

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.

Présentation au GT QuaDoGéo – le 18.10.23

Le Green Data for Health (GD4H)

CONTEXTE :

Selon l'OMS, les facteurs environnementaux et comportementaux sont à l'origine d'au moins 15% des décès en Europe, soit environ 1,4 million de décès par an, pour la plupart évitables. Mieux décrire l'effet des facteurs environnementaux sur la santé humaine, animale et des écosystèmes (Une Seule Santé) constitue un enjeu majeur de santé publique.

De nombreuses données concernant l'état de l'environnement et les facteurs environnementaux pouvant influencer la santé en France sont produites par une grande variété d'acteurs (mesures et modélisations de la pollution de l'air, contaminants de l'eau, registres agricoles, données météorologiques, registres des ICPE...).

Pourtant, le PNSE a fait le **constat d'une sous-utilisation de ces données très riches.**

UTILISATEURS FINAUX :

- _ **Les producteurs de données environnementales**
 - _ **Les chercheurs et experts en santé environnement**
- Plus largement : acteurs locaux, privés, associations etc.

OBJECTIF :

En tant qu'action phare du PNSE4, l'objectif du GD4H est de 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.**



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



Décrire les données environnementales pertinentes et disponibles pour un **appariement rigoureux avec des données de santé**



Améliorer la **communication** et la **synergie** entre les différents acteurs de la santé environnement (enjeux d'interopérabilité)

La plateforme en ligne du Green Data for Health

The screenshot shows the main navigation menu with 'Catalogue' selected. The main heading is 'Catalogue' with the subtitle 'Des jeux de données environnementales destinés à un usage en santé-environnement'. A search bar is visible with the text 'Jeux de données environnementales (127)'. Below the search bar, there are filters for 'LCSQA' and 'INERIS'. The first data entry is 'GEOD'AIR - Système national de gestion des données d'observation de la qualité de l'air', with details on its spatial coverage (FRA), temporal coverage (2013 - en cours), and granularities. The second entry is 'Agriculture biologique 2020 - nombre d'opérateurs engagés en agriculture biologique', with details on its spatial coverage (Non renseigné) and temporal coverage (2013 + 2014 + 2015 + 2016 + 2017 + ...).

This screenshot shows a detailed view of an environmental data entry. It includes a table with fields like 'Santé environnement', 'Milieu', 'Voie d'exposition', and 'Air extérieur'. A 'Version beta' notice is present, stating that the platform is under development and soliciting feedback from data producers and researchers. A 'Contactez-nous !' section encourages users to report missing data or provide general remarks. The footer includes the French government logo and the European Union logo.

This screenshot displays the 'Green Data For Health' service page, which aims to 'Valoriser la donnée environnementale pour la recherche en santé environnement'. It features three main service cards: 'Catalogue de données' (to help find environmental data), 'Aide juridique' (to answer questions about data sharing), and 'Communauté donnée santé environnement' (to foster synergy between actors). The page also includes the French government logo and the European Union logo.

Lien : <https://gd4h.ecologie.gouv.fr/>

L'offre de service du Green Data for Health (CGDD/Ecolab) – Octobre 2023

Gouvernance

Animation

Animation de la communauté de la donnée en santé-environnement

Conduite de cas concrets

_ Cas d'usage Urbanisme favorable à la santé - en cours

Appel à projets HDH x GD4H

_ 1° appel à projets – en cours

_ 2° appel à projets – à venir

Challenge GD4H

_ GT interopérabilité des données (CNIG) – en cours

_ Formation sur les données santé-environnement (HDH) – en cours

Catalogue de ressources

Ouverture au grand public du catalogue de données en ligne.

_ Enrichissement et mise à jour en partie automatisée – en cours

_ Introduction des réutilisations – en cours

Aide juridique

_ **Autoformation** : une ressource juridique en ligne

_ **Autodiagnostic** : outil juridique d'autodiagnostic – en cours

_ Une charte pour favoriser l'ouverture des données – en cours

- ✓ **Communauté de 450 membres** actifs.
2° édition de la communauté : **170 participants**.
- ✓ **AAP : 4 projets de recherche** lauréats accompagnés.
- ✓ **Challenge : 15 défis** cadrés et 15 outils réalisés au printemps 2023. Près de **150 acteurs mobilisés** dans les défis.
- ✓ **Catalogue de données : 150 ressources** référencées.
- ✓ **Plateforme en ligne :**
<https://gd4h.ecologie.gouv.fr/>
Depuis fin 2022, près de **21 000 consultations** par 4 500 visiteurs différents.
- ✓ **Ressourcerie : 20 fiches pédagogiques**.
- ✓ **Outil d'autodiagnostic : 1 commun numérique** en cours de cadrage.

