



HAL
open science

Plan de gestion des données du projet SENOVIE France

Anne Gosselin, Karna Coulibaly, Joseph Larmarange, Clémence Schantz

► To cite this version:

Anne Gosselin, Karna Coulibaly, Joseph Larmarange, Clémence Schantz. Plan de gestion des données du projet SENOVIE France. Institut de Recherche pour le Développement (IRD). 2024. hal-04563793

HAL Id: hal-04563793

<https://hal.science/hal-04563793v1>

Submitted on 30 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

PGD du projet "SENOVIE France"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré" fourni par Science Europe.

Renseignements sur le plan

Titre du plan	PGD du projet "SENOVIE France"
Version	Version initiale
Objet/périmètre du plan	Ce plan de gestion de données couvre les collectes prévues dans le volet France du projet SENOVIE (enquête biographique quantitative et enquête qualitative).
Domaines de recherche (selon classification de l'OCDE)	5.4 Sociologie, 3.3 Sciences de la santé, 5.9 Autres sciences sociales
Langue	français
Date de création	2024-03-29
Date de dernière modification	2024-04-26
Licence	Nom Creative Commons Attribution Non Commercial Share Alike 4.0 International
	URL http://spdx.org/licenses/CC-BY-NC-SA-4.0.json
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	<ul style="list-style-type: none">• Archive des publications issues du projet : https://hal.science/SENOVIE• Site web du projet : https://www.senovie.org/

Renseignements sur le projet

Titre du projet

SENOVIE France · Mobilités thérapeutiques et cancers du sein

Acronyme

SENOVIE FR

Résumé

En France, on compte 61 124 nouveaux cas de cancer du sein en France en 2023 avec une prévalence estimée en 2017 de 913 089 personnes. C'est ainsi le cancer le plus fréquent chez les femmes, avec une incidence en hausse depuis 1990, due à la fois à l'évolution du dépistage mais plus récemment à une hausse modérée en rapport avec l'évolution des facteurs de risque (le tabagisme par exemple).

Les inégalités sociales face au cancer du sein sont documentées, mais il n'existe pas de données sur les femmes immigrées or il y a des raisons de penser que leur trajectoire peut être différente de celles des femmes nées en France (caractéristiques et âge de survenue du cancer, problèmes d'accès aux soins, isolement, etc.)

Le projet SENOVIE France vise à mieux comprendre les répercussions du cancer du sein sur les trajectoires de vie des femmes, avec trois objectifs:

- Comprendre **comment le cancer du sein** (diagnostic, traitement et reconstruction mammaire) **affecte la vie des femmes** dans de nombreux domaines (conséquences sociales du cancer du sein sur des domaines tels que la famille ou la situation professionnelle/financière,...)
- **Étudier les itinéraires thérapeutiques** des femmes atteintes d'un cancer du sein en Ile-de-France et comprendre leurs déterminants sociaux.
- Observer comment le genre (relations de couple, sexualité, image de soi) peut influencer les trajectoires médicales et sociales des femmes

Pour répondre à ces questions, une méthodologie mixte sera mise en œuvre (enquête biographique quantitative, entretiens semi-directifs) dans trois services hospitaliers d'Ile de France.

Sources de financement

- Institut National Du Cancer :
- Institut Convergence Migrations :

Date de début

2024-03-01

Date de fin

2025-12-31

Partenaires

- National Institute for Demographic Studies <https://ror.org/02cnsac56>
- Centre population et développement <https://ror.org/002yg6s88>
- Centre Hospitalier Saint-Denis <https://ror.org/05ed8xr15>
- Hôpital Saint-Louis <https://ror.org/049am9t04>
- Centre hospitalier Robert-Ballanger <https://ror.org/057wh9709>

Produits de recherche :

1. Enquête SENOVIE France (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Centre hospitalier Robert-Ballanger	Centre hospitalier Robert-Ballanger - https://ror.org/057wh9709	• Collecteur de données
Centre Hospitalier Saint-Denis	Centre Hospitalier Saint-Denis - https://ror.org/05ed8xr15	• Collecteur de données
Centre Population & Développement (Ceped UMR 196)	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	• Collecteur de données
Coulibaly Karna - https://orcid.org/0000-0001-5038-8523	National Institute for Demographic Studies - https://ror.org/02cnsac56	• Gestionnaire de données • Gestionnaire de données
Direction des relations internationales et des partenariats	National Institute for Demographic Studies - https://ror.org/02cnsac56	• Autre
Direction du Développement des	Institut de Recherche pour le	• Autre

