

PLAN DE GESTION DE DONNÉES (PGD)



Décrire, analyser et comprendre les effets de l'introduction de l'autodépistage du VIH en Afrique de l'Ouest à travers l'exemple du programme ATLAS en Côte d'Ivoire, au Mali et au Sénégal

version du 5 août 2019

Ce plan de gestion de données est consultable publiquement sur <https://dmp.opidor.fr/>.

Il suit le plan recommandé par le Digital Curation Centre (DCC - <http://www.dcc.ac.uk/>).

1. Collecte des données	2
1.1. Quelles données allez-vous collecter ou créer ?	2
1.2 Comment les données seront-elles collectées ou créées ?	2
Données d'enquêtes qualitatives	2
Données d'enquêtes quantitatives téléphoniques	3
Données de l'enquête des temps et mouvements	3
Données financières	3
2. Documentation et Métadonnées	4
2.1 Quelle documentation et quelles métadonnées seront associées aux données ?	4
Données d'enquêtes qualitatives	4
Données d'enquêtes quantitatives téléphoniques	4
Données de l'enquête des temps et mouvements	4
Données financières	4
3. Aspects éthiques et juridiques	4
3.1 Comment allez-vous gérer les aspects éthiques ?	4
Données d'enquêtes qualitatives	5
Données d'enquêtes quantitatives téléphoniques	5
Données de l'enquête des temps et mouvements	5
Données financières	6
3.2 Comment allez-vous prendre en compte les aspects de droits d'auteur (copyright) et de propriété intellectuelle ?	6

4. Stockage et sauvegarde	6
4.1 Comment les données seront-elles stockées et sauvegardées pendant le projet de recherche ?	6
Données d'enquêtes qualitatives	6
Données d'enquêtes quantitatives téléphoniques	7
Données de l'enquête des temps et mouvements	7
Données financières	7
4.2 Comment allez-vous gérer l'accès et la sécurité ?	7
5. Sélection et préservation	7
5.1 Quelles données doivent être conservées, partagées et/ou préservées ?	7
Données d'enquêtes qualitatives	8
Données d'enquêtes quantitatives téléphoniques	8
Données de l'enquête des temps et mouvements	8
5.2 Quel est le plan de préservation à long terme des jeux de données ?	8
Partage des données	8
6.1 Comment allez-vous partager les données ?	8
6.2 Existe-t-il des restrictions au partage des données ?	8
Responsabilités et Ressource	9
7.1 Qui sera responsable de la gestion des données ?	9
7.2 De quelles ressources aurez-vous besoin pour mettre en oeuvre votre plan ?	9

1. Collecte des données

1.1. Quelles données allez-vous collecter ou créer ?

Les différents volets de recherche et les collectes associées à chacun d'eux sont détaillés dans le protocole du projet consultable sur <https://atlas.solthis.org/>.

La nature des données collectées est de quatre types : (i) données d'enquêtes qualitatives (enregistrements audios ou prise de notes d'entretiens individuels et de groupe, en face à face ou par téléphone) ; (ii) données d'enquêtes téléphoniques quantitatives (enquête coupons) ; (iii) données d'enquêtes des temps et mouvements ; (iv) données financières (rapports financiers, collecte spécifique auprès de sites dispensateurs).

1.2 Comment les données seront-elles collectées ou créées ?

Données d'enquêtes qualitatives

Il s'agit d'entretiens individuels et de groupe. Les entretiens seront enregistrés (audio, format MP3 ou équivalent) avec le consentement des participants. Pour les entretiens individuels, si l'enquêté-e refuse l'enregistrement audio, l'enquêteur-riche procédera à une prise de notes manuelles.

Chaque entretien sera numéroté en fonction de l'enquête ou de la sous-enquête à laquelle il correspond et du pays de collecte. Le numéro d'entretien sera reporté sur le formulaire de consentement.

Les entretiens seront menés par des anthropologues seniors avec l'appui d'assistant-e-s de recherche formées aux sciences sociales et à la conduite d'entretien. L'équipe qualitative se réunira à intervalles réguliers pour coordonner les activités de terrain et vérifier la qualité des entretiens.

Les enregistrements audio et les prises de note feront l'objet d'une retranscription (fichier texte avec mise en forme) en français (avec traduction le cas échéant pour les entretiens menés dans une autre langue). Lors de la retranscription, les entretiens seront codés (changement des noms et prénoms par exemple) afin de les pseudonymiser.