CATALOGUE DE DONNÉES

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 la conformité des valeurs et de leur croisement entre elles et avec d'autres portails (**interopérabilité**)

150 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
--------	-------------------	--	---	---	--------------------------	------------------------

OUTILS JURIDIQUES

OBJECTIFS DES OUTILS JURIDIQUES

Permettre à un utilisateur sans connaissance juridique de :

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

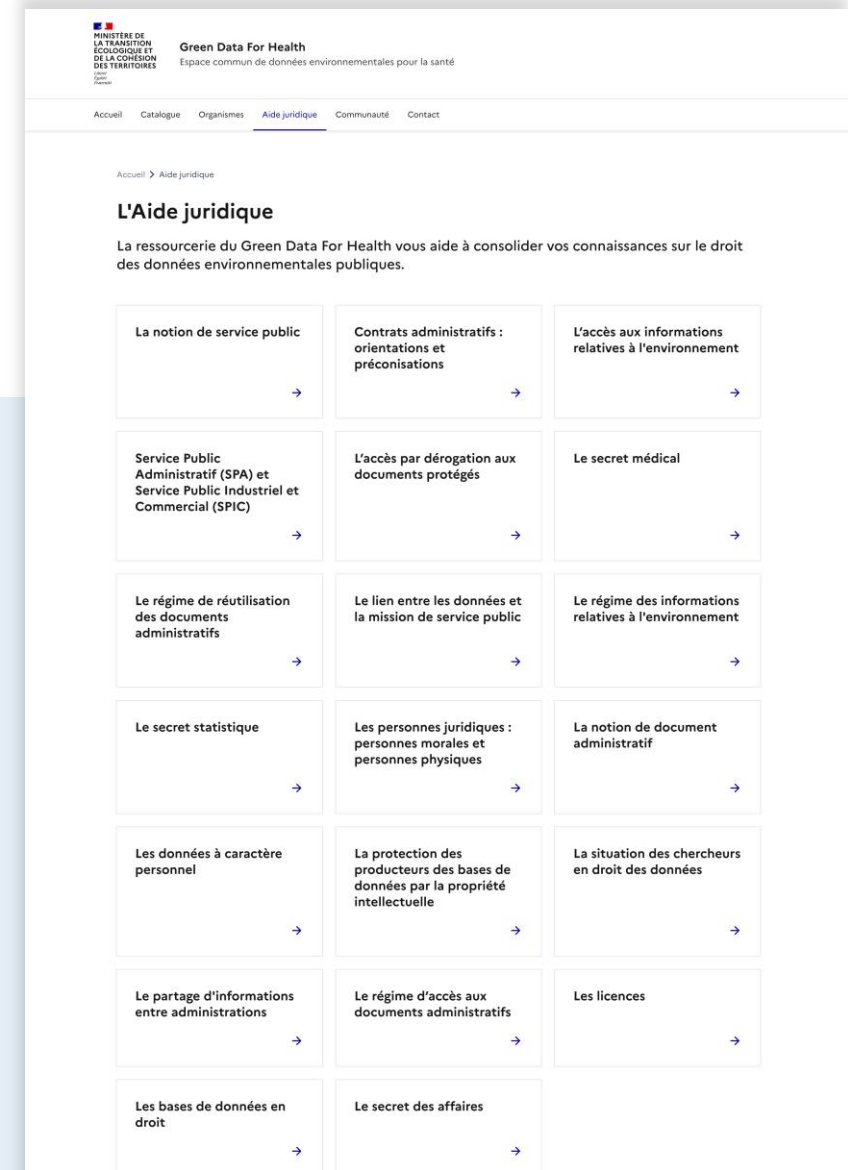
FOCUS - RESSOURCERIE

20 fiches pédagogiques accessibles en ligne vulgarisant les grands

principes du droit des données : <https://gd4h.ecologie.gouv.fr/juridique>

Les fiches portent 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...)



