

Groupe de travail « Données géolocalisées en santé »

Commission « Besoins et usages » du CNIG

Compte rendu de la 1^o réunion en plénière

18 juillet 2023 16h00 – 18h00 (visioconférence)

Ordre du Jour

16h00 – 16h30 / Présentation du mandat et recueil des commentaires généraux.

16h30 – 17h15 / Séquence 1 : consolidation du plan de travail sur le livrable 1.

17h15 – 17h30 / Séquence 2 : consolidation du plan de travail sur le livrable 2.

17h30 – 18h00 / Séquence 3 : consolidation du plan de travail sur le livrable 3.

Participants :

Hubert Holin (DGEC/BQA), Jérémie Letonnelier (Cerema), Benjamin Levy Maya (CGDD/Ecolab), Alexandre DESCAMPS (DGS/VSS), Thomas Mettey (IGN), Matthieu Brient (OpenDataFrance), Christelle Diodat (ARS, projet Atlasanté), Benoit ROCQ (Atmo Hauts de France), Gwendal LIBESSART (Cerema), Marie Ramon-Daré (CGDD/Ecolab/GD4H), Adeline MARTIN (CNES-MEDES), Pierre Laulier (CGDD/Ecolab/CNIG), Benoit David (CGDD/Ecolab), Sébastien LESTERLE (DGS), Mathilde Husson, interne en médecine (DGS/VSS2), Arnaud Bonnard (DREES), Jacques Patris (Atmo France), Perrine de Crouy-Chanel (Santé publique France), Marie GOMBERT (IGN), Paul Grignon (CGDD/ Ecolab/GD4H), Fei GAO (Health Data Hub), Sabine Host (ORS Ile-de-France), Laurent Létinois (Ineris), Frédéric Minodier (INSEE), Charlotte Lepitre (Atmo France), Damien MALTETE (Cerema), Corinne Mandin (IRSN), Anna Lungarska (INRAE), Julie Chataigner (OFB), Camille Renoux (DGS), Grégoire Rey (France Cohortes),

Marianne Billard (DNS), Soizic Urban-Boudjelab (DGS/VSS2), Elodie Plancque (Altasanté).

1. Présentation du mandat et recueil des commentaires généraux

Proposition d'un travail en 3 axes :

- **Livrable 1** : Investigation des possibilités (opportunités et contraintes) de géolocalisation des données de santé.
- **Livrable 2** : Collecte et synthèse des besoins et usages clés pour les acteurs de la santé-environnement.
- **Livrable 3** : Identification des ressources structurantes pour mobiliser la donnée en santé-environnement (bases de données structurantes, nomenclatures, schémas de données, autres standards et ressources de référence pertinents).

Pour l'ensemble des axes, les contributions ont été effectuées sur le support MIRO ci-après : https://miro.com/app/board/uXjVM11c0Uc=

2. Séquence 1 : Consolidation du plan de travail sur le Livrable 1

4 axes de travail ont successivement été complétés sur le support MIRO.

2.1. Les apports et opportunités de la géolocalisation des données en santé

Différents types d'apports ont été relevés concernant la géolocalisation des données en santé.

Le croisement entre données de santé et données environnementales peut permettre de mettre en regard **des données de santé avec des données caractérisant un territoire**. La géolocalisation permet de croiser les données de santé avec les données environnementales, facilitant ainsi des études fines. L'introduction de la dimension spatialisée peut en définitive permettre :

- De faire le **lien entre plusieurs types de données d'exposition (approche multi-expositions)** recueillies sur une même zone géographique. Il peut également être bénéfique de combiner différentes sources et modalités de collecte de données pour une meilleure compréhension des enjeux (données d'études de terrain, données citoyennes...).
- L'**identification des déterminants de santé** grâce à la cartographie des données de santé, à l'établissement d'analyses corrélationnelles et causales.
- La **mise en place d'indicateurs** : l'élaboration des indicateurs de santé pertinents.

- **La réalisation d'études d'impact** environnement/santé sur la base d'analyses causales. Les données géolocalisées permettent de mettre en évidence l'« effet territoire » sur la santé.