Données d'enquêtes quantitatives téléphoniques

L'enquête téléphonique (WP Enquête Coupons) sera réalisée par la société Ipsos Côte d'Ivoire disposant d'une plateforme CATI (Collecte Assistée par Téléphone et Informatique) à Abidjan. Trois numéros verts seront mis en place (Côte d'Ivoire, Mali, Sénégal). Tous les appels seront routés vers le plateau CATI via des passerelles mises en place dans chaque pays (voix sur IP). Lors des phases de réception d'appels (phase 1 de l'enquête), les appels entrants seront automatiquement orientés vers un enquêteur ivoirien, malien ou sénégalais en fonction de la provenance de l'appel. Avant le démarrage de la collecte, les enquêteurs auront suivis une formation spécifique à l'enquête et à son questionnaire, menée à la fois par les superviseurs d'Ipsos Côte d'Ivoire et l'équipe de recherche.

Le programme PACCI, institution de recherche ivoirienne basée à Abidjan et partenaire du projet ATLAS, développera une plateforme dédiée à la capture des données, la gestion des rendez-vous téléphoniques, la production de tableaux de bord de suivi de l'enquête et l'exportation des données collectées dans un format textuel (CSV).

Des fichiers analytiques dédiés à l'analyse scientifique seront générés à partir de cette plateforme de collecte. Ces fichiers seront générés dans un format ouvert (fichiers texte de type CSV). Ces fichiers analytiques de l'enquête coupons ne contiendront pas de données personnelles.

La contrôle qualité de la collecte des données sera assuré par la mise en place de contrôles de cohérence au moment de la saisie, la double-écoute par les superviseurs d'entretiens conduits par les enquêteurs et des vérifications sur la base analytique en cours d'enquête. L'enquête coupons étant organisée en 3 vagues, une analyse poussée des données de la première vague sera réalisée avant la conduite des vagues 2 et 3 et les procédures opérationnelles seront ajustés en fonction.

Données de l'enquête des temps et mouvements

Cette enquête sera menée par deux économiste juniors supervisés par deux économistes seniors. Les observations de terrains et le remplissage des questionnaires seront effectués par les deux économiste juniors sur support papier. Les données collectées seront par la suite saisies par ces derniers au format Excel.

L'équipe d'économistes se réunira à intervalle régulier pour vérifier le bon déroulement des enquêtes et la qualité des données collectées.

Données financières

La même équipe d'économistes ira collecter des données financières auprès d'un échantillon de structures sanitaires et communautaires engagées dans la distribution de l'autodépistage du VIH dans le cadre du programme ATLAS. Les données collectées seront elles aussi au format Excel.

2. Documentation et Métadonnées

2.1 Quelle documentation et quelles métadonnées seront associées aux données ?

Données d'enquêtes qualitatives

Chaque retranscription d'entretien sera accompagnée d'un rapport d'entretien incluant : le contexte de passation de l'entretien, les caractéristiques sociodémographiques et une présentation de la ou des personnes interrogées et d'un résumé.

Données d'enquêtes quantitatives téléphoniques

Les fichiers analytiques générés à partir de la plateforme de collecte seront documentés selon la spécification DDI (*Data Documentation Initiative* <http://www.ddialliance.org/>) dédiée à la documentation de données d'enquêtes quantitatives en sciences humaines et sociales. La documentation des variables sera effectuée avec le logiciel *Nesstar Publisher* du *Norwegian Centre for Research Data*. Ce logiciel permet notamment de produire un dictionnaire détaillé des variables et un export des fichiers de données labellisées au format Stata et SPSS. Une copie des données au format CSV (fichiers texte) sera systématiquement conservée.

Données de l'enquête des temps et mouvements

Les fichiers de l'enquête des temps et mouvements seront également documentés selon la spécification DDI.

Données financières

La collecte s'appuiera à la fois sur les différents documents de rapportage produits par la structure générant le site de dispensation (rapports financiers, rapports d'activités...) ainsi que une grille en cours de développement par l'équipe d'économistes.