OUTILS JURIDIQUES

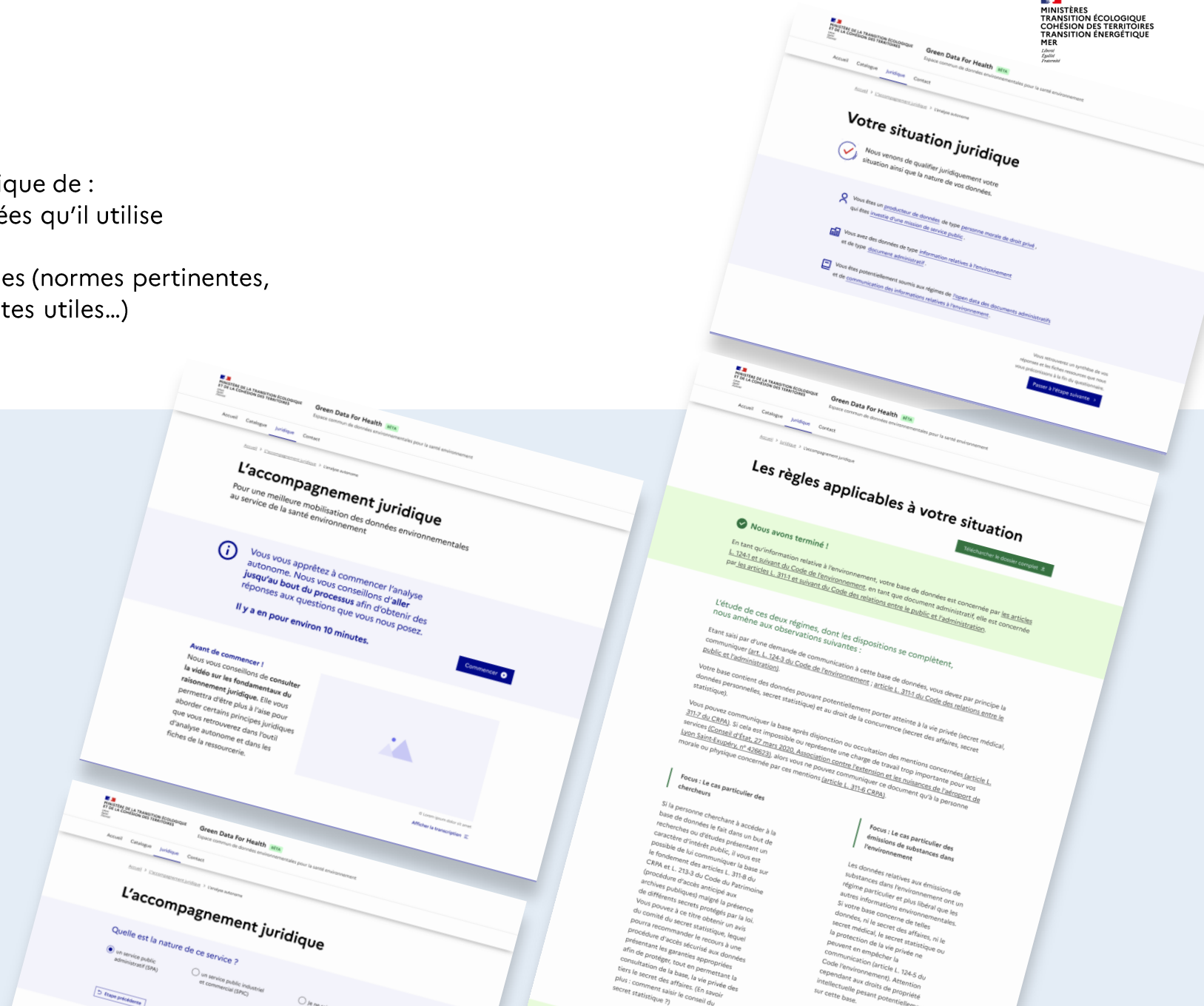
OBJECTIFS DES OUTILS JURIDIQUES

Permettre à un utilisateur sans connaissance juridique de :

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

FOCUS – OUTIL JURIDIQUE D'AUTODIAGNOSTIC (en cours)

Le développement d'un **outil juridique d'autodiagnostic** (projet en cours) proposant de réaliser soi-même un diagnostic juridique grâce à une suite de questions/réponses (arbre de décision) permettant de qualifier juridiquement sa base de données et d'identifier le régime applicable.



