

**[TITRE COMPLET DE LA RECHERCHE]**

**Sommaire du protocole scientifique**

[**PRÉSENTATION DE L’ÉQUIPE PROJET [1 PAGE MAXIMUM]**](#_heading=h.30j0zll)5

[**OBJECTIFS ET FINALITÉS [2 - 4 PAGES MAXIMUM]**](#_heading=h.naq8wt60i932)6

[Contexte, objectif(s) et justification de l’étude](#_heading=h.6fp7ife5ig56) 6

[Justification du respect de l’éthique](#_heading=h.pflbil6sfrq5) 6

[Justification de l’intérêt public](#_heading=h.9qyrc96005c4) 6

[Publication des résultats et valorisation](#_heading=h.gpdr4t8skkkt) 6

[**MÉTHODOLOGIE [3 - 5 PAGES MAXIMUM]**](#_heading=h.adc6t9m4vg0z)7

[Design de l’étude](#_heading=h.42xcfldzewtk) 7

[Description et justification de la population d’étude](#_heading=h.nwwml3kzvn1s) 7

[Taille de la population](#_heading=h.oucs5ji4u46a) 7

[Sources de données](#_heading=h.ubehqw53weui) 7

[Variables](#_heading=h.al8a34qnqraf) 83

[Préparation de données](#_heading=h.ebkzk0rywe7m) 83

[Méthodes, traitements et analyses des données](#_heading=h.3i6jdop8v0gj) 9

[Limites de l’étude](#_heading=h.vxmhh5ujv63) 94

[Calendrier prévisionnel et faisabilité du projet](#_heading=h.310tmdqe3ndv) 10

[**PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES [10 PAGES MAXIMUM] (Section destinée à la CNIL)**](#_heading=h.qa4eqzgkhjce)11

[Respect des droits des personnes concernées](#_heading=h.l9y0egk082h4) 116

[Confidentialité et sécurité des données](#_heading=h.tks4jinhbwbg) **Erreur ! Signet non défini.**0

[Gestion du risque de réidentification](#_heading=h.janvqp1pfxzq) 120

[Support de données](#_heading=h.eudpuntjj3fb) 121

[Circuit des données en cas d’appariement](#_heading=h.ms42vx4jxx06) 122

[**ANNEXES**](#_heading=h.c83bjdw1yuhg)14

# PRÉSENTATION DE L’ÉQUIPE PROJET [1 PAGE MAXIMUM]

Les établissements du Groupement de Coopération Sanitaire des Hôpitaux Universitaires du Grand Ouest (GCS HUGO) ont constitué leur entrepôt de données de santé (EDS) à partir d’une plateforme technologique identique : le système eHOP (« entrepôt-Hôpital »).

**A l’échelle locale**, chaque établissement a mis en place une structure dédiée, le Centre de données cliniques. Cette structure assure l’intégration des données patients issues des différentes applications du système d’information hospitalier et apporte l’expertise et les services nécessaires à leur exploitation. Le **Réseau interrégional des Centres de Données Cliniques (RiCDC)** du Grand Ouest coordonne, sous l’égide du GCS HUGO, la mise en œuvre et l’exploitation des gisements de données massives en santé des CHU et des Centres de Lutte contre le Cancer du Grand Ouest.

**A l’échelle inter-régionale**, le GCS HUGO s’est doté d’une plateforme technologique pour le traitement des données massives en santé pour les projets multicentriques, la plateforme Ouest Data Hub (ODH). Elle apporte, aux porteurs de projets des établissements adhérents, des outils pour le traitement de leurs données de recherche (pseudonymisation, appariement, nomenclatures de référence, logiciels et méthodes statistiques).

Chaque projet est évalué par le Comité scientifique et éthique (CSE) de la plateforme ODH. Le CSE évalue le projet au regard de l’intérêt public, de la qualité scientifique, des garanties éthique et déontologique et du respect des dispositions réglementaires relatives à la protection des personnes (RGPD) et aux traitements de données de santé à caractère personnel (loi Informatique et libertés du 6 janvier 1978 modifiée par le décret 2019-536 du 29 mai 2019). Il est composé d’un expert scientifique de chaque établissement, d’un responsable valorisation, d’un représentant des usagers, d’un éthicien, d’un épidémiologiste, d’un médecin DIM et d’un DPO.

Le projet *Nom Projet* sera mené sur la plateforme ODH. Le responsable de traitement est “*nom du chu promoteur”*. Le responsable de la mise en oeuvre est “*nom équipe de recherche*”.