3. Aspects éthiques et juridiques

3.1 Comment allez-vous gérer les aspects éthiques ?

Le protocole de recherche a été évalué et autorisé par 4 comités d'éthique :

- OMS : Ethical Research Committee, date :
- Côte d'Ivoire : Comité National d'Ethique des Sciences de la vie et de la santé, date : 28 mai 2019
- Mali : Comité d'Éthique de la Faculté de Médecine et de Pharmacie de l'université de Bamako , date :
- Sénégal : Comité National d'Ethique pour la Recherche en Santé, date : 26 juillet 2019

L'ensemble des formulaires de consentement et des notices d'informations sont disponibles en annexe du protocole, sur le site <https://atlas.solthis.org/>

Les formulaires de consentement seront conservés pendant 5 ans après la fin du projet (soit jusque fin 2026) dans une armoire sous clé dans les locaux de Solthis situés à Abidjan pour les formulaires de

consentement collectés en Côte d'Ivoire, Bamako pour les formulaires collectés au Mali et Dakar pour les formulaires collectés au Sénégal.

Données d'enquêtes qualitatives

Avant tout entretien en face à face, un formulaire de consentement éclairé sera signé par la personne enquêtée. Un consentement spécifique sera demandé pour les enregistrements audio. Afin de permettre aux enquêté·e·s d'exercer leur droit d'accès, de rectification et d'opposition, le numéro unique de l'entretien sera reporté sur le formulaire de consentement remis à l'enquêté.

Pour les entretiens menés par téléphone, le consentement sera demandé verbalement à l'enquêté et son anonymat préservé. Un formulaire de recueil du consentement verbal sera rempli par l'enquêteur·rice. Les données de contacts de la personne interrogée seront détruites juste après la passation de l'entretien téléphonique. Le numéro de l'entretien sera précisé à la personne enquêtée et il lui sera expliquée qu'elle doit le conserver si elle veut pouvoir exercer ultérieurement son droit d'accès, de rectification et/ou d'opposition.

Les enregistrements audio des entretiens seront conservés le temps de leur transcription et de leur pseudonymisation. Chaque chercheur·e s'engage à ce que les enregistrements audios soient détruits au plus tard à la fin du projet (fin 2021).

Comme mentionné sur les notices d'information, les données pseudonymisées seront conservées pour une durée maximum de 5 ans après la fin du projet, soit jusque fin 2026 au plus tard. Seules des données anonymisées feront l'objet d'un archivage de longue durée (voir le chapitre "Sélection et préservation").

Données d'enquêtes quantitatives téléphoniques

Étant donnée la nature de l'enquête (appels anonymes sur une ligne téléphonique gratuite), le consentement de participation à l'étude sera demandé oralement. Le recueil du consentement sera notifié et horodaté dans la base de données.

Les données collectées seront par nature anonymes, sauf si la personne enquêtée fournit un ou plusieurs numéros de téléphone dans les cas suivant : (i) souhaite être rappelée ultérieurement pour poursuivre l'entretien ; (ii) est éligible à la phase 2 de l'enquête (questionnaires complémentaire par rappel trois mois plus tard), accepte d'être recontactée et de laisser un numéro de contact ; (iii) accepte d'être recontactée pour un entretien qualitatif complémentaire et de laisser un numéro de contact. En l'absence de numéro de téléphone, la granularité des données collectées ne permettra pas l'identification des personnes par recoupement des informations. Les numéros de téléphones collectés seront stockés dans une table spécifique de la base de données et supprimés à la fin de l'enquête (au plus tard fin 2021).

Si la personne enquêtée a indiqué un numéro de téléphone, elle pourra exercer son droit d'accès, de rectification et/ou d'opposition (identification possible via le numéro de téléphone) pendant la durée de l'étude (2019-2021). Si la personne enquêtée n'a pas indiqué de numéro de téléphone à l'enquêteur, l'exercice du droit d'accès, de rectification et/ou d'opposition ne sera plus possible une fois l'entretien terminé.

Données de l'enquête des temps et mouvements

Un consentement éclairé et écrit sera demandé aux agent·e·s dispensateurs sélectionné·e·s pour cette étude. Un numéro de participation sera mentionné sur le formulaire de consentement afin de leur permettre d'exercer leur droit d'accès, de rectification et/ou d'opposition.

Comme mentionné sur les notices d'information, les données brutes seront conservées pour une durée maximum de 5 ans après la fin du projet, soit jusque fin 2026 au plus tard. Seules des données ayant

fait l'objet d'une anonymisation feront l'objet d'un archivage de longue durée (voir le chapitre "Sélection et préservation").