OUTILS JURIDIQUES

FOCUS – PROJET DE CHARTE*

PROJET

Une charte à laquelle chaque producteur pourrait décider d'adhérer :

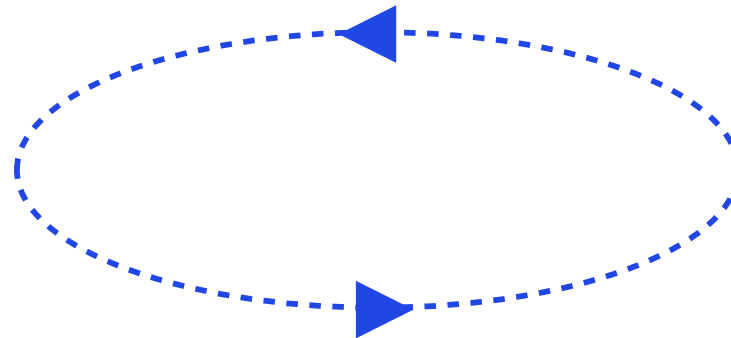
- **Pédagogique** répondant à un besoin d'acculturation juridique
- Formalisant un **engagement réciproque**, « gagnant-gagnant »

PÉRIMÈTRE DE DONNÉES

La charte pourrait s'appliquer à l'ensemble des données auquel le cadre de l'open data s'applique.

PRODUCTEUR DE DONNÉES

- **Valoriser la démarche d'ouverture** des données par les producteurs.
- Permettre aux producteurs de **bénéficier des retours des ré-utilisateurs/quant aux usages afin de favoriser la montée en qualité** ultérieure des jeux de données.



RÉUTILISATEUR DE DONNÉES

- S'assurer d'un **retour du producteur quant à la qualité de la donnée (montée en qualité)**.
- Pour les **créateurs de réutilisations**, valoriser les **productions** issues des jeux de données.
- Pour l'ensemble des ré-utilisateurs, **mettre à disposition un ensemble de réutilisations** de données.

GD4H

- Contribuer à la démarche d'**ouverture** et de **valorisation** des données.
- **Enrichir le catalogue avec des ressources** (données de qualité et des réutilisations associées).
- **Consolider l'animation de la communauté** de la donnée en s-e.

