



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

CNIG

Conseil national
de l'information
géolocalisée

Géolocalisation des données de santé

ACTION 18 DU PNSE4

Données de santé géolocalisées : usages et enjeux

Conseil plénier du Conseil National de l'Information Géolocalisée (CNIG) - 30 mai 2023

Une définition large de la donnée de santé

Pour la CNIL, et d'après le RGPD, les données de santé sont :

« Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne »

Données de santé par nature

Les données de santé par nature : antécédent médicaux, maladie, ...

Données de santé par destination

Les données qui deviennent des données de santé **en raison de l'utilisation** qui en est faite.

Les données qui, **croisées avec différentes bases de données**, deviennent des données de santé.

Ex : Croisement de la tension avec une mesure de l'effort, du poids avec le nombre de pas, ...

A quelle maille géographique mobiliser les données de santé ?

Maille de restitution souhaitée

Une détermination **par l'usage** : l'unité spatiale choisie pour l'étude vise avant tout à être pertinente vis-à-vis de la problématique adressée.

Une restitution à une maille géographique plus agrégée :

- _ Permet d'augmenter l'hétérogénéité à l'intérieur de l'unité spatiale et favorise la possibilité de comparaison de plusieurs groupes
 - _ Emet l'hypothèse d'une répartition homogène sur la zone, et implique ainsi une perte d'information.
-

Contraintes à prendre en compte

- _ Les **caractéristiques de la donnée**, dont la maille géographique la plus fine disponible est hétérogène/dépendante du jeu de données.
 - _ Les **enjeux juridiques** associés aux données de santé et aux données géolocalisées.
-

Usages de la donnée de santé au service de la santé-environnement

l'importance de l'information géolocalisée

Type d'opération	Définition	Application santé-environnement
Géocodage.	Transformation d'une référence géographique implicite (e.g. adresse postale) en référence géographique explicite (coordonnées spatiales).	Localisation des patients à partir de l'adresse de résidence. NB : outil : https://adresse.data.gouv.fr/api-doc/adresse . L'API adresse permet de faire du géocodage et géocodage inverse.

Usages de la donnée de santé au service de la santé-environnement

l'importance de l'information géolocalisée

Type d'opération	Définition	Application santé-environnement
Géocodage.	Transformation d'une référence géographique implicite (e.g. adresse postale) en référence géographique explicite (coordonnées spatiales).	Localisation des patients à partir de l'adresse de résidence. NB : outil : https://adresse.data.gouv.fr/api-doc/adresse . L'API adresse permet de faire du géocodage et géocodage inverse.
Analyse de distances/surfaces.	Identification des objets les plus proches, calcul de superficies, création de zones tampon.	Calcul d'indicateurs de proximité. Ex : sélection des patients habitant à moins de X mètres d'une infrastructure, calcul du nombre de lieux à risque dans un rayon donné...

Usages de la donnée de santé au service de la santé-environnement

l'importance de l'information géolocalisée

Type d'opération	Définition	Application santé-environnement
Géocodage.	Transformation d'une référence géographique implicite (e.g. adresse postale) en référence géographique explicite (coordonnées spatiales).	Localisation des patients à partir de l'adresse de résidence. NB : outil : https://adresse.data.gouv.fr/api-doc/adresse . L'API adresse permet de faire du géocodage et géocodage inverse.
Analyse de distances/surfaces.	Identification des objets les plus proches, calcul de superficies, création de zones tampon.	Calcul d'indicateurs de proximité. Ex : sélection des patients habitant à moins de X mètres d'une infrastructure, calcul du nombre de lieux à risque dans un rayon donné...
Superposition ou jointure spatiale	Sélection d'entités en fonction de relations spatiales.	Agrégation de plusieurs mesures de polluants au sein d'une même zone.

Usages de la donnée de santé au service de la santé-environnement

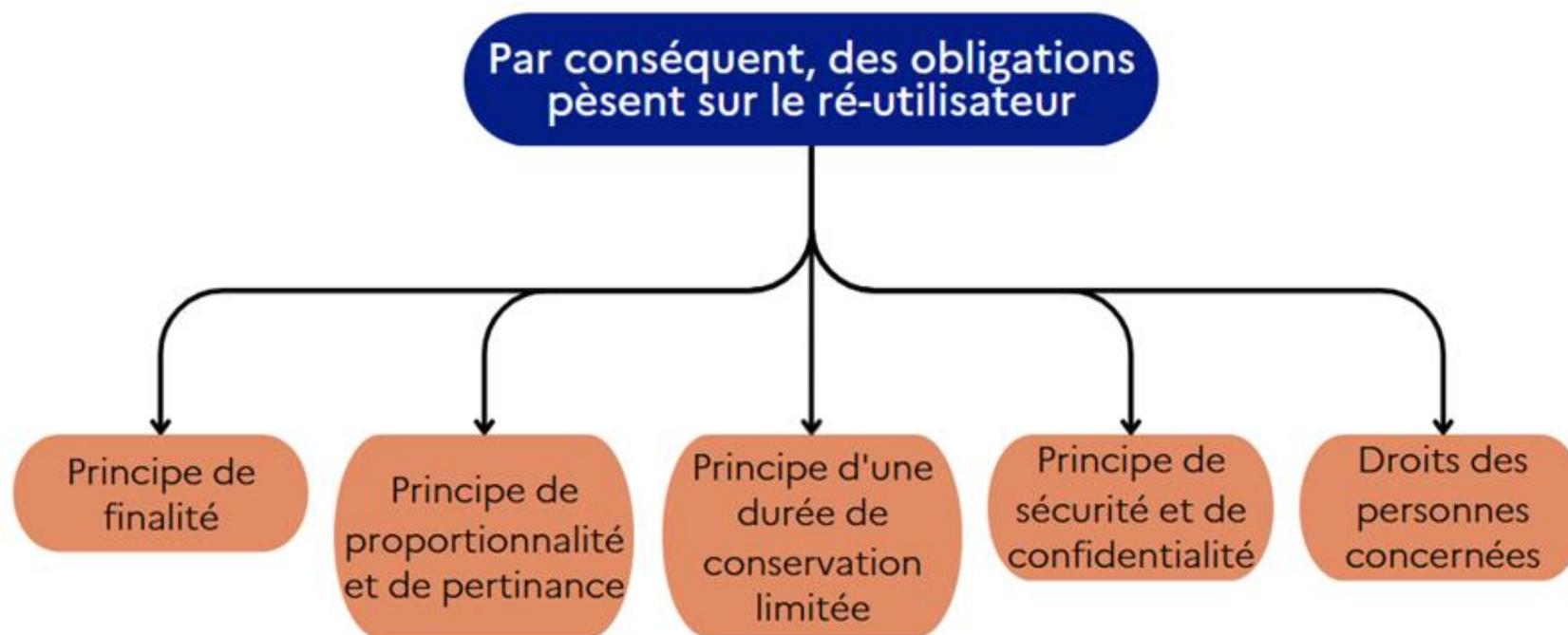
l'importance de l'information géolocalisée