Le dpo du responsable de traitement est “*Nom, adresse postale, téléphone, mail* »

Equipe projet

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervenant** | **Structure** | **Fonction** | **Rôle** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **CDC Partenaires** | Etablissements partenaires |  |  |

# OBJECTIFS ET FINALITÉS [2 - 4 PAGES MAXIMUM]

## Contexte, objectif(s) et justification de l’étude

## Justification du respect de l’éthique

Proposition : Cette étude est une étude sur données de santé collectées pour le soin dans des entrepôts de données de santé autorisés par la CNIL. La confidentialité des données des personnes concernées est assurée par les processus de pseudonymisation et d’échange de données mis en œuvre.

Compte-tenu du nombre de personnes concernées une information individuelle n’est pas envisageable. Afin de donner, aux patients concernés, la possibilité de s’opposer à l’utilisation de leurs données, une communication grand public sera organisée. L’étude sera présentée sur les sites web du GCS HUGO et des établissements participants. Des mécanismes ont été prévus pour que les données des personnes qui le désirent soient retirées.

Le projet a reçu un avis favorable du Comité Scientifique et Ethique de la plateforme ODH. La finalité ne présente pas de risque psychologique ou de stigmatisation pour les patients.

*Expliciter le bénéfice attendu pour le patient et la collectivité. Repris dans le paragraphe Justification de l’intérêt public.*

Les membres de l’équipe projet n’ont pas déclaré de liens d’intérêt en rapport avec le projet ou *clarifier les liens d’intérêt.*

## Justification de l’intérêt public

## Publication des résultats et valorisation

# 

# MÉTHODOLOGIE [3 - 5 PAGES MAXIMUM]

## Design de l’étude

## Description et justification de la population d’étude

## Taille de la population

## 

## Sources de données

Sources 1

Bases de données : Entrepôts de données de santé eHop des CHU de *à compléter*

Producteur : CHU de *à compléter*

Légalité de la base L’EDS est autorisé par la CNIL depuis date , autorisation n°*XXXXX* de données : délibération *numéro*

*CHU de Nantes : Délibération n° 2018-295 du 19 juillet 2018, Demande d’autorisation n° 2129203*

*CHU de Rennes : Délibération n°2020-028 du 27 février 2020, Demande d’autorisation n°2212496*

*CHU de Tours : Dossier d’autorisation n°2212853 - Accord tacite 12/11/2019*

*CHU d’Angers : Décision DT-2020-003*

*Institut de Cancérologie de l’Ouest : Autorisation n°2200363 – Autorisation tacite*

*CHU de Brest : dossier déposé à la CNIL fin avril, en cours d’instruction à la CNIL, demande n°2217775*

Collecte des données Extraction des données de l’EDS par programme

Données collectées *liste des variables et* *période d’extraction*

Le chaînage des données des patients répondant aux critères de l’étude qui seraient passés dans plusieurs des établissements participant au projet est réalisé à partir des variables suivantes : NIR assuré, sexe, date de naissance, rang de naissance, numéro d’étude.

Sources 2

Base de données : Système de gestion administrative dues CHU de *à compléter*

Producteur : CHU de *à compléter*

Légalité de la base Autorisation CNIL pour les applicatifs des SIH

de données :

Collecte des données Extraction des données par programme

Données collectées NIR assuré, sexe, date de naissance, rang de naissance

Un traitement du NIR est réalisé localement dans chaque établissement. Une fonction de hachage lui est appliquée. Il est traité par la DSI de chaque établissement par une personne en charge de la gestion de l’application administrative du système d’information de l’établissement soumise au secret professionnel dans un environnement sécurisé qui sera la seule à accéder à cette donnée. Il n’est pas exporté. Le pseudonyme issu de son traitement fait l’objet d’une deuxième pseudonymisation sur la plateforme interrégionale.

L'utilisation du NIR est justifiée car il  permet un chaînage fiable et simple des  données des patients qui ont été pris en charge dans plusieurs établissements dans des conditions de sécurité. Les autres méthodes d'appariement ne permettent pas un chaînage exhaustif et peuvent être entachées d’erreur. Elles peuvent nécessiter un retour à des données quasi-identifiantes (dates, commune de naissance par exemple) ou à des données sensibles (pathologies, actes médicaux) pour confirmer l’appariement. Elles ne mettent pas à l’abri d’erreurs, doublons de patients ou chaînage incorrect. Ces erreurs pourront être source de biais dans l'interprétation des résultats d’analyse.

*Le cas échéant justifier la collecte du code commune de résidence, les dates exactes de soins ou de décès, le mois et année de naissance.*

*Autre source le cas échéant*

## Variables

Cf guide CESREES

## Préparation de données

Y inclure la *méthodologie de ciblage pour les variables non structurées*

*Localement*

Les données sont pseudonymisées selon la procédure décrite dans les paragraphes 3.4 Source des données et 4.2.3 Circuit des données.

