

# GREEN DATA FOR HEALTH : UN ESPACE COMMUN DE DONNÉES ENVIRONNEMENTALES AU SERVICE DE LA SANTÉ-ENVIRONNEMENT

**Marie Ramon-Daré (1),  
Paul Grignon (2)**

1. Cheffe de projet design  
et innovation/Green Data  
for Health, Ecolab du  
Commissariat général au  
développement durable

[marie.ramon-dare@developpement-durable.gouv.fr](mailto:marie.ramon-dare@developpement-durable.gouv.fr)

2. Chef de projet DATA-IA/  
Green Data for Health,  
Ecolab du Commissariat  
général au développement  
durable

[paul.grignon@developpement-durable.gouv.fr](mailto:paul.grignon@developpement-durable.gouv.fr)

**Le Green Data for Health, action 18 du 4<sup>e</sup> plan national santé-environnement (PNSE4), est un dispositif actuellement incubé à l'Ecolab, le laboratoire d'innovation au service de la transition écologique, des ministères de la Transition écologique et de la Cohésion des territoires ainsi que de la Transition énergétique. Il vise à contribuer à l'enjeu de santé publique consistant à accélérer la production de connaissances quant à l'impact des facteurs environnementaux sur la santé. Pour ce faire, son objectif est de favoriser la mobilisation des données environnementales, par les acteurs de la recherche et de l'expertise, pour un usage en santé environnement. L'offre de service du dispositif s'articule autour de trois grandes briques : un catalogue en ligne de ressources environnementales utiles et décrites pour un usage en santé-environnement, des outils juridiques permettant de vulgariser les grands principes des données environnementales et d'accompagner les producteurs dans leur démarche d'ouverture de leurs données ; et enfin des actions d'animation de la communauté de la donnée en santé-environnement telles que la conduite de cas d'usage, l'organisation d'un Data Challenge ou encore l'ouverture d'un appel à projets conjoint avec le Health Data Hub.**

### L'environnement, un déterminant majeur de notre santé

Selon l'Organisation mondiale de la santé, les facteurs environnementaux et comportementaux sont responsables d'au moins 15 % des décès en Europe, soit environ 1,4 million de décès par an, pour la plupart évitables. En effet, la pollution de l'air, de l'eau, des sols, les pesticides, les perturbateurs endocriniens, etc. sont à l'origine de multiples pathologies dont les cancers, pathologies respiratoires, allergies, asthmes, maladies cardiovasculaires, diabète, obésité.

Ainsi, les attentes citoyennes sur les questions de santé-environnement sont de plus en plus fortes. En effet, le citoyen souhaite que l'impact sanitaire des facteurs environnementaux soit évalué et anticipé. Face à ces enjeux, le 4<sup>e</sup> plan national santé-environnement (PNSE4) publié en 2021<sup>1</sup> propose des actions concrètes pour mieux comprendre et réduire les risques sanitaires liés à l'environnement. Il y a donc un réel enjeu à utiliser les données existantes, nombreuses sur le champ environnemental mais peu mobilisées, pour établir un lien de causalité entre santé et environnement afin de mieux connaître et qualifier les expositions et les effets de l'environnement sur la santé des populations et des écosystèmes. Cette démarche de facilitation de la mobilisation des données existantes avait été initiée dans les précédents plans nationaux. L'action du Green Data for Health (GD4H) vise alors à passer à l'échelle les travaux réalisés dans les PNSE antérieurs pour pérenniser la démarche.

Par ailleurs, le nouveau plan écologique du gouvernement « France nation verte », publié le 21 octobre dernier, a pour objectif de doubler la baisse des émissions de gaz à effet de serre pendant le quinquennat. Cette planification écologique est constituée de sept enjeux transverses sur lesquels elle devra s'appuyer, dont les données environnementales, ainsi que de vingt-deux chantiers thématiques, parmi lesquels « la réduction des pollutions ayant un impact sur la santé », étroitement liés aux enjeux de santé-environnement.