Traitements statistiques & modélisation	Définition	Application santé-environnement
Analyse de tendance spatiale	Utilisation de modèles pour évaluer le degré de significativité d'une tendance spatiale. NB : la cartographie permet d'observer des patterns sans toutefois démontrer leur significativité statistique.	Exemple : détection de points noirs/points chauds : lorsque des valeurs élevées sont également entourées d'autres entités présentant des valeurs élevées. ⇔ Identification de l'agrégation spatiale statistiquement significative de valeurs élevées (points chauds) et de valeurs faibles (points froids)

Rappel : enjeux juridiques liés à l'utilisation des données géolocalisées

« Pour la CNIL, [...] les objectifs parfaitement légitimes poursuivis par la politique d'ouverture des données publiques sont pleinement conciliables avec la protection de la vie privée »

Une vigilance devra être apportée quant aux **modalités de leur réutilisation** (notamment lors de croisements avec d'autres données pouvant révéler des données à caractère personnel).



Enjeux juridiques : la spécificité des données de santé

D'après l'article 9.1 du RGPD : « **Le traitement des données [...] concernant la santé [...] sont interdits** ».

Des **dérogations sont néanmoins accordées** (article 9.2 de la RGPD)

Les trois principales finalités utilisées en recherche sont :

- le **consentement explicite** (article 9.2 a),
- les **finalités de recherche scientifique** Article 9.2 j
- **l'intérêt public dans le domaine de la santé publique** (article 9.2 i).

Ces exceptions doivent être combinées à une des bases légales prévues à l'article 6.1 du RGPD.

L'article 89.1 du RGPD impose pour les traitements de données la mise en place de **mesures techniques et organisationnelles appropriées** pour la préservation des droits et libertés des personnes concernées (**pseudonymisation, ...**).

Certaines exceptions sont prévues et documentées par l'article 89.2

Pour respecter le droit des personnes, il faut notamment :

- _ **Informers les personnes concernées** sur le traitement entrepris
- _ **Permettre un droit d'opposition** avant tout traitement.

Enjeux juridiques : la spécificité des données de santé

Pour faciliter l'utilisation des données de santé, différents outils sont disponibles :

Pour un projet de recherche sur des données de santé, une **Analyse d'Impact Relative à la Protection des Données (AIPD)** est à réaliser systématiquement.

Pour la conduire, la CNIL propose un modèle de rédaction et l'outil PIA, un simulateur permettant de préparer la réalisation d'une AIPD

et

L'utilisation de la **méthodologie de référence MR-004** encadre les traitements de données dans le domaine de la santé.

Si la proposition de recherche ne peut se conduire dans le cadre de cette méthodologie, un **processus plus long pour obtenir l'accès aux données de santé doit être mis en place**

NB : En France et aux Etats Unis, les lois relatives aux données de santé ne sont pas les mêmes.

ACTION 18 DU PNSE4

GD4H - Green Data for Health

**Pour la mobilisation et valorisation, par les acteurs de la recherche et expertise,
des données environnementales au service de la santé environnement.**

Conseil plénier du Conseil National de l'Information Géolocalisée (CNIG) - 30 mai 2023

La santé-environnement : de quoi parle-t-on ?

Santé environnementale

Selon l'OMS, le terme désigne « les aspects de la santé humaine, y compris la qualité de la vie, qui sont déterminés par les facteurs physiques, chimiques, biologiques, sociaux, psychosociaux et esthétiques de notre environnement ». Elle désigne également « la politique et les pratiques de gestion, de résorption, de contrôle et de prévention des facteurs environnementaux susceptibles d'affecter la santé des générations actuelles et futures ».

Au sens large, le terme peut recouvrir l'étude des expositions aux risques d'un individu qui peuvent affecter sa santé physique ou mentale :

- Risques physiques (e.g. rayonnements, bruits)
- Risques chimiques (e.g.)
- Risques biologiques (e.g. risques infectieux)
- Risques psycho-sociaux

LA SANTÉ ENVIRONNEMENT DANS LE PNSE4

« Le PNSE 4 propose des actions concrètes pour mieux comprendre et réduire les risques liés aux **substances chimiques, aux agents physiques (comme le bruit ou les ondes) et aux agents infectieux en lien avec les zoonoses**, c'est-à-dire les pathologies qui peuvent se transmettre de l'animal à l'homme.

Il s'inscrit pleinement dans le cadre de la démarche « **Une seule santé** » »

Source : <https://www.ecologie.gouv.fr/plan-national-sante-environnement-pnse>

OBJECTIF DU GREEN DATA FOR HEALTH

(action 18 PNSE4)

Permettre une meilleure mobilisation et valorisation, par les acteurs de la recherche et de l'expertise, des données environnementales au service de la santé environnement.



Facilite la **repérabilité** et l'**accès** aux données environnementales y compris au niveau des territoires



Décrit les données environ. pertinentes et disponibles pour un **appariement rigoureux** avec des données de santé