*Sur la plateforme ODH*

L’appariement de données des patients venus dans plusieurs établissements est réalisé sur la plateforme ODH

(cf paragraphe Circuit des données).

## Méthodes, traitements et analyses des données

## Limites de l’étude

## Calendrier prévisionnel et faisabilité du projet

# PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES [10 PAGES MAXIMUM] (Section destinée à la CNIL)

## Respect des droits des personnes concernées

* + 1. **Information individuelle des patients**

Les établissements fournissant des données pour cette étude réalisent une information générale collective et individuelle concernant la réutilisation des données de soins pour la recherche : affiches, livret d’accueil, site Internet, distribution de notes d’information en systématique, courrier postal, selon les modalités décrites dans leur dossier de demande d’autorisation. La note d’information remise personnellement à chaque patient précise les modalités pour prendre connaissance des études qui seront menées.

Les patients auront connaissance de l’étude sur le site internet de l’établissement. La note d’information sur l’étude sera publiée sur le site internet de chaque établissement partenaire ainsi que sur le site internet de la plateforme ODH.

Pour les patients n’ayant pas bénéficié de l’information sur la réutilisation de leurs données de soins une demande de dérogation à l’information individuelle est sollicitée auprès de la CNIL.

* + 1. **Respect du droit des personnes**

Une procédure est définie pour permettre au patient de faire valoir ses droits. Une information sur l’étude est publiée sur le site internet de chaque établissement. Le responsable de l’étude fournit au patient dans la note d’information un contact lui permettant de faire valoir ses droits. Si le patient veut faire valoir ses droits il s’adressera à la personne contact de l’établissement qu’il a fréquenté.

Le patient pourra faire valoir ses droits dans les 3 mois qui suivent la publication de l’étude sur le site internet des établissements et le site internet du GCS HUGO. Durant cette période les pseudonymes pourront être regénérés. L’administrateur technique de la plateforme ODH communiquera au responsable du projet l’identifiant patient plateforme pour l’étude.  Ce dernier pourra alors répondre à la demande d’exercice des droits formulé par le patient. Au-delà de cette période, le responsable de traitement fait valoir les dispositions de l’article 11 du RGPD. Les secrets ne sont pas conservés après gel de la base pour des questions de sécurité. Cependant, le patient pourra fournir des informations complémentaires qui pourraient permettre de l’identifier et permettre ainsi au responsable du projet de répondre à sa demande.

Le patient pourra donc exercer son droit d’opposition pour l’étude dans les 3 mois qui suivent la publication de l’étude sur le site internet des établissements et le site internet du GCS HUGO, période au cours de laquelle les analyses n’ont pas encore été réalisées. Le responsable de traitement invoque le motif légitime et impérieux pour limiter l’exercice du droit d’opposition à cette période. Les oppositions qui surviendraient au-delà de cette période sont susceptibles de compromettre la réalisation des objectifs du projet. Si sa demande intervient avant transmission des données sur la plateforme ODH le CDC de l’établissement concerné ne transmet pas les données le concernant.

### Gestion du risque de réidentification

Les données sont pseudonymisées. Un pseudonyme propre à chaque patient de l’étude est généré. Ce numéro est commun à tous les établissements. Il permettra de chaîner les données si le patient est passé dans plusieurs établissements. Il y a deux étapes de pseudonymisation. Localement dans chaque établissement un premier pseudonyme est généré. Une fonction de hachage est appliquée sur les variables suivantes : NIR assuré, sexe, date de naissance, rang de naissance. Un sel spécifique à l’étude sera utilisé. Un deuxième pseudonyme est créé lors de l’intégration des données sur la plateforme ODH par application d’une fonction de hachage. La procédure de pseudonymisation a été validée par la CNIL pour les projets antérieurs menés sur la plateforme ODH.

Les fichiers de correspondance ne sont pas conservés sur la plateforme (cf circuit des données). Les secrets (sels utilisés pour la génération des pseudonymes) sont effacés après le gel de la base de données.

La date de naissance complète n’est pas recueillie. *Seuls le mois et l’année de naissance sont recueillis ou seul l’âge au moment de la venue est recueilli.*

Le cas échéant*. Le recueil des dates complètes est limité au strict nécessaire.*

L’utilisateur signe une charte utilisateur de la plateforme ODH dans laquelle il s’engage à ne pas essayer de réidentifier les personnes.

La base de données de l’étude est conservée 2 ans en base active après la fin des analyses. Seuls des résultats agrégés sont exportés en dehors de la plateforme ODH.

