

Rayane Hadji
Master 2 – Santé publique
Evaluation et gestion des risques en santé-environnement.
2016/2017

Mémoire de fin d'étude

Problématique :

Open Data – Big Data : Quel rôle pour le citoyen dans la gestion sanitaire et
environnementale des risques ?



Date de la soutenance : Octobre 2017

Président de Jury : M. DENAYER Doyen de l'ILIS

Directeur de mémoire : M. OCCELLI Maître de Conférences, Université Lille 2

Membre externe du Jury : M. PICHOT – DAMON Chargé de Mission Open Data, Métropole Européenne de Lille (MEL)

Faculté d'Ingénierie de Management de la Santé - ILIS

Université Lille 2 – Droit et Santé

Adresse :

42, Rue Ambroise Paré

59200 - LOOS

Téléphone :

03.20.62.37.37



Remerciements

*Mes remerciements se dirigent tout d'abord vers le **Docteur Franck-Olivier Denayer**, Doyen de la Faculté d'Ingénierie et de Management de la Santé et Président de Jury, il fut pour moi un exemple de professionnalisme à travers la qualité de ces enseignements en gestion des risques en santé-environnement. Merci à vous et à toute l'Equipe Pédagogique de l'ILIS pour votre dévotion pour faire de nous des acteurs de notre vie professionnelle.*

*Je remercie aussi vivement **Monsieur Florent Occelli**, - Maître de Conférences en santé publique et environnement à l'ILIS – pour son accompagnement, son écoute et la qualité de ses conseils tout au long de la rédaction de ce mémoire.*

*Merci à **Monsieur Etienne PICHOT – DAMON** - Chargé de Mission Open Data à la Métropole Européenne de Lille (MEL), d'avoir accepté mon invitation pour faire partie du Jury.*

*Un grand merci à tous les acteurs de la donnée que j'ai eu le privilège de rencontrer lors de mes recherches. Je pense particulièrement à **Madame Armelle Gilliard** –Spécialiste en Open Data, **Monsieur Daniel Kaplan** – Co-Fondateur de la Fondation Internet Nouvelle Génération (FING), et aux intervenants de la Conférence Transitions²*

*Enfin, je remercie chaleureusement tous mes **proches et amis** pour leur soutien au cours de mes années d'études et qui ont consacré de leur temps pour la relecture de ce Mémoire.*

Sommaire

INTRODUCTION:	1
PARTIE 1 : ACCES ET APPROPRIATION DES DONNEES SANTE-ENVIRONNEMENTALES: LE CITOYEN UTILISATEUR/PRODUCTEUR	6
I. POSITIONNEMENT STRATEGIQUE DU CITOYEN DANS LE DOMAINE DES DONNEES	6
1. LES ACTEURS AUTOUR DU CITOYEN	6
2. PLACEMENT DU CITOYEN DANS "L'ECOSYSTEME DATA".....	8
II. EN TANT QU'INDIVIDU / CITOYEN, J'AI ACCES A QUOI ?	9
1. QUI PEUT AVOIR ACCES AUX DONNEES ET OU?	9
2. COMMENT SE PROCURER DE NOUVELLES DONNEES?	11
III. CONCRETEMENT, QUELS USAGES PAR LE CITOYEN DES DONNEES SANTE-ENVIRONNEMENTALES?	12
1. CARTOGRAPHIE UN OUTIL POUR UNE GESTION DES RISQUES SANTE-ENVIRONNEMENT? (COMMENT REPRESENTER MES DONNEES)	12
2. EXEMPLES D'OUTILS DE CARTOGRAPHIE POUR UN USAGE DU CITOYEN.....	15
IV. LIMITES DE L'USAGE DES DONNEES SANTE-ENVIRONNEMENTALES PAR LE CITOYEN	19
1. ETAT DES LIEUX DES DONNEES ENVIRONNEMENTALES - ABSENCE DE CERTAINES DONNEES.....	19
2. INCERTITUDES INTRINSEQUES DE LA DONNEE	20
3. ACCESSIBILITE, USAGES ET COMPETENCES	20
CONCLUSION PARTIE 1	22
PARTIE 2: PERSPECTIVES D'AVENIR PAR LE PRISME DU CITOYEN: IMPACTS MULTISECTORIELLES DES DATAS SE	23
I. POLITIQUE	23
1. TRANSPARENCE ET EMPOWERMENT: UN MOUVEMENT POLITIQUE EMERGENT.....	23
2. INTERETS POLITIQUES POUR UN DEVELOPPEMENT DE CES TECHNIQUES DE GESTION SANITAIRE ET ENVIRONNEMENTALE.....	25
II. ECONOMIQUE	27
III. SOCIAL	29
1. UNE PRISE DE CONSCIENCE DES PROBLEMATIQUES ENVIRONNEMENTALES VIA LES DONNEES PAR LE CITOYEN.....	29
2. MIGRATION VERS UN SUIVI PERSONNALISE DES INDIVIDUS: LE QUANTIFIED-SELF OU L'INDIVIDU OBJET DATA. .	32
IV. TECHNIQUE	34
1. VERS DE NOUVELLES MANIERES DE PRODUIRE DE LA DONNEE ENVIRONNEMENTALE	34
2. CITOYEN PRODUCTEUR DE DONNEES NON STRUCTUREES: QUELLES OPPORTUNITES EN SANTE ENVIRONNEMENT?	35
V. ÉCOLOGIQUE	36
1. IMPACT SUR LA GESTION DES RISQUES EN SANTE ENVIRONNEMENT	36
2. IMPACT ENVIRONNEMENTAL DE L'USAGE DES DONNEES.....	36
VI. LEGISLATIF	38
1. DONNEES ENVIRONNEMENTALES: ETHIQUE ET RESPONSABILITES POUR LES UTILISATEURS	38
2. POSSESSION ET PROTECTION DES DONNEES (PERSONNELLES ET NON PERSONNELLES).....	38
CONCLUSION PARTIE 2	41
CONCLUSION GENERALE DU MÉMOIRE	42
BIBLIOGRAPHIE	44
TABLE DES MATIERES	48
TABLE DES ANNEXES	I
RESUME / SUMMARY	XIX