### Le Green Data for Health, un dispositif au service de la santé-environnement

Le GD4H est une initiative de structuration de données environnementales pour la santé annoncée par le 4<sup>e</sup> plan national santé-environnement (action 18), dont la préfiguration a été confiée au Commissariat général au développement durable (CGDD), à l'automne 2020. Son objectif principal est de contribuer à la production de connaissances pour nourrir la politique publique de santé-environnement quant à l'impact des facteurs environnementaux pour la santé et de stimuler l'identification et la réutilisation des données environnementales existantes, particulièrement riches mais jusqu'alors peu mobilisées, pour faire émerger des associations entre expositions environnementales et pathologies humaines.

Pour ce faire, le dispositif du Green Data for Health vise à faciliter la mobilisation et la valorisation des données environnementales, par les chercheurs et experts, pour un usage en santé-environnement :

- en facilitant l'accès aux données environnementales, y compris au niveau des territoires ;
- en décrivant les données environnementales utiles et disponibles pour être croisées avec des données de santé ;
- en améliorant l'interopérabilité et notamment le croisement des données environnementales entre elles et avec les données de santé ;
- en stimulant un appariement rigoureux des données environnementales avec les données de santé.

1 <https://solidarites-sante.gouv.fr/IMG/pdf/pnse4.pdf>

Une phase d'investigation a été menée par le CGDD (Ecolab et Service de la donnée et des études statistiques) dès la fin de l'année 2020 auprès de l'écosystème des producteurs et utilisateurs de données environnementales et de santé, dont les chercheurs, organismes publics d'expertise, associations, décideurs, acteurs privés, société civile. L'investigation a permis de recentrer le périmètre du dispositif en prenant pour cible, dans un premier temps, les acteurs de la recherche et de l'expertise en santé-environnement, et de définir le périmètre de données concernées. L'objectif d'information pour le citoyen est quant à lui adressé par une autre action du PNSE4, Recosanté.

La phase d'investigation a également facilité l'identification fine des besoins des utilisateurs ciblés, allant plus loin que l'accès centralisé aux données. Pour permettre aux acteurs de l'expertise et de la recherche de mieux mobiliser la donnée environnementale, l'offre de service du GD4H doit répondre aux trois enjeux majeurs ci-après.

#### Repérabilité et accès aux données environnementales

La nature elle-même de la donnée environnementale étant la plupart du temps ouverte par défaut (convention Aarhus, directive INSPIRE), elle constitue un différenciant majeur par rapport aux données de santé, qui sont souvent caractérisées par un accès plus restreint. Toutefois, les acteurs de la recherche académique ou de l'expertise ont besoin d'un appui quant à l'identification des données disponibles sur leur champ de recherche et de leur emplacement. Lorsque ces dernières sont produites par des organismes publics, elles tendent à être plus facilement accessibles car diffusées en ligne, bien que cela ne soit pas systématiquement mis en œuvre. Cependant, il existe certains cas de figure où la donnée est soumise à des restrictions (secret statistique, secret économique, secret médical ou encore données personnelles) et pour lesquels les réutilisateurs des données ont besoin d'être aiguillés dans l'identification des détenteurs de la donnée ainsi que de processus d'accès à mettre en œuvre.

#### Transformation des données environnementales : faciliter leur mobilisation et leur appariement avec des données de santé

En effet, des lacunes dans les données collectées ainsi que la différence entre la nature des données collectées et ce qui est utilisable en santé-environnement, constituent un obstacle important, souvent chronophage, et parfois bloquant pour des croisements avec des données de santé. L'enjeu consiste alors, par exemple, à passer de la donnée de vente, d'émission ou de concentration des polluants à la donnée géolocalisée d'exposition à ces substances ou nuisances. La granularité des données collectées n'est par ailleurs pas toujours adaptée à une bonne interopérabilité avec des données de santé géolocalisées.