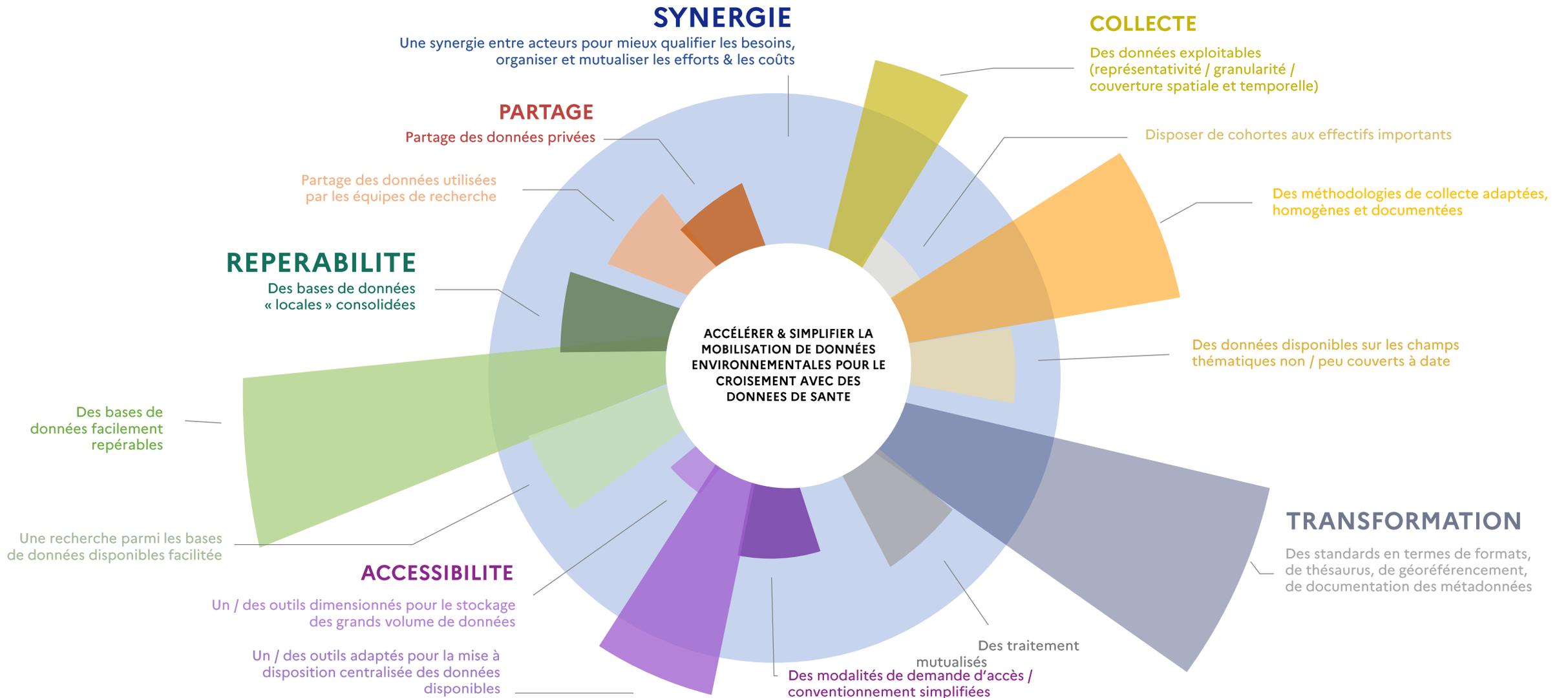


Améliore la **communication** et la **synergie** entre les acteurs de la santé environnement (notamment sur les enjeux d'interopérabilité)

UTILISATEURS FINAUX DU GD4H

- _ Chercheurs et experts en santé environnement
- _ Producteurs de données environnementales

Bilan de la phase d'investigation auprès des utilisateurs finaux du Green Data for Health*



*après entretiens, ateliers et cas concrets auprès des utilisateurs cibles de la recherche et expertise

L'offre de service du Green Data for Health (CGDD/Ecolab) en fin avril 2023

Gouvernance

Animation

- Animation d'une **communauté** multi-acteurs (plénières, première newsletter).
- Conduite de **cas concrets**.
- **Appel à projets HDH x GD4H**.
- **Challenge GD4H**.

Catalogue de ressources

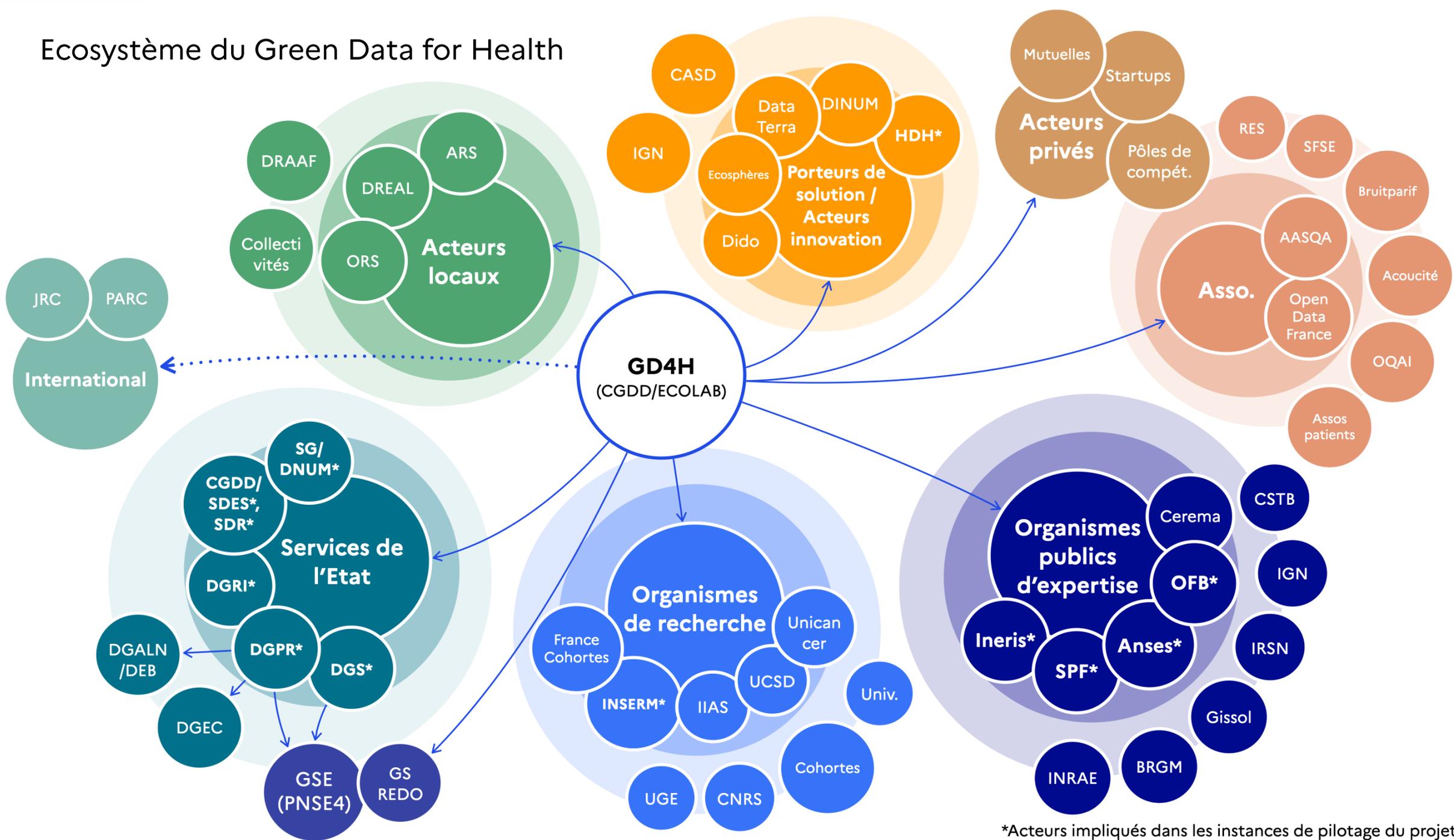
Ouverture au grand public du catalogue de données en ligne.