**Projet en cours de cadrage susceptible d'évoluer.*

ANIMATION

Challenge Green Data for Health

CONTEXTE

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

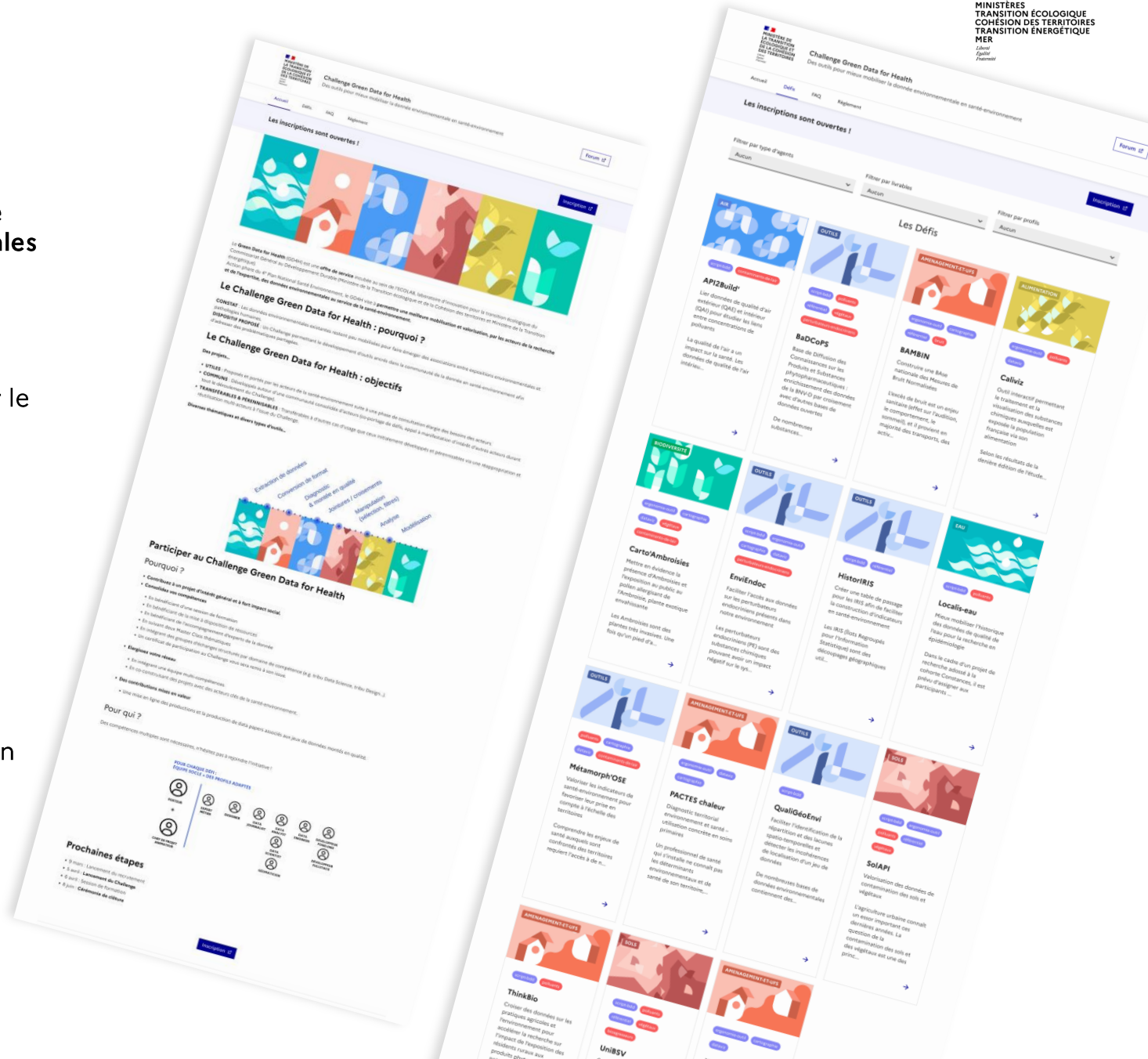
OBJECTIFS

Faciliter la mobilisation des données environnementales par le développement d'outils :

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

L'INITIATIVE

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

Challenge Green Data for Health

LE RESULTAT

Près de **150 acteurs** experts de la santé-environnement et bénévoles experts du numérique mobilisés autour de **15 défis**.

LES 15 DÉFIS



ALIMENTATION

↳ **Caliviz** : Un outil interactif permettant le traitement et la visualisation des substances chimiques auxquelles est exposée la population française via son alimentation.



AIR

↳ **API2Build** : Un script pour lier les données de qualité d'air extérieur (QAE) et intérieur (QAI).



OUTIL TRANSVERSE

↳ **Histor'IRIS** : Une table de passage pour les IRIS afin de faciliter la construction d'indicateurs en santé-environnement.

↳ **QualiGéoEnvi** : Une application web d'analyse et de conversion des données environnementales temporelles et géographiques.

↳ **EnviEndoc** : Une carte interactive et un outil d'extraction de données sur les perturbateurs endocriniens.

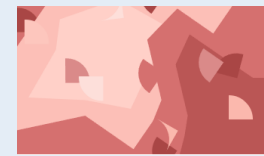
↳ **Base C3PO** : Une base de connaissances sur les produits phytopharmaceutiques à partir de sources ouvertes.

↳ **Métamorph'OSE** : Une base de données sur les caractéristiques des territoires.



BIODIVERSITÉ

↳ **Carto'Ambroisies** : Une base de donnée et carte interactive sur les données liées à l'ambroisie.



SOLS

↳ **SolAPI** : Un script de traitement des données brutes sur la pollution des sols.

↳ **UniBSV** : Une base de données exploitable sur les bulletins de santé des végétaux.



EAU

↳ **Localis-eau** : Un script d'appariement des unités de distribution (UDI) de l'eau à des données géographiques.



AMENAGEMENT & Ufs*

↳ **ThinkBio** : Un script d'appariement des données de vente des produits phytosanitaires à des données environnementales.

↳ **Urbanité** : Une base de donnée et une cartographie pour la comparaison de caractéristiques environnementales et sociales démographiques au niveau de l'IRIS.

↳ **BAMBIN** : Une BAsE nationale des Mesures de bruit Normalisées.

↳ **PACTES Chaleur** : Une cartographie des populations à risque face aux vagues de chaleurs.

* Page des défis du Challenge GD4H :

<https://gd4h.ecologie.gouv.fr/defis>

* Code source des défis :

<https://gitlab.com/data-challenge-gd4h/>

Groupe de travail « données géolocalisées en santé » (CNIG)

CONSTAT

Le **manque d'interopérabilité entre données environnementales et sanitaires** constitue un frein important à la mobilisation des données au service de la santé-environnement.