Usages Numériques Innovants (DDUNI)	Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Autre • Autre
Fusfeld Lindsay	National Institute for Demographic Studies - https://ror.org/02cnsac56	<ul style="list-style-type: none"> • Délégué de protection des données (selon RGPD)
Gosselin Anne - https://orcid.org/0000-0002-6876-8249	National Institute for Demographic Studies - https://ror.org/02cnsac56	<ul style="list-style-type: none"> • Personne contact • Responsable • Responsable • Responsable de la protection des données • Responsable de la qualité des données • Responsable du plan • Responsable du traitement et de l'analyse des données
Hôpital Saint-Louis	Hôpital Saint-Louis - https://ror.org/049am9t04	<ul style="list-style-type: none"> • Collecteur de données
Huma-Num	Huma-Num - https://ror.org/04ces3204	<ul style="list-style-type: none"> • Autre
Institut de Recherche pour le Développement	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Institution d'accueil
Institut National des Études Démographiques	National Institute for Demographic Studies - https://ror.org/02cnsac56	
Larmarange Joseph - https://orcid.org/0000-0001-7097-700X	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Membre du projet
Merquiol Aurélie	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Délégué de protection des données (selon RGPD)
Mortalité, santé, épidémiologie (UR05)	National Institute for Demographic Studies - https://ror.org/02cnsac56	<ul style="list-style-type: none"> • Collecteur de données
Responsable Sécurité des Systèmes d'Information (RSSI) - IRD	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Autre
Schantz Clémence - https://orcid.org/0000-0003-2825-4531	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Coordinateur du projet • Responsable • Responsable • Responsable
Service des Enquêtes et Sondages	National Institute for Demographic Studies - https://ror.org/02cnsac56	<ul style="list-style-type: none"> • Autre
Service Informatique de l'Ined	National Institute for Demographic Studies - https://ror.org/02cnsac56	<ul style="list-style-type: none"> • Autre
Service Partenariats et Valorisation de la DR IDF	Institut de Recherche pour le Développement - https://ror.org/05q3vnk25	<ul style="list-style-type: none"> • Autre

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

PGD du projet "SENOVIE France"