Outils juridiques

- Ouverture au grand public de la ressource juridique en ligne.
- Conception en cours d'un outil juridique d'autodiagnostic.

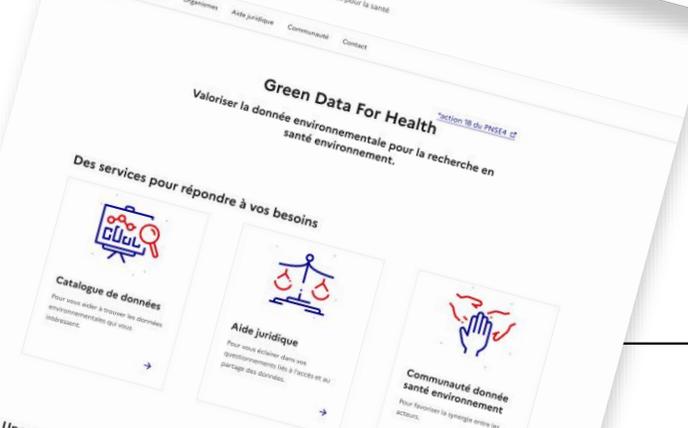
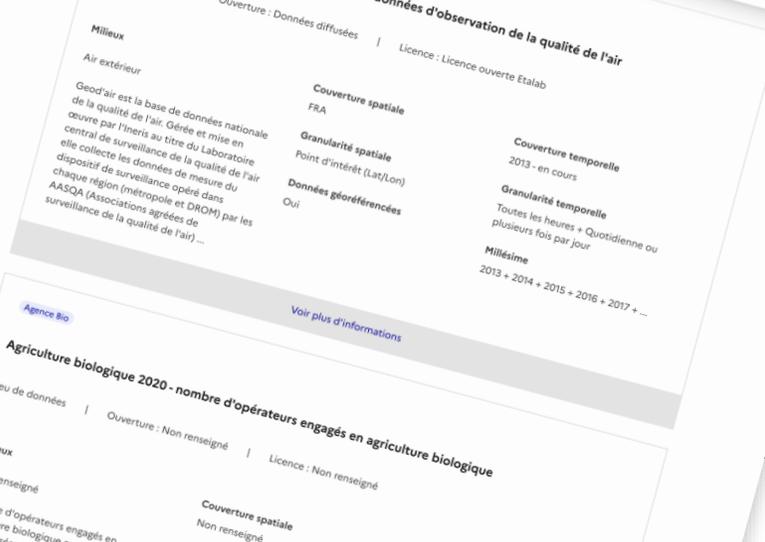
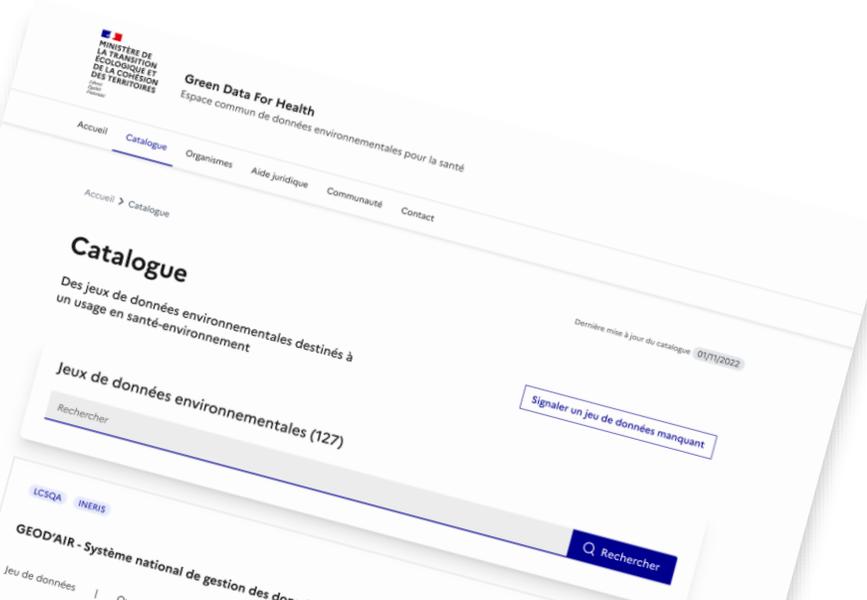
- ✓ **Communauté de 450 membres actifs.**
1^o édition de la communauté de la donnée s-e : **130 participants**.
2^o édition de la communauté de la donnée s-e : **200 participants**.
- ✓ **AAP : 4 projets de recherche** sélectionnés pour être accompagnés.
- ✓ **Challenge : environ 15 défis** cadrés pour une implémentation technique au printemps 2023.
- ✓ **Catalogue de données : 130 ressources** référencées.
- ✓ **Plateforme en ligne : <https://gd4h.ecologie.gouv.fr/>**
Tous onglets confondus, la plateforme en ligne a fait l'objet, en 4 mois, de : **12 000 consultations par 2 600 visiteurs différents**.
- ✓ **Ressourcerie : 20 fiches pédagogiques**.
- ✓ **Outil d'autodiagnostic : 1 commun numérique** en cours de cadrage.

Ecosystème du Green Data for Health



*Acteurs impliqués dans les instances de pilotage du projet

Le Green Data for Health : Une plateforme en ligne accessible au grand public



DÉMO : navigation sur la
plateforme en ligne
<https://gd4h.ecologie.gouv.fr/>

Catalogue

Modélisation des données descriptives du catalogue GD4H

La modélisation des données consiste à définir les champs qui décrivent une ressource - JDD, BDD, portail - avec des métadonnées descriptives qui :
_ doivent être utiles et pertinentes (**qualification**) afin d'être facilement identifiées (recherche et filtre) et utilisées (**accessibilité**)
_ doivent suivre des référentiels pour s'assurer de l'uniformité et la conformité des valeurs et de leur croisement entre elles et avec d'autres portails (**interopérabilité**)

127 jeux/bases/portails de données environnementales, 54 producteurs/référents publics de données environnementales

MÉTADONNÉES GÉNÉRALES

Titre de la ressource	Sigle de la ressource	Organisme(s) producteur(s)	Type de ressource	Description de la ressource
-----------------------	-----------------------	----------------------------	-------------------	-----------------------------