Il a également été souligné qu'il peut également être bénéfique de combiner différentes sources et modalités de collecte de données pour une meilleure compréhension des enjeux (données d'études de terrain, données citoyennes...).

In fine, les travaux facilités ou permis par une meilleure interopérabilité entre données environnementales et sanitaires peuvent ainsi être mis au service des apports suivants :

- **Améliorer la connaissance des territoires sur la santé-environnement (diagnostics de territoires) pour les décideurs publics et pour les professionnels de santé.** Les diagnostics territoriaux et portraits de territoires peuvent constituer une base de connaissances pour orienter les diagnostics et les décisions pour les acteurs impliqués en santé-environnement. De tels outils pourraient ainsi aider les professionnels de santé dans leur diagnostic.
- **Améliorer les parcours de soin.** La géolocalisation des données peut permettre de personnaliser l'offre et le suivi de la santé en fonction de la localisation, améliorant ainsi le parcours du patient. L'identification des bassins de patientèle associés aux variables environnementales identifiées peut ainsi être un levier pour des actions ultérieures.
- **Adapter les politiques de santé aux territoires** (e.g. politiques environnementales et agricoles). Cela peut permettre d'identifier et de prioriser les enjeux de santé sur un territoire donné et à une échelle donnée, et de mettre ainsi en place des politiques territoriales d'offre de soins adéquates.
- **Effectuer des analyses de risques et gérer des situations de crise** (e.g. gestion d'un épisode de sécheresse).
- **Etudier et réduire les inégalités territoriales de santé.** Cela peut permettre d'analyser les différentiels socio-économiques et d'exposition environnementale, et d'identifier notamment des zones critiques. Les inégalités sociales et environnementales, et leurs retranscriptions en termes sanitaires, peuvent ainsi être étudiées en croisant les différentes sources de données.
- **Sensibiliser et impliquer les citoyens :** La géolocalisation des données de santé peut permettre de mieux sensibiliser les citoyens aux effets de l'environnement sur leur santé.

2.2. Les freins et limites rencontrés concernant la mobilisation des données géolocalisées de santé, notamment aux service de la santé environnement

- **Repérabilité et diversité des sources de données** : les données proviennent d'une diversité de sources. Il peut parfois être difficile de savoir où sont les données disponibles.
- **Accès aux données** : l'accès aux données géolocalisées peut être difficile en raison de leur dispersion, de l'absence de centralisation et de la multiplicité des producteurs ayant des standards différents. Certaines données financées par des fonds publics ne sont pas systématiquement ouvertes en open data, limitant ainsi leur utilisation dans la recherche et les politiques publiques. L'identification des acteurs et l'hétérogénéité des standards adoptés par la multiplicité des producteurs ont également été mentionnées.
- **Sensibilité et sécurité des données** : La géolocalisation des données de santé soulève des préoccupations concernant la protection des données à caractère personnel, et plus généralement des enjeux de sensibilité des données. La confidentialité des données et l'acceptabilité sociale de la géolocalisation des données sont donc importants à considérer. Le Règlement général sur la protection des données (RGPD) encadre le traitement de ces informations, mais leur sécurité reste une préoccupation.
- **Géocodage : qualité, documentation et mobilisation de l'information géocodée** :
Un besoin de documentation du géocodage. Le niveau de précision de la géolocalisation impacte la capacité de croisement des données. Cependant, il peut être difficile de disposer de métadonnées complètes ou d'une documentation explicative sur certaines données, rendant complexe l'analyse et la compréhension des résultats.
La qualité du géocodage : La qualité des données géolocalisées est essentielle pour des analyses pertinentes. Cependant, il peut y avoir des problèmes d'adresses non géocodées, de géocodage abandonné ou de manque de circuit de correction pour améliorer la qualité des informations de localisation. La stabilité dans le temps des géocodages (comme les codes communes) peut poser des défis pour le suivi à long terme. Aussi, la diversité des outils de restitution des données peut également compliquer la visualisation et l'analyse.
Le besoin d'outils et de frameworks dédiés pour accélérer les analyses (e.g. PostGis).

