L'outil permet-il d'identifier facilement les activités de collecte présentant des risques ?

- L'outil permet-il de catégoriser manuellement une activité de collecte comme contenant des données personnelles non sensibles, des données sensibles, des données démographiques identifiables (DII), etc. ?
- L'outil offre-t-il la possibilité d'identifier automatiquement des données comme personnelles (via par exemple l'existence d'un champ nom, adresse, etc.) ?
- L'outil permet-il, pour une même activité de collecte, d'avoir des accès différents par type de champs (des champs de données pseudonymisés accessibles à certains utilisateurs et des champs de données personnelles uniquement accessibles à d'autres) ?

5 L'outil contribue-t-il de par sa conception à mettre en œuvre des bonnes pratiques de gestion des données ?

- L'outil permet-il une collecte adaptée du consentement en fonction des contextes (enregistrement audio par exemple, enregistrement de signatures, impression automatique d'un reçu et des modalités de contact ultérieur, etc.) ?

- L’outil permet-il une suppression manuelle aisée des données tant au niveau de l’activité de collecte qu’au niveau de chacun des enregistrements (suppression de masse possible) ? L’outil permet-il une suppression automatique des données stockées en local sur les terminaux ?

- L’outil facilite-t-il le suivi des périodes de rétention des données : est-il par exemple possible d'affecter à chaque activité de collecte une date d’expiration avec rappel automatique pour suppression des données ? Une fonction d’archivage (avec des accès limités) est-elle disponible ?

- L’outil permet-il de facilement pseudonymiser des données ?

- L’outil avertit-il l’utilisateur lorsque celui-ci réalise des actions “risquées” telles que l’export ou le partage public des données ?

- L’outil permet-il d’exercer facilement les demandes spécifiques d’une personne concernée au sujet de ses données (accès, suppression, arrêt du traitement etc.) en facilitant la recherche à travers certains champs (nom, adresse) ?

- Le chiffrement ("*encryption*" en anglais) de certaines activités ou champ de données est-il facile à mettre en œuvre, de telle manière qu’il ne décourage pas l’utilisateur ? Les exports de données sont-ils également automatiquement chiffrés ?

- L’outil aide-t-il à anonymiser automatiquement certains champs tels que le floutage des visages en cas de collecte de photo, la dé-identification des coordonnées GPS, etc. ?

6 L’outil contribue-t-il de par sa conception à faciliter le travail de l’administrateur de l’outil ?

- L'administrateur a-t-il bien accès à un “*log*” informatique centralisé (ou *audit trail*) permettant d’identifier les actions de chaque utilisateur (consultation, téléchargement, modification, etc.) et d’identifier les problèmes en cas d’éventuelle fuite de données ?

- L’administrateur est-il alerté en cas de comportement suspect d’un utilisateur (téléchargement de large quantité de données, tentatives d’accès infructueuses etc.) ?

- L’administrateur est-il aidé dans sa gestion des utilisateurs (désactivation automatique par exemple d’utilisateurs inactifs pendant un certain temps, possibilité de transfert des droits facilitée en cas de congés ou *turn over* RH) ?

- L’administrateur est-il aidé dans sa gestion des activités de collecte (identification des activités devant être supprimées car expirées, liste des activités étant publiquement accessibles, etc.) ?

7 L’outil est-il suffisamment sécurisé pour héberger des données personnelles ?

- Un chiffrement des données est-il possible de bout en bout tant au niveau du serveur (*at rest*), que pendant le transfert des données (*in transit*) ou au niveau des terminaux (mobiles ou navigateurs) ?

- Le type de chiffrement utilisé est-il connu et satisfaisant ?

- La clé de chiffrement est-elle bien détenue uniquement par l’organisation et non accessible par le prestataire ?

- Les serveurs du prestataire sont-ils bien certifiés (ISO, NIST, HDS, HIPAA, etc.) et disposent-ils d'un suivi adapté ?
- Une évaluation externe des mesures de sécurité déployées par l'outil et/ou le prestataire est-elle disponible (rapport d'audit sécuritaire de l'outil, outil open source dont le code a été revu par une communauté de développeurs, etc.) ?
- Des mises à jour de l'outil sont-elles faites régulièrement ?
- Les mesures d'authentification sont-elles compatibles avec celles de l'organisation (mot de passe fort imposé, déconnexion automatique après un laps de temps, authentification à deux facteurs pour certains utilisateurs, accès possible à travers un VPN, etc.) ?
- Les serveurs stockant les données sont-ils bien automatiquement répliqués/sauvegardés à une fréquence adaptée ? Le prestataire dispose-t-il d'un plan de continuité des activités ? Le niveau de service (SLA - *Service Level Agreement*) tel que le temps de disponibilité des serveurs ou le temps pour combler une vulnérabilité sont-ils satisfaisants ?

RESSOURCES DISPONIBLES EN LIGNE

1 **Ressources détaillant des principes génériques au sujet de la protection des données pour la collecte de données sur mobile :**

- **Data Security with Mobile Forms** *(en anglais)*

2 **Ressources contenant quelques éléments de comparaison entre différents outils au sujet de la protection des données :**

- **How to Choose a Mobile Data Collection Platform** *(en anglais)*
- **Benchmarking de solutions de collecte de données sur mobile** (version 2017 en anglais, version 2021 en français)
- **What electronic tools are appropriate to meet the needs of outpatient programs of Médecins Sans Frontières?** *(en anglais)*

Merci de votre lecture !

Les icônes utilisées dans ce document ont été créées par DinosoftLabs, Eucalyp, Freepik, monkik et Smashicons, disponibles sur www.flaticon.com.

Avec la participation de MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES Liberté Égalité Fraternité

AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

H2H Network Humanitarian Action Support

UKaid from the British people

La version originale de cette publication a bénéficié du soutien de l'Agence Française de Développement (AFD) et du H2H Fund du H2H Network, ce dernier étant soutenu par UK aid du gouvernement britannique. Sa traduction en français a pu être réalisée grâce à la participation du Centre de crise et de soutien du Ministère de l'Europe et des Affaires étrangères (CDCS). Néanmoins, les idées et opinions présentées dans ce document ne représentent pas nécessairement celles du H2H Network, de UK aid, du CDCS et de l'AFD.