MÉTADONNÉES JURIDIQUES ET TECHNIQUES

Url d'accès à la ressource	Url de téléchargement de la ressource	Point de contact	La ressource est-elle accessible ?	Tarifcation	Nom de la licence	Format de la ressource
----------------------------	---------------------------------------	------------------	------------------------------------	-------------	-------------------	------------------------

MÉTADONNÉES TEMPORELLES ET GÉOGRAPHIQUES

Couverture temporelle	Millesime	Granularité temporelle	Fréquence de publication	La collecte est-elle terminée ?	La ressource est elle géoréférencée ?	Couverture spatiale	Granularité spatiale
-----------------------	-----------	------------------------	--------------------------	---------------------------------	---------------------------------------	---------------------	----------------------

MÉTADONNÉES SANTÉ ENVIRONNEMENT

Milieu	Voie d'exposition	Recherche, utilisation ou impact en SE	Type d'agents présents dans l'environnement	Nom de l'agent présent dans l'environnement	Jeux de données associés	Réutilisation associée
--------	-------------------	----------------------------------------	---------------------------------------------	---------------------------------------------	--------------------------	------------------------

Aide juridique

Objectifs et vue globale

OBJECTIFS

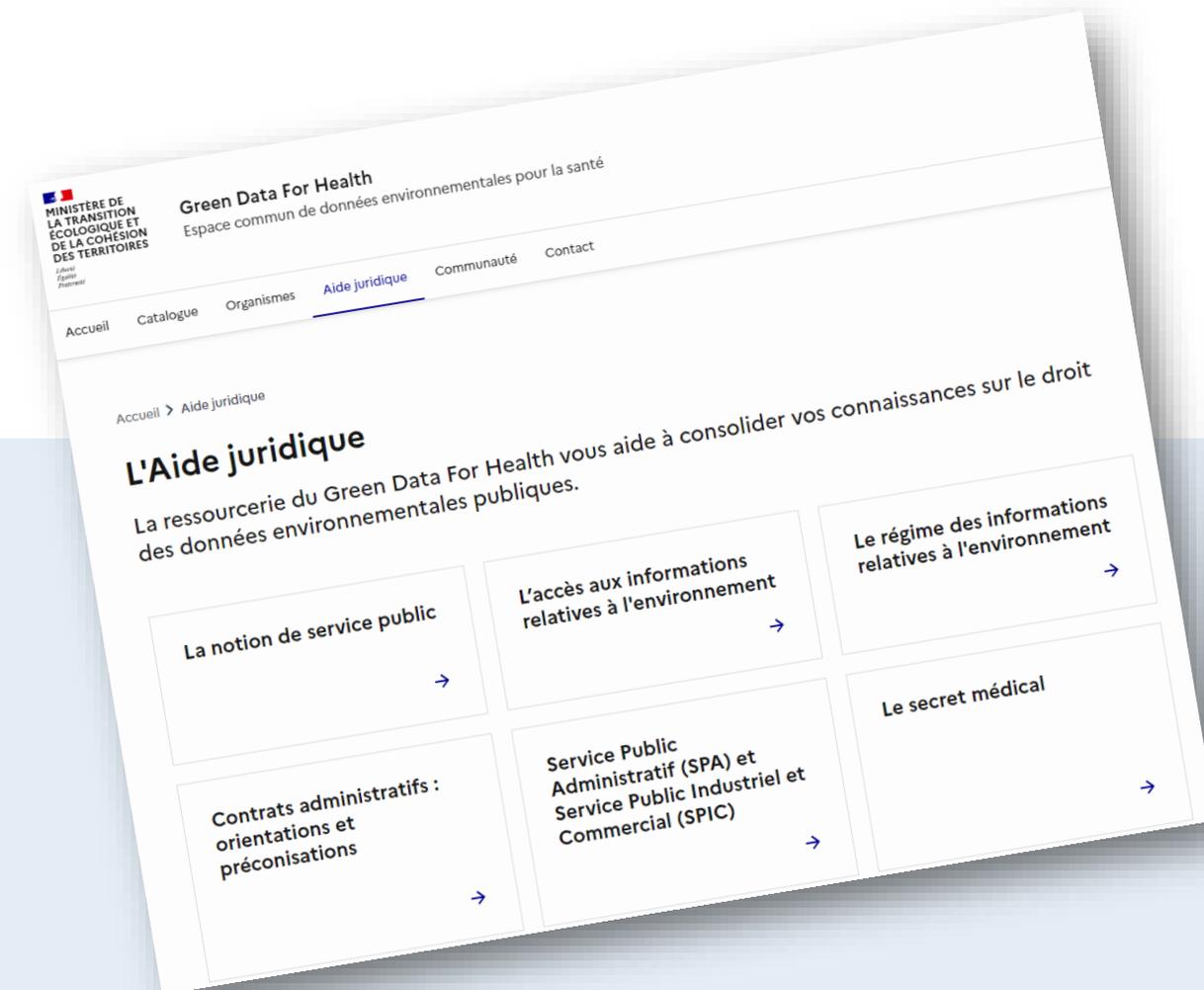
- _ Améliorer l'accessibilité des BDDs environnementales par les chercheurs et leur partage par les producteurs
- _ Permettre une montée en connaissance de la communauté sur les aspects juridiques liés à la donnée env.
- _ Construire un commun réutilisable pour d'autres projets du MTE et d'autres organismes publics et ministère

RESSOURCERIE

- _ **20 fiches juridiques** accessibles en ligne sur la plateforme

OUTIL JURIDIQUE AUTO-DIAGNOSTIC

- _ Outil à destination des producteurs et gestionnaires de bases de données : cadrage du besoin, maquettage, et détermination du contenu juridique.
- _ A venir : cadrage comme un commun numérique et implémentation technique.



Animation

Un socle : la communauté de la donnée en santé-environnement



QUELS ACTEURS ?

- _ **Chercheurs, experts et décideurs** en santé-environnement
- _ **Responsables de données, gestionnaires, financeurs, référents** de bases de données de santé ou environnementales pertinentes pour la santé-environnement



QUELS OBJECTIFS ?

- _ **Susciter des collaborations** en intégrant les besoins et priorités de l'écosystème.
- _ Mettre en place des **projets concrets autour d'objectifs partagés**
- _ **Faire monter en compétences** la communauté sur des sujets d'intérêt conjoints

Animation

Appel à projets conjoint avec le Health Data Hub : soutenir la recherche et l'innovation en santé environnement par la donnée.