- **Hétérogénéité des données** : Les données géolocalisées proviennent de diverses sources, ce qui peut entraîner des problèmes de temporalité, de granularité géographique et de comparabilité des données. Parmi les différents problèmes d'hétérogénéité induits, a notamment été mentionné le fait que la qualité des adresses peut varier selon les territoires, ce qui peut biaiser les analyses (en particulier entre territoires ruraux et urbains ou dans les territoires d'outre-mer).
- **Manque de standards** :
Référentiels communs pour certaines données. L'absence d'un référentiel commun pour certaines données géolocalisées et le manque de référentiels sur certains champs d'informations (e.g. la surface agricole utile), peuvent compliquer l'analyse et l'utilisation des données.
Méthodes de construction d'indicateurs partagées. Le besoin d'indicateurs de santé spécifiés a été mentionné.

2.3. Les leviers, opportunités et solutions à envisager pour favoriser la mobilisation des données de santé géolocalisées, notamment au service de la santé-environnement

- **Mettre en place des initiatives d'animation/d'accompagnement comme des appels à projet** : Lancer des appels à projets spécifiques axés sur l'utilisation des données de santé géolocalisées dans le domaine de la santé-environnement. L'initiative du Space Climate Observatory (SCO) a notamment été mentionné.
- **Produire de nouvelles données** en s'appuyant sur la donnée citoyenne et sur le développement de l'IOT et des objets connectés.
- **Mieux repérer données environnementales et sanitaires**. Il a été mentionné le fait de promouvoir un catalogue sur les données de santé, de poursuivre l'alimentation de catalogues sources permettant de montrer l'existence des données, de les documenter, et de les articuler avec des réutilisations au travers d'une porte d'entrée visible et unique. Il a également été mentionné le fait d'améliorer la documentation des données de santé sous l'aspect de la maille géographique mobilisable (finesse, qualité). Le fait de faire le lien entre données et réutilisations a en particulier été mentionné pour les grandes cohortes de patients.
- **Partager des méthodes et outils pour exploiter et visualiser les données géolocalisées** : favoriser le partage des méthodes et des outils d'exploitation de données géolocalisées, et d'outils de visualisation à différentes échelles.

- **Enrichir la qualité de la géolocalisation des données** : Améliorer les informations de localisation dans les bases de données médico-administratives pour accroître la finesse géographique des données. Enrichir les systèmes d'information en santé avec des données géolocalisées plus précises, particulièrement en milieu urbain où cela n'affecte pas la confidentialité.
- **Mettre à disposition les jeux de données sous un format géographique** permettant une remobilisation aisée (e.g. PostGis).
- **Appliquer les principes FAIR** (découverte, accès, interopérabilité, réutilisation) pour faciliter le partage et la réutilisation des données géolocalisées.
- **Explorer différents leviers juridiques pour favoriser l'ouverture et l'interopérabilité des données** : Établir des règles et des clauses spécifiques dans les marchés publics pour garantir l'ouverture et l'interopérabilité des données géolocalisées, en encourageant l'utilisation de standards.
- **Explorer différents leviers juridiques pour gérer au mieux les enjeux de sensibilité des données** : renforcer le consentement des personnes concernant la géolocalisation de leurs données, renforcer les contrôles sur la sécurisation de ces informations, utiliser des leviers tels que l'anonymisation et la pseudonymisation.
- **Mettre en place des standards : référentiels et indicateurs de santé** : Standards pour mieux mobiliser la donnée. Mutualiser et réutiliser les référentiels existants autant que possible pour éviter de créer de nouvelles données par défaut. Mettre en place un référentiel descriptif des données et des métadonnées associées pour faciliter leur utilisation.
Indicateurs de santé. Créer des indicateurs de santé.
- **Mettre en place des géostandards** : adopter des standards (ex. COVADIS) et des technologies ouvertes pour assurer l'interopérabilité des données géolocalisées.

Le fait d'appréhender cette question de l'identification des solutions en explorant les autres solutions explorées à l'échelle européenne a également été évoquée.