Description des données et collecte ou réutilisation de données existantes

Description générale du produit de recherche

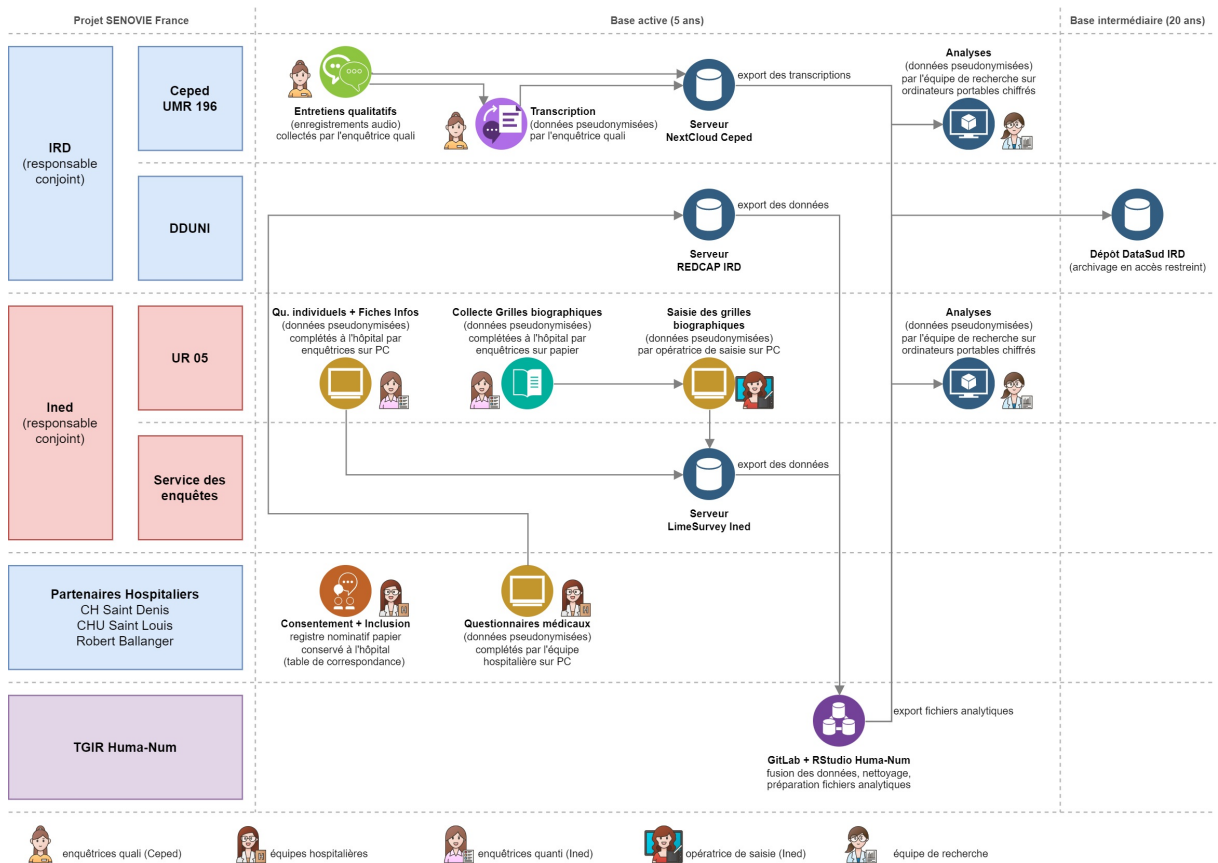
Nom	Enquête SENOVIE France
Description	L'enquête SENOVIE France est une enquête en méthodes mixtes comportant deux volets: <ul style="list-style-type: none">• Volet quantitatif. Enquête rétrospective, biographique, quantitative réalisée auprès de 1050 femmes vivant avec un cancer du sein et suivies dans trois services d'oncologie franciliens. Cette enquête comporte un questionnaire "patiente" et une grille biographique passés en face-à-face par des enquêtrices spécialement formées ainsi qu'un questionnaire médical rempli par l'équipe des services d'oncologie concernés.• Volet qualitatif. Enquête par entretiens semi-directifs auprès de femmes vivant avec un cancer du sein et suivies en Île-de-France. Ces entretiens font l'objet d'un enregistrement audio puis d'une transcription pseudonymisée.
Type	Jeu de données
Mots clés (texte libre)	cancer du sein; trajectoires; migration; itinéraire thérapeutique; genre; inégalités sociales de santé
Langue	français
Contient des données personnelles ?	Oui
Contient des données sensibles ?	Oui
Prend en compte des aspects éthiques ?	Oui

Est-ce que des données existantes seront réutilisées ?

Justification	Dans l'enquête biographique, certaines données du dossier médical des participantes seront collectées au travers d'un questionnaire médical dédié.
----------------------	--

Comment seront produites/collectées les nouvelles données ?

Nom de la méthode	Enquête SENOVIE France
Description	<u>Schéma récapitulatif de la collecte et du traitement des données</u>



Volet quantitatif

Les participantes seront recrutées dans trois structures hospitalières (CH Saint-Denis, CHU Saint-Louis, Hôpital Robert Ballanger).

Un registre est établi pour chaque professionnel de santé qui inclut. Il permet de vérifier la qualité de l'échantillon (critères d'éligibilité, taux de participation, représentativité de l'échantillon). Le registre établit la liste de toutes les patientes éligibles présentes à la consultation. Il est composé de deux volets, un volet anonyme et un volet nominatif qui sera détaché avant la remise des registres aux membres de l'équipe de recherche. Le registre est rempli par le professionnel de santé au cours de la consultation. Il indique pour chaque patiente éligible présente si elle accepte ou refuse de participer ainsi que quelques données sociodémographiques.