L'INITIATIVE

Soutenir l'innovation par la donnée en levant les freins à la mobilisation de la donnée environnementale et sanitaire afin de stimuler la recherche dans le domaine de la santé-environnement :

- ↳ **Sélection de projets de recherche sur des thématiques à fort impact de santé publique** (bruit, perturbateurs endocriniens, bassins industriels).
- ↳ **Accompagnement (humain et financier) des projets lauréats** pour favoriser les synergies entre données environnementales et de santé.

RÉSULTATS

Sélection de **4 projets lauréats**.

* **perturbateurs endocriniens :**

MRC-PE porté par l'université de Lille et qui vise à étudier l'association entre le profil environnemental des territoires et l'incidence de l'insuffisance rénale chronique terminale (IRCT) à l'échelle nationale.

* **2° projet sur les bassins industriels :**

BIS, porté par Santé Publique France et l'Ineris qui visera à déterminer l'association entre la proximité de grands bassins industriels et l'état de santé de la population riveraine. Une étude sera également menée pour déterminer les corrélations géographiques entre les caractéristiques des communes et l'asthme de l'enfant.

* **projets sur le bruit :**

SOMNIBRUIT porté par Bruitparif, l'ORS Ile de France et le Centre du Sommeil et de la Vigilance de l'Hôtel Dieu. Le projet a pour objectif de mieux connaître et quantifier les effets du bruit environnemental sur le sommeil en réalisant une étude écologique sur 10 millions d'habitants.

Noises porté par le CHU de Bordeaux et le Centre de lutte contre le cancer-institut Bergonié. L'objectif est de cartographier et de décrire la distribution spatiale des indicateurs de maladies cardio-neuro-vasculaires, socioéconomiques et environnementaux (pollution de l'air et bruit) pour les territoires de Bordeaux et de l'Euro Métropole de Strasbourg.

Animation

Le Challenge du Green Data for Health

CONTEXTE

La donnée environnementale reste peu mobilisée pour faire émerger des associations entre expositions environnementales et pathologies humaines.

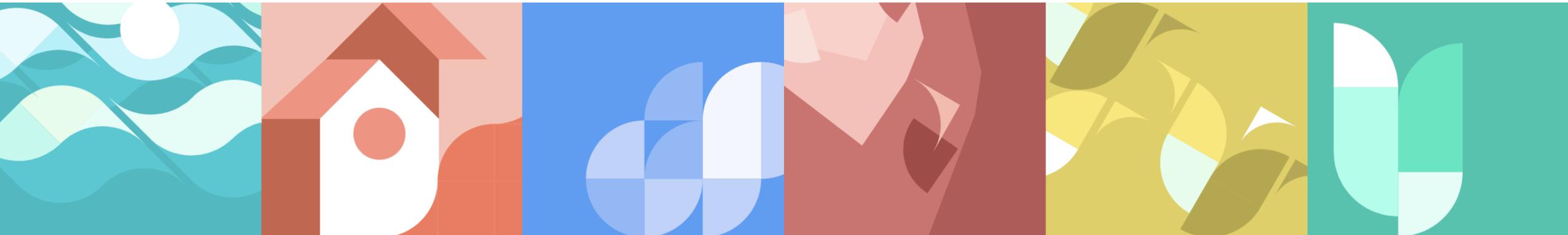
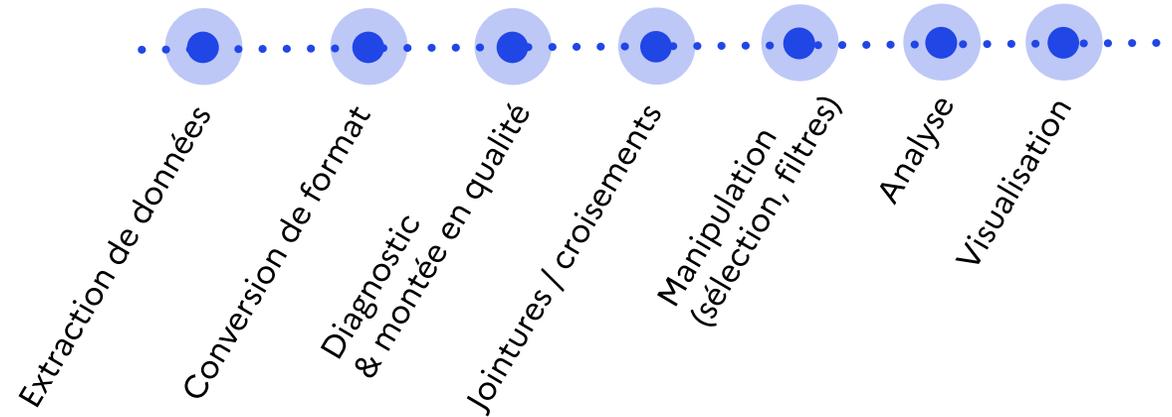
OBJECTIFS

Pour faciliter la mobilisation des données environnementales, développer des outils :

- **Utiles** : proposés puis portés par les acteurs de la santé-environnement.
- **Communs** : développés autour d'une communauté consolidée d'acteurs.
- **Transférables et pérennisables**

ACTIONS

Le développement d'une quinzaine de défis entre avril et juin 2023, et une mise en ligne des productions à l'issue du Challenge.



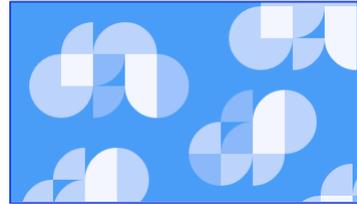
Animation

Les défis du Challenge du Green Data for Health



BIODIVERSITÉ

- ↳ Plateforme de visualisation des données de l'Ambroisie



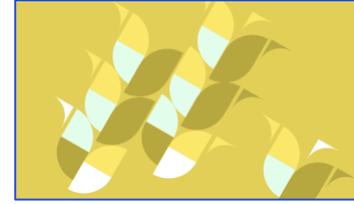
AIR

- ↳ Script de jointure des données de qualité d'air extérieur (QAE) et intérieur (QAI).



OUTIL TRANSVERSE

- ↳ Table de passage sur les IRIS
- ↳ Outil de diagnostic de qualité des données (dimension spatio-temporelle)
- ↳ Plateforme de visualisation sur les perturbateurs endocriniens.
- ↳ Base de diffusion des connaissances sur les produits et substances phytopharmaceutiques
- ↳ Plateforme de visualisation d'indicateurs de santé-environnement.



ALIMENTATION

- ↳ Plateforme de visualisation des données d'exposition aux substances chimiques dans l'alimentation de la population française.



SOLS

- ↳ Base de données exploitable sur la contamination des sols et végétaux.
- ↳ Base de données exploitable à partir des bulletins de santé des végétaux.