Le GD4H vise alors à épauler les réutilisateurs des données environnementales lors de l'identification des modalités et processus d'accès à ces dernières. Il est également envisagé d'ajouter au catalogue des éléments descriptifs permettant de préciser les limites et biais potentiels des données recensées. Par ailleurs, le GD4H ne se substitue pas aux échanges entre producteurs et réutilisateurs mais sert de médiateur afin de faciliter la synergie entre acteurs.

### Création d'une communauté et stimulation d'échanges entre acteurs de l'écosystème

La phase d'incubation du dispositif, portée par l'Ecolab du CGDD (MTECT-MTE) depuis l'automne 2021, vise à affiner et mettre en œuvre l'offre de service du GD4H. Cette dernière est articulée autour de trois briques principales, dans le but d'adresser tout ou partie des enjeux identifiés précédemment :

- un catalogue en ligne<sup>2</sup> pointant vers environ 130 jeux, bases et portails de données publiques environnementales utiles et décrites (métadonnées) pour un usage en santé-environnement ;
- deux outils d'accompagnement juridique en cours de conception : des fiches vulgarisant les fondamentaux du droit de la donnée environnementale (accessibles en ligne) ainsi qu'un outil de formation et d'autodiagnostic pour les producteurs/gestionnaires de données (en cours de conception) ;
- l'animation de la communauté de la donnée en santé-environnement, notamment autour d'actions de montée en qualité des données pour une meilleure mobilisation de ces dernières (sessions en plénière de la communauté, appel à projets conjoint avec le Health Data Hub « La donnée pour la recherche et l'innovation en santé-environnement », organisation d'un Data Challenge visant à développer des outils permettant de mieux mobiliser la donnée environnementale en santé-environnement).

L'offre de service est portée et validée grâce à une gouvernance partagée dès le démarrage de l'action du PNSE4. Cette gouvernance, impliquant les organismes publics de recherche, d'expertise et agences sanitaires, permet en effet de donner et valider les grandes orientations du dispositif.

### Communauté de la donnée en santé-environnement mobilisée autour d'actions d'animation

Afin de fédérer la communauté des acteurs de la recherche et de l'expertise en santé-environnement, des actions d'animation sont menées dans le cadre de l'incubation du Green Data for Health. Parmi ces dernières, des cas d'usage sont conduits sur des problématiques ciblées sous-tendant la mobilisation de données environnementales sur un sujet de santé-environnement. Un Data Challenge qui se déroulera en avril et mai 2023 vise également à mettre en œuvre des projets communs utiles pour une montée en qualité et meilleure mobilisation des données environnementales utiles dans le périmètre de la santé-environnement.

#### Cas d'usage sur les données de qualité de l'eau potable

Lors de l'investigation des problématiques liées à la mobilisation des données environnementales, il avait été identifié la difficulté d'exploiter les données de qualité de l'eau potable remontées des ARS et disponibles à partir de la base SISE-Eaux (à retrouver sur DataGouv ainsi que sur le SIE EauFrance). Les chercheurs et experts avaient exprimé un besoin d'accéder aux concentrations par substance dans l'eau potable à une zone géographique adéquate (ex. : échelle communale ou adresse). Actuellement, les données de la base SISE-Eaux permettent aux chercheurs et experts de bénéficier d'informations quant à la concentration d'une ou de plusieurs substances par unité de distribution (UDI). Toutefois, le contour géographique de ces UDI n'est pas directement renseigné dans la base SISE-Eaux, et une jointure avec une autre source, les données du portail AtlaSanté, est apparue nécessaire. Ces constats ont mené à produire une interface permettant d'extraire les données de qualité de l'eau potable à une maille géographique appropriée, mais également pour visualiser simultanément le périmètre des unités de distribution concernées par les données utilisées.