Pour chaque patiente sollicitée pour l'étude, le professionnel invite la patiente à lire attentivement la notice d'information de l'étude. À l'issue de l'acceptation de l'enquêtée pour participer à l'étude, le professionnel de santé remplit deux exemplaires du formulaire de consentement et les fait signer à l'enquêtée. Le professionnel de santé garde un exemplaire et donne le 2^e exemplaire à la patiente. Pour chaque patiente acceptant de participer à l'enquête, le professionnel de santé colle une étiquette comportant un numéro d'identifiant sur un carton d'anonymat qu'il remet à la patiente. Ce carton d'anonymat sera donné par la patiente à l'enquêtrice. C'est le numéro porté sur ce carton (ou code patiente) qui sera utilisé comme seul identifiant pour les différents questionnaires d'enquête (questionnaire patiente et grille biographique, questionnaire médical). L'enquêtrice, de même que l'équipe de recherche, ne connaîtra donc pas l'identité de la patiente.

Les formulaires de consentement et le volet nominatif des registres (format papier) seront conservés par les équipes hospitalières.

Pendant les douze mois de collecte, nous collecterons les données suivantes :

- **Questionnaire patiente et grille biographique.**
Le questionnaire patient et la grille biographique comportent les modules suivants :
 - 1) rétrospective sur toute la vie (historique résidentiel, professionnel),
 - 2) histoire des relations et enfants,
 - 3) nationalité, arrivée en France et séjour (pour les femmes immigrées),
 - 4) protection maladie,
 - 5) violences,
 - 6) cancer du sein,
 - 7) période Covid,
 - 8) conséquences économiques,
 - 9) santé, sexualité et bien-être.

La collecte sera réalisée dans le cadre d'un entretien en face à face avec une enquêtrice dédiée, recrutée et formée par l'Ined. La passation du questionnaire aura lieu dans les locaux des partenaires hospitaliers du projet (CH Saint-Denis, CHU Saint-Louis, Hôpital Robert Ballanger) dans des bureaux qui garantissent la confidentialité de l'entretien. Les questionnaires individuels seront saisis en temps réel sur un serveur sécurisé LimeSurvey de l'Ined. Les grilles biographiques seront collectées sur papier avant d'être saisies par un opérateur dédié sur le même serveur LimeSurvey.

- **Questionnaire médical.** Le questionnaire médical contient des questions sur les caractéristiques du cancer et des traitements suivis par les femmes. Il sera complété par le personnel hospitalier des centres partenaires. La saisie se fera sur le serveur sécurisé REDCap de l'IRD.

Volet qualitatif

L'objectif de ce volet qualitatif est de décrire en profondeur la manière dont s'articulent les trajectoires sanitaires, sociales et migratoires des femmes, et leur implication dans la prise en charge de femmes immigrées dans les services d'oncologie français.

Ce volet vise à répondre aux questions suivantes : quelles femmes circulent et dans quelles conditions ? Quels sont les défis et les obstacles qu'elles rencontrent ? Quelles sont les difficultés administratives, matérielles et sociales auxquelles elles sont confrontées ? Quels leviers peuvent être actionnés pour lever ces obstacles, sur quels acteurs peuvent-elles s'appuyer ? Quelles sont les expériences professionnelles, familiales, conjugales, affectives et corporelles de ces femmes ?

Pour répondre à ces questions, nous réaliserons des entretiens qualitatifs avec des femmes venant d'Afrique subsaharienne en Île-de-France pour leur traitement (après avoir été diagnostiquées dans leur pays d'origine) ainsi que des femmes immigrées dont le diagnostic a été fait en France. Nous avons pour objectif de réaliser une quarantaine d'entretiens en Île-de-France. Le nombre exact pourra varier selon le principe de saturation des données. Les entretiens seront enregistrés, pseudonymisés et retranscrits. Ils seront stockés sur un serveur sécurisé NextCloud du Ceped. Les entretiens seront analysés avec une démarche compréhensive qui vise à analyser les expériences telles qu'elles ont été vécues par ces femmes.