EAU

- ↳ Montée en qualité de données : facilitation de l'exploitation de l'historique d'une base de données sur la qualité de l'eau (SISE-Eaux)



AMENAGEMENT & Ufs*

- ↳ Outil de diagnostic territorial en santé : vagues de chaleur et soins primaires.
- ↳ Base de données de déterminants environnementaux de santé.
- ↳ Script de croisement de données sur les pratiques agricoles et l'environnement
- ↳ Base de données nationale des mesures de bruit normalisées.

*Urbanisme favorable à la santé

Besoins exprimés par les acteurs de la santé-environnement*

_ Développer l'articulation entre mobilisation de données et retours d'expériences/cas concrets

"Des ateliers visant à recueillir des retours d'expériences concrets autour de projets et de thématiques très ciblés sur des territoires délimités (échelles régionale, départementale, des EPCI, des communes, des quartiers)"

Ateliers

"Un rdv annuel sur le sujet avec des présentations d'exemples de projet ou études réalisées avec des données croisées ou des retours d'expériences d'actions réalisées, des échanges ou brainstorming pour faire émerger des idées."

Rendez-vous annuel

"Un catalogue de projets en santé environnement alimenté par les utilisateurs et modéré par le gd4h"

Catalogue de cas d'usage

_ Mieux identifier les acteurs et faciliter leur mise en relation

"Un canal de discussion direct avec les producteurs de données pour les chercheurs. Un sorte de forum amélioré."

Canal de discussion

"Schématiser les liens entre les acteurs en rappelant les missions de chacun"

Organigramme

PROPOSITION – Création d'un groupe de travail sur la santé-environnement

Objectifs – identifier les principaux usages et besoins correspondants pour mieux mobiliser les données environnementales et sanitaires en santé-environnement.

Etape 1

IDENTIFIER LES OUTILS EXISTANTS

Identification des référentiels, standards et données structurantes en santé-environnement.

SYNTHETISER LES BESOINS ET USAGES CLES POUR LES ACTEURS DE LA SANTE-ENVIRONNEMENT

Maille spatiale, maille temporelle...

Etape 2

IDENTIFIER LES MANQUES

Mise en évidence des manques en termes de standardisation (référentiels, standards, données structurantes)

PRIORISER LES MANQUES

Etape 3

IDENTIFIER DES CAS D'USAGE POUR IMPLEMENTATION

Mise en œuvre de travaux sur quelques cas d'usage spécifiques.

- _ Définition d'un cahier des charges.
- _ Définition d'une stratégie de déploiement et d'adoption.

Proposition d'organisation

ACTEURS

- _ Co-animation par un acteur de la santé et de l'environnement.
- _ Socle du groupe de travail : principaux organismes producteurs et ré-utilisateurs de données en santé-environnement.

CALENDRIER

- _ Une première réunion du GT d'ici l'été pour définir le calendrier et les livrables.

ANNEXES

Conseil plénier du Conseil National de l'Information Géolocalisée (CNIG)- 30 mai 2023

Annexe 1 : Enjeux juridiques à l'utilisation des données géolocalisées

Principe de finalité

Le responsable d'un fichier ne peut enregistrer et utiliser des informations sur des personnes physiques que **dans un but bien précis, légal et légitime**

Principe de proportionnalité et de pertinence

Les informations enregistrées doivent être **pertinentes et strictement nécessaires au regard de la finalité** du fichier

Principe d'une durée de conservation limitée

Il n'est pas possible de conserver des informations sur des personnes physiques dans un fichier pour une durée indéfinie. Une **durée de conservation précise** doit être fixée, en fonction du type d'information enregistrée et de la finalité du fichier

Principe de sécurité et de confidentialité

Le responsable du fichier doit **garantir la sécurité des informations qu'il détient**. Il doit en particulier veiller à ce que seules les personnes autorisées aient accès à ces informations

Pour les données géolocalisées de santé, il convient de faire attention aux informations que l'on peut obtenir suite au croisement de différentes bases de données (API adresse, données INSEE,...)

Annexe 2 : l'aide juridique au sein du GD4H

Focus : ressource de fiches juridiques

OBJECTIFS

Les fiches pédagogiques doivent permettre à un utilisateur sans connaissance juridique de :

- _ Mieux comprendre la nature juridique des données qu'il utilise
- _ Cerner le régime applicable à ces données
- _ Accéder rapidement à des informations juridiques (normes pertinentes, jurisprudences ou décisions d'autorités compétentes utiles...)

OFFRE

RESSOURCERIE <https://gd4h.ecologie.gouv.fr/juridique>

_ **20 fiches pédagogiques** vulgarisant les grands principes du droit des données en ligne notamment sur :

- La **qualification juridique des données** (document administratif, information relative à l'environnement...)
- Le **régime applicable à ces données** (régime d'open data, réutilisation des données publiques...)
- Des **notions juridiques transverses** (le secret statistique, les données à caractère personnel...)

The screenshot shows the 'Aide juridique' page of the Green Data For Health website. The page is titled 'L'Aide juridique' and includes a brief introduction: 'La ressource de Green Data For Health vous aide à consolider vos connaissances sur le droit des données environnementales publiques.' Below this, there is a grid of 20 clickable boxes, each containing a legal topic and a right-pointing arrow. The topics are: La notion de service public; Contrats administratifs : orientations et préconisations; L'accès aux informations relatives à l'environnement; Service Public Administratif (SPA) et Service Public Industriel et Commercial (SPIC); L'accès par dérogation aux documents protégés; Le secret médical; Le régime de réutilisation des documents administratifs; Le lien entre les données et la mission de service public; Le régime des informations relatives à l'environnement; Le secret statistique; Les personnes juridiques : personnes morales et personnes physiques; La notion de document administratif; Les données à caractère personnel; La protection des producteurs des bases de données par la propriété intellectuelle; La situation des chercheurs en droit des données; Le partage d'informations entre administrations; Le régime d'accès aux documents administratifs; Les licences; Les bases de données en droit; Le secret des affaires.

Annexe 2 : l'aide juridique au sein du GD4H

Focus : ressource de fiches juridiques > Exemple de fiche



Le secret statistique

! Infos pratiques

Les informations protégées par le secret statistique peuvent être diffusées

_ Au public, qu'après avoir été agrégées ou après les délais de communication des archives publiques atteints (75 ans ou 25 ans après le décès de la personne) ;

_ Aux chercheurs par anticipation, qu'après un avis du Comité du secret statistique ;

— Quelques éléments d'introduction

Le secret statistique est bien connu de tous les statisticiens, c'est une forme de secret professionnel qui s'applique à ceux qui sont chargés de recueillir et d'exploiter des statistiques publiques. Le rôle historique des statisticiens publics a été important, notamment au sortir de la seconde guerre mondiale. Dès lors, un statut particulier a vu le jour en France par la loi n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Décryptage d'un métier pivot et d'un secret professionnel particulier.