OBJECTIF & MISSIONS DU GROUPE DE TRAVAIL

Faciliter la mobilisation et le croisement des données environnementales et sanitaires en identifiant, afin de prioriser les manques et axes à privilégier pour permettre une meilleure interopérabilité des données :

- _ **les besoins et usages** associés à la mobilisation et au croisement de ces données
- _ **les ressources structurantes** (référentiels et standards existants) en santé-environnement

ANIMATION

Lancement du GT par le Green Data for Health (CGDD/Ecolab). **Co-animation avec le Health Data Hub sur les livrables 1 et 2.**

MEMBRES

Organismes producteurs et réutilisateurs de bases de données environnementales et sanitaires.

Profils d'experts métier (santé-environnement) et de la donnée.

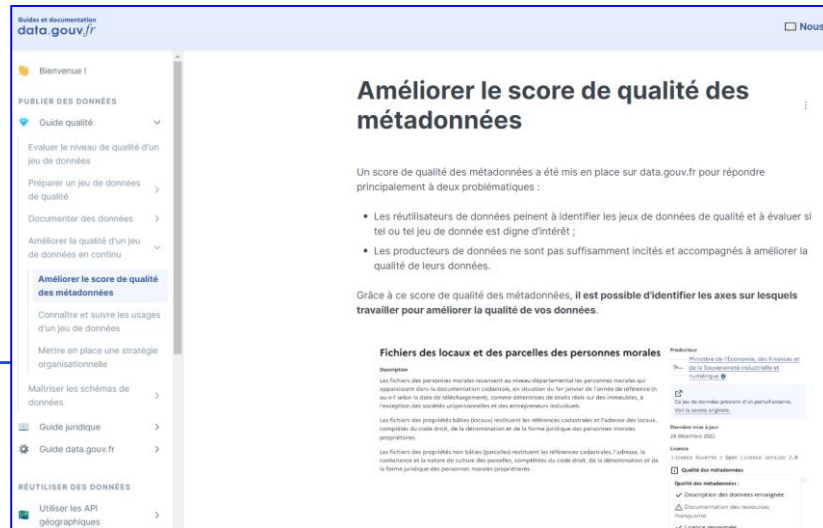
* **Ouverture du GT à tout acteur intéressé.**

Lien vers la page du mandat : <https://cnig.gouv.fr/gt-ethique-de-l-usage-de-la-donnee-et-de-l-a26117.html>

La qualité des données au sein du Green Data for Health

LA QUALITÉ DES MÉTADONNÉES & DE LA DOCUMENTATION

- Levier 1 : la documentation des jeux de données du catalogue.
- Levier 2 : l'intégration d'un indicateur sur la qualité des métadonnées.



Indice qualité des métadonnées de data.gov

LA QUALITÉ INTRINSÈQUE DES JEUX DE DONNÉES

- Levier 1 : Permettre de faire un commentaire sur les différentes ressources du catalogue.
- Levier 2 : Proposer une charte open data (projet en cours de cadrage).

FOCUS – PROJET DE CHARTE*

PROJET

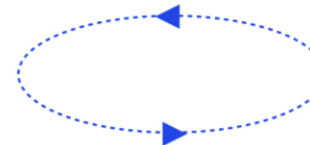
- Une charte à laquelle chaque producteur pourrait décider d'adhérer :
 - Pédagogique répondant à un besoin d'acculturation juridique
 - Formalisant un engagement réciproque, « gagnant-gagnant »

PÉRIMÈTRE DE DONNÉES

La charte pourrait s'appliquer à l'ensemble des données auquel le cadre de l'open data s'applique.

PRODUCTEUR DE DONNÉES

- Valoriser la démarche d'ouverture des données par les producteurs.
- Permettre aux producteurs de bénéficier des retours des réutilisateurs/quant aux usages afin de favoriser la montée en qualité ultérieure des jeux de données.



RÉUTILISATEUR DE DONNÉES

- S'assurer d'un retour du producteur quant à la qualité de la donnée (montée en qualité).
- Pour les créateurs de réutilisations, valoriser les productions issues des jeux de données.
- Pour l'ensemble des réutilisateurs, mettre à disposition un ensemble de réutilisations de données.

GD4H

- Contribuer à la démarche d'ouverture et de valorisation des données.
- Enrichir le catalogue avec des ressources (données de qualité et des réutilisations associées).
- Consolider l'animation de la communauté de la donnée en s-e.

*Projet en cours de cadrage susceptible d'évoluer.

Comment prendre en compte tous les enjeux de qualité des données au sein du catalogue en ligne ?

Quels leviers pour accroître la qualité des données ?

Quels retours les plus utiles pour les producteurs pour favoriser la montée en qualité de leurs données ?