Données financières

Toutes les copies papier des enregistrements seront conservées par les équipes d'étude dans des armoires verrouillées. Les copies électroniques des données seront stockées dans des ordinateurs protégés par mot de passe ou en accès limité et protégé sur le serveur Microsoft SharePoint de Solthis. L'accès aux dossiers sera limité aux équipes d'économistes et aux membres de l'équipe ATLAS, selon les besoins.

Les données collectées brutes seront conservées pour une durée maximum de 5 ans après la fin du projet, soit jusque fin 2026 au plus tard. Seules des données ayant fait l'objet d'un traitement spécifique pour éviter l'identification des structures enquêtées (données agrégées par exemple) feront l'objet d'un archivage de longue durée (voir le chapitre "Sélection et préservation").

3.2 Comment allez-vous prendre en compte les aspects de droits d'auteur (copyright) et de propriété intellectuelle ?

Les droits de propriété intellectuelle sur les données du projets et leurs dérivés restent acquises par les différentes parties ayant créé ou produit ces données. Cependant, les différents membres du consortium Atlas se sont engagés à partager les données produites entre eux.

Toute publication ou communication scientifique réalisée dans le cadre du projet ATLAS et portant sur les données du projet devra être discutée en amont par le comité de pilotage scientifique du projet et s'inscrire dans le plan de valorisation du projet.

Les données de recherche anonymisée (voir section 5. Sélection et Préservation) qui feront l'objet d'un archivage en accès libre seront distribuées sous licence *Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International* (<https://creativecommons.org/licenses/by-sa/4.0/>).

Les institutions de recherche du projet se sont par ailleurs toute engagées contractuellement à fournir une sous-licence non-exclusive, irrévocable, mondiale, libre de droits et sous-licence à l'OMS pour le compte d'UNITAID pour utiliser les données anonymisées (voir section 5. Sélection et Préservation) du projet et les publications scientifiques produites à des fins non commerciales de santé publique, d'éducation et de recherche.

4. Stockage et sauvegarde

4.1 Comment les données seront-elles stockées et sauvegardées pendant le projet de recherche ?

Données d'enquêtes qualitatives

Les enregistrements audio des entretiens seront conservés sur l'ordinateur personnelle du/de la chercheur-e chargée de la conduite de l'entretien et protégés par mot de passe.

Les entretiens retranscrits et pseudonymisés seront déposés sur la plateforme <https://cloud.atlasti.com/> dédiée à l'analyse collaborative de données qualitatives et notamment à la codification thématique d'entretiens. Cette plateforme est un produit développé et commercialisé par la société *Scientific Software Development GmbH* spécialisée dans les logiciels d'analyses qualitatives scientifiques. La plateforme est sécurisée et les accès individuels et privés. Selon la politique de

confidentialité du service (<https://atlasti.com/privacy/>), la plateforme "ATLAS.ti Cloud" respecte le règlement général sur la protection des données (RGPD) de l'Union Européenne. Les communications seront chiffrées (https), les accès individuels (compte utilisateur) avec gestion des droits.

Données d'enquêtes quantitatives téléphoniques

La plateforme de collecte des données sera hébergée sur un serveur informatique sécurisé localisé dans les locaux du programme PACCI. L'accès à la plateforme (enquêteurs, superviseurs, chercheurs) se fera uniquement au travers d'une connexion chiffrée (https) et d'un compte utilisateur individualisé (les droits de chaque acteur pouvant être définis séparément).

Les analyses statistiques seront réalisées à partir de fichiers analytiques anonymisés extraits de cette plateforme. Les analyses seront effectuées avec le logiciel R et hébergées sur l'infrastructure de la TGIR Huma-Num (<https://www.huma-num.fr/>) qui mets à disposition un serveur R Studio Pro (<https://r-tools.huma-num.fr>) et un gestionnaire de versions Git (<https://gitlab.huma-num.fr>) situés sur des serveurs sécurisés du Centre national de la recherche scientifique (CNRS) en France. Les communications sont chiffrées (https), les accès individuels et la gestion des droits par utilisateur.