Tout en envisageant des leviers permettant de mieux appréhender le caractère géolocalisé de la donnée de santé, l'enjeu est par ailleurs de sensibiliser aux enjeux de cette utilisation et de garantir que celle-ci soit justifiée (nécessaire) et sécurisée lorsque des enjeux de sensibilité viennent à se poser.

2.4. Les grands axes à développer pour le livrable 1

Les grands axes à développer pour le premier livrable ont fait l'objet d'un premier travail de collecte des idées durant cette première réunion du GT. Des mots clés ont ainsi été recueillis sur les différents axes proposés, soit sur :

- Le **référencement des ressources** (nature de la donnée de santé et maille géographique disponible)
- La **mobilisation souhaitée** pour celles-ci (objectif/usage, type de traitement, maille géographique nécessaire)
- Les **contraintes et freins** de mobilisation rencontrés (juridiques, techniques, organisationnels)
- Les **leviers** associés.

Les éléments recueillis figurent sur le support MIRO et alimenteront le travail de définition de structure pour le livrable 1 initié par les animateurs du groupe de travail.

En complément, plusieurs axes complémentaires à développer ont été proposés :

- Effectuer un état des lieux des travaux et cas d'usages.
- Prendre en compte les enjeux scientifiques.
- Effectuer un état de l'art des SI et référentiels existants.
- Introduire un glossaire, une partie présentant une définition des mots clés (e.g. notions de donnée agrégée, de donnée sensible).
- Introduire un regard sur les pratiques internationales.

Des acteurs souhaitant se mobiliser sur ce livrable et des suggestions de contact supplémentaires ont été indiqués sur le support MIRO lors de la réunion.

3. Séquence 2 : Consolidation du plan de travail sur le livrable 2

Différents axes ont été proposés pour structurer la collecte des besoins et des usages clés :

- Des questions portant sur les **problèmes de mobilisation des bases de données environnementales et sanitaires en santé-environnement** (questions de recherche non-résolues du fait de freins rencontrés se rapportant aux données, croisements de données problématiques...).

Quelques axes ont déjà été proposés :

- La **diversité de référentiels** de données géolocalisées,
- La **problématique de repérabilité/connaissance des données disponibles** (notamment données environnementales)
- La **problématique d'accès** (notamment problématiques de sensibilité de la donnée).

- Plusieurs **problématiques relatives au traitement et à la manipulation** des données (gestion des problèmes d'hétérogénéité, expertise métier spécifique à chaque jeu de données, gestion des petits effectifs et méthodes de lissage).
- Des questions se rapportant aux **outils utiles et aux besoins d'outils** pour mieux mobiliser les données environnementales et sanitaires.

Apports proposés pour cette dimension :

- Il pourrait être opportun de préciser la question en mentionnant les **types de traitement** (dont le pas de temps et la maille spatiale attendue),
- Il pourrait être opportun de préciser la question en mentionnant les **acteurs effectuant les traitements**. En complément, il pourrait être intéressant de questionner la possibilité pour les équipes data en place de traiter l'information géographique (compétence à disposition en interne ou non).
- Il pourrait être opportun de préciser la question en mentionnant des **exemples d'outils** pour chaque dimension :
 - Pour repérer la donnée : catalogues (en particulier, les catalogues de données environnementales et sanitaires qu'il pourrait être intéressant de mobiliser en lien avec des données de santé).
 - Pour accéder à la donnée : questionner les modalités d'accès aux données (API, manuel...).
 - Pour croiser et traiter les données (scripts de de croisement et de traitement de données géolocalisées, SIG, outils d'analyse spatiale dans R, cartouche spatiale dans base PostGRE, FME, etc).
- Il pourrait être opportun de questionner la **volumétrie et les contraintes de ressources de traitement**.
- Des questions portant sur l'**état des lieux des standards et besoins associés** (bases de données structurantes, nomenclatures, schémas de données notamment).

Précisions et dimensions proposées dans le cadre de la consultation :

notions de tables de correspondance/tables de passage, distinction entre bases de données structurantes et jeux de données de référence, notion de dictionnaire de données, format de consommation des données, scénarios d'échange de données.