Nature des données Observation

Equipements, plateaux techniques utilisés

- Serveur LimeSurvey de l'Ined : <https://survey.web.ined.fr/>
- Serveur REDCap de l'IRD : <https://redcap.ird.fr/>
- Serveur NextCloud du Ceped : <https://cloud.ceppep.org/>

Documentation et qualité des données

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Description

Volet quantitatif

Les fichiers analytiques (tables de données) feront l'objet d'une labellisation (étiquettes de variables et étiquettes de valeurs). Un dictionnaire des variables et des codes sera produits ainsi qu'une notice de description des différentes tables.

Volet qualitatif

Une notice décrivant les différents entretiens, leur contexte de passation et les caractéristiques des personnes enquêtées sera produite.

Références associées

- Protocole de recherche : <https://hal.science/SENOVIE>

Code langue des métadonnées

français

Logiciel de documentation

DDI Tools, DdiEditor

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Description

Différentes mesures seront mises en œuvre pour contrôler de la qualité des données :

- Contrôle au fur et à mesure de la collecte du bon remplissage des questionnaires, débriefs hebdomadaires
- Double saisie d'une partie des grilles biographiques pour repérer les erreurs de saisie
- Tests de cohérence intra bases et entre les différentes bases

Exigences légales et éthiques, code de conduite

Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description

Le traitement des données à caractère personnel a fait l'objet d'une inscription au registre des traitements de l'IRD et d'une analyse d'impact (AIPD). Ce traitement est par ailleurs conforme à la MR-003 de la CNIL.

Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ?
Quelle est la législation applicable en la matière ?

Description

Une convention de reversement a déjà été signée entre l'Ined et l'IRD, afin de verser le budget de la collecte à l'Ined qui en est chargé. À cette convention de reversement s'ajoutera un avenant sur la gestion des données pour préciser les responsabilités de chaque acteur.

Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Description

Il s'agit d'une étude dans le domaine de la santé qui implique la personne humaine. Elle est réalisée dans le respect de la déclaration d'Helsinki et en conformité avec la loi Jardé (dispositions relatives aux recherches non interventionnelles (RIPH 3)) Elle a été approuvée par le Comité de Protection des Personnes Est-1 (numéro national 2023-A01311-44) en novembre 2023, et une déclaration de conformité à la MR 003 a été faite.

Information et consentement des participantes

Volet quantitatif

Après présentation et remise d'une notice d'information, en cas d'acceptation, un consentement papier est signé et conservé dans les services hospitaliers.

Volet qualitatif

Un consentement est demandé avant la réalisation de tout entretien, et un consentement spécifique est demandé pour les enregistrements audios.

Traitement et analyse des données

Comment et avec quels moyens seront traitées les données ?

Description

Un schéma récapitulatif de la collecte et du traitement des données est disponible dans la section 1.3

Volet quantitatif

Les questionnaires individuels et les grilles biographiques seront saisies sur un serveur LimeSurvey hébergé et géré par le service des enquêtes de l'Ined. Le questionnaire médical sera lui collecté sur une instance REDCap hébergée et gérée par la DDUNI de l'IRD.

Les données d'enquêtes (pseudonymisées) seront exportées depuis ces deux plateformes puis déposées dans un répertoire privé sur le serveur GitLab de la TGIR Huma-Num. Un premier traitement des données sera réalisé sur la plateforme RStudio hébergée par Huma-Num afin de procéder à la fusion des tables, leur nettoyage, la production d'indicateurs et la génération de fichiers de données analytiques pour les équipes de recherche.

Pour les analyses, les membres de l'équipe de recherche pourront récupérer les fichiers analytiques à partir du serveur GitLab (accès sécurisés par des comptes individuels et une gestion individuelle des droits). Les fichiers seront copiés sur les ordinateurs professionnels des membres de l'équipe de recherche. Tous ces ordinateurs sont paramétrés pour chiffrer les données stockées sur leur disque dur. Un compte individuel est nécessaire pour accéder au disque dur.