1. Définition du secret statistique

Le secret statistique entend garantir, d'une part, la **confidentialité de la vie personnelle et familiale pour les personnes physiques**, d'autre part la **protection du secret des affaires pour les entreprises** (cf. Fiche « Les données à caractère personnel »). Cette protection est également garantie au niveau européen. Le secret statistique est défini à l'article 6 de la loi de 1951 et prévoit que les renseignements obtenus par l'intermédiaire des questionnaires révélant d'une manière générale des faits et comportements de nature privée des personnes physiques et les ne peuvent être communiqués à des tiers, y compris à des fins de lutte contre la fraude fiscale. Ainsi le droit prévoit une protection particulière pour les personnes qui fournissent des informations détaillées lorsqu'elles participent à des enquêtes.

C'est ce qu'on appelle le critère de destination : **les informations ont été récoltées dans un cadre particulier, celui très codifié des études statistiques**, c'est ce qui les fait entrer dans les dispositions du secret statistique. Il convient

de souligner que l'article 6 a été complété en 1986 en permettant aux services statistiques d'enrichir les données d'enquête en obtenant un accès facilité aux données des autres administrations. Cet accès est réservé uniquement « à des fins statistiques ».

En pratique, le juge considère que le secret statistique fait partie des « autres secrets protégés par la loi » prévu à l'article L. 311-5 du CRPA. Ainsi l'administration devra étudier préalablement dans quel cadre les informations ont été collectées avant d'envisager de communiquer le document en occultant les mentions couvertes par le secret ou lorsque le demandeur a suivi une procédure particulière. En effet, le secret statistique s'observe surtout au moment de la diffusion des résultats issus de l'exploitation des informations recueillies dans le cadre d'enquête ou à des fins statistiques.

— Illustrations d'informations non-communicables

La communication du document mentionnant le prix de l'eau commune par commune est refusée dès lors que les prix ont été obtenus par l'intermédiaire d'une enquête statistique (Avis 20010545).
Idem pour les chiffres de production d'un site éolien établis par une structure du groupe EDF Énergies Nouvelles dans le cadre d'une enquête annuelle approuvée par un arrêté du ministre. (Avis 20150102).

2. Accès aux données protégées par le secret statistique

L'accès aux données issues d'enquêtes et d'études statistiques connaît différentes variantes.

Évidemment les résultats des enquêtes statistiques ont vocation à être diffusés et connus. Ainsi à l'appui des analyses, des données peuvent être publiées. Ces données sont agrégées selon des règles strictes autour des unités statistiques retenues (la maille) puis celle du poids des individus dominants (aucune information ne sera communiquée si une personne est surreprésentée, plus de 85% du total pour les impôts).

Le Guide du secret statistique proposé par l'INSEE décrit les pratiques nécessaires pour respecter le secret statistique. En substance après retraitement, aucune donnée publiée ne doit concerner moins de 3 entités à la fois, aucune donnée publiée ne doit concerner une seule entreprise pour plus de 85 % du total et donc aucune donnée publiée ne doit permettre l'identification directe ou indirecte des personnes. C'est finalement une application du principe général qui prévoit que l'administration anonymise les informations protégées avant toute communication.

Autrement les informations individuelles collectées dans le cadre d'études sont des archives publiques, elles suivent donc le même régime. Dès lors, les informations figurant dans des questionnaires peuvent faire l'objet d'une communication sans anonymisation après un délai de 75 ans après la date de réalisation de l'enquête ou de 25 ans après le décès de la personne concernée.

Par exemple voir le Conseil 20152823 du 30/07/2015 favorable à la communication libre des archives de l'occupation française en Allemagne et en Autriche relative à la dénazification.

Enfin le droit (la loi notamment) peut prévoir des exceptions permettant de communiquer les informations protégées, par exemple en matière fiscale ou bien dans certains cadres précis. Ainsi depuis 2016, il existe une procédure particulière permettant aux chercheurs de demander l'accès à des bases de données protégées, dès lors que leurs travaux présentent un intérêt public.

L'administration sollicitée pour fournir un accès aux données peut saisir le Comité du secret statistique qui formulera un avis sur la demande et pourra recommander un mode d'accès sécurisé aux données. Il convient de souligner que cette procédure nécessite pour le demandeur de présenter un projet de recherche.

— Illustrations

Le Comité du secret statistique formule des avis sur de nombreux travaux menés par des administrations et des laboratoires de recherches. Chacun des avis présente le porteur, la thématique et les bases de données nécessaires à l'étude. Comme par exemple une étude sur l'impact du Crédit impôt recherche sur la R&D et l'innovation des entreprises françaises, menée par l'Institut des politiques publiques jusqu'en avril 2023 qui nécessite d'accéder à une vingtaine de bases de données d'enquêtes.

En parallèle de la réglementation française, existent des modalités d'accès aux données collectées par l'Union Européenne à des fins scientifiques. C'est alors le règlement n° 831/2002 qui s'applique pour accéder aux informations protégées par le secret statistique disponibles au niveau de l'Union européenne. Il prévoit une procédure similaire à celle qui existe en France.

3. Conclusion

Le secret statistique est donc un secret professionnel qui peut tout à fait s'opposer à la communication d'information produites ou reçues lorsqu'elles sont un matériau utilisé dans le cadre des études statistiques. Pour autant, cela ne signifie pas que toutes les données et informations manipulées par les services statistiques ministériels sont automatiquement protégées. Ces services peuvent mettre leur savoir-faire en matière d'exploitation et d'analyse de données à disposition d'autres administrations lorsqu'ils interviennent dans un autre cadre que celui des enquêtes.

Annexe 2 : l'aide juridique au sein du GD4H

Focus : outil juridique d'autodiagnostic (OJAD)

OBJECTIF

Développement de l'outil en le pensant comme un **commun numérique**.

Exigence de **réplicabilité** et de **transférabilité** à d'autres cas d'usage.

- Premier champ d'expérimentation en santé-environnement (producteurs et gestionnaires de bases de données).
- Extensions futures à d'autres utilisateurs et cas d'usage.

EXTENSIONS POSSIBLES

_ Propriété intellectuelle : module dédié (fil de questions pour répondre aux enjeux)

_ Duplication de la méthode côté réutilisateurs