<sup>2</sup> <https://gd4h.ecologie.gouv.fr/>

### Challenge Green Data for Health

Le Green Data for Health organisera au premier semestre 2023 un Challenge<sup>3</sup> visant à permettre le développement d'outils/de projets communs, contribuant à faciliter la mobilisation des données environnementales en santé-environnement. La collecte des points de difficultés et besoins des acteurs pour mieux mobiliser la donnée environnementale en santé-environnement, ainsi que le recueil des idées de défis à développer, est effectuée par le Green Data for Health auprès de son écosystème d'acteurs afin d'identifier des intérêts convergents et des synergies, et afin de faire émerger ainsi des projets communs, rassemblant potentiellement plusieurs acteurs. L'ambition est de permettre le développement d'outils utiles pour un grand nombre d'acteurs impliqués sur les sujets de santé-environnement, que ces outils puissent ensuite être transférables et remobilisables pour d'autres cas d'usage et mis en ligne en accès ouvert et pérennisé pour être utilisés pleinement à l'issue du Challenge. Une quinzaine de défis sera ainsi sélectionnée et les outils correspondants développés au cours du premier semestre 2023.

### Appel à projets conjoint avec le Health Data Hub

L'appel à projets intitulé « La donnée au service de la recherche et de l'innovation en santé-environnement » a été lancé conjointement par le Health Data Hub et le Green Data for Health (CGDD) à l'automne dernier afin de stimuler et faciliter les croisements des données environnementales et sanitaires, en levant une partie des difficultés d'identification et d'accès aux données et des freins liés aux enjeux d'interopérabilité.

L'accompagnement humain, technique et financier proposé sur 18 mois visera à soutenir l'innovation et la recherche sur des facteurs d'exposition environnementale à fort enjeu de santé publique (bassins industriels, substances perturbatrices endocriniennes et bruit). Ainsi, bien moins souvent traité en recherche que la pollution atmosphérique, le bruit constitue toutefois la deuxième cause de morbidité en Europe, après cette dernière selon l'OMS. Les quatre projets de recherche lauréats de l'appel à projets ont été annoncés le 9 mars dernier lors de la seconde édition de la communauté de la donnée en santé-environnement, organisée par l'Ecolab du CGDD, qui a rassemblé près de 200 participants en présentiel et à distance<sup>4</sup>.

### Une ambition forte portée par les acteurs de l'écosystème

L'expérimentation des premières briques de l'offre de service du Green Data for Health actuellement incubées au sein du CGDD, a permis d'identifier les axes à poursuivre dans le cadre du déploiement et de la pérennisation du dispositif. Ainsi, afin de poursuivre l'ambition portée par l'écosystème d'une réutilisation facilitée des données environnementales, il est prévu de consolider le référencement des données existantes et de mettre en avant des outils utiles. Aller au-delà du référencement en accompagnant la montée en qualité de données et la production de nouvelles données et outils est également attendu de l'écosystème. Enfin, poursuivre l'accompagnement des producteurs et gestionnaires de données dans l'ouverture de ces dernières sera indispensable.

3 <https://challenge.gd4h.ecologie.gouv.fr/>

4 <https://www.ecologie.gouv.fr/annonce-des-laureats-lappel-projets-donnee-recherche-et-linnovation-en-sante-environnement-du-health>

Toutefois la pérennisation du dispositif, actuellement piloté par une gouvernance multi-acteurs comprenant les organismes publics d'expertise et de recherche (INSERM, Santé publique France, Anses et Ineris), ne peut s'opérer sans une gouvernance partagée pérennisée.

D'autres actions sont également envisagées afin de renforcer l'ancrage du dispositif GD4H dans les stratégies multiscalaires (régional, national et européen), telles que développer le réseau de collaboration du GD4H avec les organismes de recherches et les agences sanitaires, intégrer le dispositif au niveau européen et coopérer avec les initiatives régionales dans le cadre des plans régionaux santé-environnement.