### Support des données

L’analyse des données aura lieu dans un espace projet dédié sur la plateforme ODH. La plateforme est hébergée au CHU de Nantes Hébergeur de données de santé certifié. La base projet n’est mise à disposition que des utilisateurs habilités par le responsable de traitement dans cet espace sécurisé. Un contrôle d‘accès logique avec authentification forte est mis en place. Une démarche d’homologation sécurité de la plateforme ODH a été initiée. En signant la charte utilisateur de la plateforme l’utilisateur s’engage à assurer la confidentialité des données, à mettre en place les mesures de sécurité sur son poste de travail.

Les mesures de sécurité sont détaillées dans le PIA du projet.

### Circuit des données en cas d’appariement

*Décrire le circuit local le cas échéant (pour des données qui ne seraient pas issues des EDS).*

**Les différentes étapes :**

Étape 1 -  Constitution du jeu de données à partir des  EDS des CDC partenaires (Annexe 1)

Chaque responsable de CDC extrait de son EDS local le jeu de données correspondant au besoin de l’étude conformément aux critères d’extraction définis dans le protocole. Les données sont pseudonymisées avec un identifiant local EDS pour l’étude (Id\_pat\_etude\_cdc1).

Un pseudonyme propre à chaque patient de l’étude (id1\_hugo\_étude) est généré par un automate (boîte noire « hugo\_pseudo »). Ce numéro est commun à tous les établissements. Il permettra de chaîner les données si le patient est passé dans plusieurs établissements. Pour générer ce numéro une fonction de hachage (algorithme SHA-256) est appliquée sur les variables suivantes : NIR assuré, sexe, date de naissance, rang de naissance. Un sel spécifique à l’étude sera utilisé. La procédure de pseudonymisation a été validée par la CNIL.

Le CDC met en entrée de l’automate un fichier avec la correspondance entre l’identifiant local EDS de l’étude (Id\_pat\_etude\_cdc1) et l’identifiant patient (IPP) dans le système d’information hospitalier (SIH). La DSI récupère de l’automate le fichier des IPP (fich2) et génère un fichier avec les variables qui feront l’objet de la pseudonymisation (fich3). Il met ce fichier en entrée de l’automate. Le pseudonyme est généré (id1\_hugo\_étude), le fichier de correspondance est créé (fich4), il est chiffré avec la clef publique du projet dans hugo\_pseudo et il est transféré automatiquement sur le serveur FTP de l’espace projet sur la plateforme ODH.

Chaque CDC transmet son jeu de données pseudonymisées sous forme chiffrée via un réseau sécurisé (VPN).

Les fichiers sont supprimés après validation de l’intégration des données dans la base de données de l’étude sur la plateforme inter-régionale.

Étape 2 - Intégration des jeux de données dans un espace projet chez  l’hébergeur (Annexe 2)

Un espace projet dédié, cloisonné et sécurisé est créé chez l’hébergeur.

Etape 2.1 Pour chaque établissement le fichier de correspondance et le jeu de données dénommés ci-après fichiers source sont réceptionnés sous forme chiffrée sur le serveur SFTP de l’espace projet.

Etape 2.2 Les fichiers source sont déchiffrés. Après déchiffrement du fichier de correspondance fich4 un nouvel identifiant anonyme (id2\_hugo\_etude) est créé à partir de l’id1\_Hugo\_etude par application d’une fonction de hachage. Le jeu de données est rattaché au nouvel identifiant. Les numéros identifiants affectés antérieurement sont supprimés. Il n’y a pas de table de correspondance conservée. Un nouveau jeu de données est ainsi créé.

Etape 2.3 Lors de la phase d’intégration le datamanager HUGO effectue les tâches de contrôle qualité des jeux de données (formats, nomenclatures….). Les données sont ensuite insérées dans une base postgres sur un serveur dédié pour mise à disposition à l’équipe projet et le cas échéant chaînées sur  le numéro identifiant anonyme (id2\_hugo\_etude).

Après intégration les jeux de données sont supprimés.

Etape 3 - Analyse par l’équipe projet

Le datamanager de l’équipe projet effectue les contrôles qualité des données sur la base projet. Les fichiers source sont détruits sur le serveur FTP après contrôle qualité des données.

Les données seront analysées par les personnes habilitées définies dans le protocole scientifique. Seuls les résultats agrégés seront exportés par les utilisateurs

Etape 4 – Export des résultats

Seuls les résultats des analyses sont exportés sous forme de statistiques agrégées. Les données individuelles ne sont pas transférées en dehors de la plateforme.

Etape 5 - Elimination des données

Les données seront conservées en base active 3 ans. *à modifier si durée prévue est différente*

Au-delà les données seront supprimées de la base.

# ANNEXES

Annexe 1 : Circulation des données étape 1



Annexe 2 Circulation des données étape 2