Données de l'enquête des temps et mouvements

Les fichiers Excel seront hébergés sur la plateforme Microsoft SharePoint de l'ONG Solthis (Solthis est le porteur du programme ATLAS). L'accès aux fichiers sera limité aux membres de l'équipe de recherche impliqués dans l'analyse économique. Les communications sont chiffrées (https) et les droits individuels (comptes utilisateurs).

Les analyses statistiques seront, quant à elle, menées sur la plateforme de la TGIR Hum-Num (voir point précédent).

Données financières

Les données collectées seront elles aussi au format Excel, et hébergées sur la plateforme Microsoft SharePoint de Solthis.

4.2 Comment allez-vous gérer l'accès et la sécurité ?

L'ensemble des serveurs informatiques retenus (serveurs internes ou prestations fournies par des tiers) sont sécurisés. Toutes les communications sont chiffrées et les droits d'accès sont systématiquement gérés individuellement (voir section précédente).

5. Sélection et préservation

5.1 Quelles données doivent être conservées, partagées et/ou préservées ?

Les données personnelles (données nominatives ou données permettant l'identification des personnes par recoupement d'information) seront conservées au maximum pendant 5 ans après la fin du projet, soit jusque fin 2026 au plus. De même, les formulaires de consentement seront conservés pendant la même période afin de permettre aux personnes enquêtées d'exercer leur droit d'accès, de rectification et/ou d'opposition (voir section 3.1).

Seules des données anonymisées générées à partir des données brutes feront l'objet d'un archivage.

Données d'enquêtes qualitatives

Une procédure d'anonymisation des retranscriptions pseudonymisées des entretiens sera mise en place. Les rapports d'entretiens seront repris un à un et tous les passages contenant des informations susceptibles d'identifier la personne interrogée ou celles qu'elle mentionne seront supprimés (avec indication de la suppression). En cas de doute, la suppression sera favorisée. Les rapports d'entretiens anonymisés seront validés collégialement par les équipes de recherche qualitatives.

Données d'enquêtes quantitatives téléphoniques

L'anonymisation des données sera effectué au moment de la génération de fichiers de données analytiques à partir de la plateforme de collecte des données.

Données de l'enquête des temps et mouvements

Des fichiers analytiques anonymisés seront produits en fin de projet. Dans la mesure où l'enquête a été menée auprès d'une population spécifique (agent·e·s dispensateurs) de taille réduite, des outils d'aide à l'anonymisation des données tels que *ARX Data Anonymization Tool* (<https://arx.deidentifier.org/>) seront utilisés pour guider la sélection et la recodification des variables.

Données financières

De manière similaire, des fichiers analytiques anonymisés (ne permettant pas d'identifier les sites de dispensation) seront générés. Au besoin, le recours à des données agrégées sera envisagé.

5.2 Quel est le plan de préservation à long terme des jeux de données ?

Les fichiers de données anonymisées seront déposés sur le dépôt de données Zenodo : <https://www.zenodo.org/communities/atlas-research/>. Zenodo est un dépôt gratuit et public, hébergé par le CERN et appuyé par l'Union Européenne.

Partage des données

6.1 Comment allez-vous partager les données ?

Les données déposées sur Zenodo disposeront chacune d'un DOI (Digital Object Identifier) permettant leur identification pérenne au cours du temps. Ces DOI seront notamment utilisés dans les publications scientifiques générées à partir de ces données.

6.2 Existe-t-il des restrictions au partage des données ?

Pendant la durée du projet, les fichiers de données seront déposés en accès à la demande, le temps que les équipes de recherche publient les résultats de leurs analyses.

Une fois le projet terminé et les articles correspondants publiés, les fichiers anonymisés de données quantitatives seront en accès libre et distribués sous licence *Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International* (<https://creativecommons.org/licenses/by-sa/4.0/>).

Concernant les données qualitatives, considérant que le processus d'anonymisation décrit précédemment pourrait être source d'erreur et que l'identification de certaines des personnes enquêtées pourrait toujours être théoriquement possible par croisement d'informations, l'accès aux

données archivées sera restreint. L'accès, via la plateforme Zenodo, se fera sur demande, par des équipes académiques et sera soumis à la signature d'un engagement de confidentialité.

Responsabilités et Ressource

7.1 Qui sera responsable de la gestion des données ?

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package.

L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données.

Un organigramme détaillé des rôles et responsabilités sera développé en début de projet.

7.2 De quelles ressources aurez-vous besoin pour mettre en oeuvre votre plan ?

Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.