Volet qualitatif

Les enregistrements audios des entretiens individuels seront déposés sur le NextCloud sécurisé du Ceped hébergé en France sur <https://cloudeazy.com/>. Ils ne seront accessibles qu'aux enquêtrices du volet qualitatif, avec comptes individuels. Une transcription textuelle des fichiers audio, avec pseudonymisation des noms et des lieux, sera réalisée par les enquêtrices du volet qualitatif. Les transcriptions pseudonymisées des entretiens seront également déposées sur le NextCloud du Ceped et rendus accessibles aux membres de l'équipe de recherche pour analyses.

Equipements, plateaux techniques

- NextCloud du Ceped : <https://cloud.ceped.org/>
 - Serveur GitLab Huma-Num : <https://gitlab.huma-num.fr/>
 - Serveur RStudio d'Huma-Num : <https://r-tools.huma-num.fr/>
-

Stockage et sauvegarde des données pendant le processus de recherche

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage

Voir le schéma récapitulatif dans la section 1.3 et le descriptif de la section *Traitement et Analyse des données (4.1)*.

Pour le volet quantitatif, la base active sera stockée sur 3 serveurs (serveur LimeSurvey de l'Ined, serveur REDCap de l'IRD, et le serveur GitLab / RStudio d'Huma-Num). Pour le volet qualitatif, les données seront stockées sur le serveur NextCloud du Ceped.

Mesures prises pour la sécurité des données

L'ensemble des outils utilisés (LimeSurvey Ined, REDCap IRD, GitLab Huma-Num, RStudio Huma-Num, NextCloud Ceped) sont hébergés en France sur des serveurs sécurisés, chiffrés et offrant une redondance des données.

Tous ces outils ne sont accessibles via un navigateur web qu'au travers d'une connexion chiffrée (HTTPS). Chaque outil permet la création de comptes individuels et une gestion des droits d'accès utilisateur par utilisateur. Chaque utilisateur n'aura accès qu'aux outils et aux données qui le concernent.

Il sera expressément demandé aux membres de l'équipe de recherche de ne pas enregistrer leurs mots de passe sur leurs ordinateurs. Ainsi en cas de perte ou de vol d'ordinateur, la sécurité des données n'est pas engagée.

Les connexions entre un ordinateur portable et un des services se font par le biais de jetons (*token*) dédiés. Il sera aussi possible de suspendre les droits d'un appareil particulier, par exemple en cas de perte ou vol d'un ordinateur.

L'ensemble des ordinateurs des membres de l'équipe de recherche sera configuré pour que les disques durs soient chiffrés et accessibles uniquement après identification.

Pour plus de détails, se référer à l'analyse d'impacts.

Partage des données et conservation à long terme

Comment les données seront-elles partagées ?

Modalités de partage**Base active (5 ans)**

Les données seront conservées en base active pendant 5 années (voir schéma récapitulatif dans la section 1.3).

À l'issue de cette période, toutes les données directement identifiantes (formulaires papiers conservés au sein des structures hospitalières, enregistrements audios) seront détruites. De même, les données sources sur les serveurs LimeSurvey, RedCAP, Huma-Num, NextCloud seront également détruites.

Base intermédiaire (20 ans)

Les formulaires de consentement prévoient la possibilité de réutilisation des données pour d'autres études exclusivement à des fins scientifiques. Les participantes peuvent se tenir informées des nouvelles utilisations via le portail HAL dédié au projet SENOVIE : <https://hal.science/SENOVIE>.

Pour les participantes ayant donné leur accord, un archivage des bases analytiques sera effectué sur la plate-forme DataSud de l'IRD pour partage à la communauté des chercheurs. Seules des données pseudonymisées feront l'objet d'un archivage en base intermédiaire pour une durée de 20 ans.

Cet archivage sera en accès restreint, sous réserve d'une demande motivée et justifiée et de l'avis favorable des responsables scientifiques de l'enquête (Anne Gosselin et Clémence Schantz).

Potentiel de réutilisation

dans un cadre académique uniquement

Entrepôt/Catalogue de données

- DataSuds : <https://cat.opidor.fr/index.php/DataSuds> ()

Comment les données seront-elles conservées à long terme ?**Justification**

L'archivage sera porté par l'IRD, dans le cadre de la plateforme Data Sud, voir 6.1 et schéma en section 1.3.

Date de début

2029-01-01