Des axes de questionnement complémentaires ont été suggérés :

- Ajouter une question complémentaire sur l'utilisation des données de santé et la maille géographique nécessaire : données de santé produites et mobilisées par l'acteur, information/précision géographique supplémentaire nécessaire et usages associés.
- Profiter de la consultation pour faire émerger de nouveaux cas d'usage (e.g. sans contrainte de mobilisation de données, quel serait le cas d'usage rêvé).
- Intégrer les enjeux de qualité des données au sein de la consultation (e.g. ajout d'un item sur le sujet).

Des acteurs souhaitant se mobiliser sur ce livrable et des suggestions de contact supplémentaires ont été indiqués sur le support MIRO lors de la réunion.

4. Séquence 3 : Consolidation du plan de travail sur le livrable 3

La première réunion du groupe de travail a permis de converger sur la thématique et le périmètre à envisager pour les différents sous-groupes à constituer pour identifier par thématique les ressources structurantes/standards en termes de mobilisation de la donnée environnementale et sanitaire.

Les 6 sous-groupes identifiés sont les suivants :

4.1. Thématique Eaux

Deux sous-thématiques :

- Qualité de l'eau (à envisager au sens large, avec une déclinaison eau potable, eau de baignade...). Cette dimension pourra notamment inclure les problématiques de contrôle sanitaire des eaux destinées à la consommation humaine et des eaux de baignade.
- Stress hydrique (accès à l'eau potable, prélèvements et usages, milieux humides).

Précisions sur le périmètre : Prise en compte des différentes eaux (potable, de baignade, de mer), notions de qualité et problèmes liées aux milieux humides.

Acteurs concernés/suggérés : IGN et OFB (acteurs impliqués dans BD TOPAGE, Système d'Information sur l'Eau), Eau France, ARS, ORS, CNES (SWOT¹), DGS (EA/EA4), DEB, Office International de l'Eau (OiEau).

4.2. Thématique Sols

¹ Mission spatiale étudiant l'eau à la surface de la Terre.

Deux sous-thématiques :

- Qualité et contamination des sols
- Occupation des sols (présence de végétation, caractérisation de la végétation – e.g. vignes et vergers, terres arables, prairies et friches agricoles).

Précisions sur le périmètre : attention aux recoupements entre les thématiques Eaux et Sols.

Acteurs concernés/suggérés : INRAE, INERIS, BRGM, Gis Sol, IGN.

4.3. Thématique Bruit

Acteurs concernés/suggérés : collectivités locales, Cerema, Bruitparif, acteurs des projets CENSE et NoiseCapture.

4.4. Thématique Rayonnements ionisants et non-ionisants :

Acteurs concernés/suggérés : ANFR, IRSN.

4.5. Thématique Qualité de l'air

Deux sous-thématiques :

- Qualité d'air intérieur
- Qualité d'air extérieur

Précisions sur le périmètre : Les pollens et allergies pourront être considérés dans le cadre de ce sous-groupe.

Acteurs concernés et suggérés : DGPR, LCSQA, INERIS, Atmo France, AASQAs, OQAI, Météo-France, SpF (Programme Air et Santé).

4.6. Thématique Risques alimentaires :

Précisions sur le périmètre : attention aux recoupements avec la thématique Eaux.

Acteurs concernés/suggérés : Anses, DGEC/BQA.

AUTRES PRECISIONS TRANSVERSALES

Dimensions transverses à prendre en compte dans chaque sous-groupe :

- Prisme « Changement et les risques climatiques ». Prendre en compte au sein de chaque sous-groupe les standards utiles à considérer pour apprécier les risques climatiques et les effets du changement climatique sur la thématique.
- Prisme « Facteurs socio-économiques ». Prendre en compte au sein de chaque sous-groupe les standards utiles à considérer pour

apprécier la vulnérabilité socio-environnementale/les inégalités socio-environnementales de santé sur la thématique.

Autre précision transversale en termes de périmètre : Rester plutôt sur une donnée chronique plutôt qu'accidentelle dans un premier temps.