Glossaire

Par ordre alphabétique :

Algorithme: Programme de traitement de données constitué de suite d'instruction ayant pour objectif l'obtention d'un résultat

Big Data : outils et méthodes permettant de collecter, traiter et d'extraire de l'information à un très grand volume de données

CADA : Commission d'Accès aux Documents Administratifs

Cloud (computing) : Traduction de "nuage", représente un réseau informatique partagé (sur internet) et configurable permettant d'utiliser les capacités de stockage et de calculs selon ses besoins

CNIL: Commission national informatique et liberté

Crowdsourcing: Système de recueil d'information en ligne dans lequel un groupe d'individus consolide ensemble une base de données.

CSV: Comma - separated values (traduction de valeurs séparées par une virgule) est un format de fichier informatique utilisé dans les tableurs

Dark net: partie du réseau internet non référencée et accessible uniquement à l'aide de logiciels spécifiques d'anonymisation contenant du contenu illégal

Data center: lieu regroupant les serveurs informatiques dans lesquels sont stockés les données

Data-literacy: compréhension de l'utilisation de la donnée

Data-visualisation: visualisation des données sous forme graphique

Donnée : Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement (en anglais : data)

Donnée non-structurée: données pouvant pas être inscrites dans des bases de données classiques. (exemple les données textuelles et les images)

DREAL : Direction Régionale de l'Environnement de l'Aménagement et du Logement

EDGI: Environmental Data & Governance Initiative

Empowerment : mouvement politique ayant pour objectif la réappropriation du pouvoir politique par la population

EPI : Environmental Performance index

FING : Fondation Internet Nouvelle Génération

Gb (Go) : Giga bit (Giga octet) - unité de mesure de la mémoire de stockage informatique

Google street view : service Google de navigation permettant de voir les rues à 360°

Hackaton: Court évènement organisé ayant pour objectif la création de nouveaux outils numériques seul ou en groupe sur un sujet donné

Homo-data-sapiens : concept selon lequel l'homme est défini par les données qu'il produit

IGN: Institut National de l'Information Géographique et forestière

Infolab : projet visant à promouvoir l'utilisation des données par tous les acteurs de la société

INSPIRE : Directive Européenne demandant aux acteurs publics d'harmoniser la collecte et la diffusion de données environnementales d'ordre géographiques

JSON: JavaScript Object Notation, format de fichier informatique

Méta-donnée: informations complémentaires décrivant une donnée (nom du créateur, date...)

Open Data : système d'ouverture et d'accès des données publiques

Open source: système permettant l'accès et l'utilisation libre des programmes

Quantified-self : concept des méthodes et outils permettant de faire des mesures de variables relatives à son mode de vie

RATP : Régie Autonome des Transports Parisiens

SIG: Système d'information géographique

SIGLES: Systèmes d'Information Géographique et Lien Environnement – Santé

SNCF : Société Nationale des Chemins de fer Français

Think-tank: organisation privée qui regroupe un ensemble de professionnels experts autour d'un même sujet

USEPA: US Environment Protection Agency

XLS: Format de fichier informatique correspondant au fichier de tableur Excel

Table des illustrations

PYRAMIDE DE LA DONNEE	2
EXEMPLE D'ARBRE DE DECISION - RECHERCHE DIVERS TYPES DE DONNEES ENVIRONNEMENTALES	10
CARTOGRAPHIE DES CAS DE CHOLERA PAR LE DOCTEUR JOHN SNOW DANS LE QUARTIER DE SOHO (1855)	13
SCHEMA DE SUPERPOSITIONS DES COUCHES GEOGRAPHIQUES EN SIG (SOURCE: INSTITUT EN CONSEIL ET D'ETUDES EN DEVELOPPEMENT DURABLE - ICEDD - ICEDD.BE)	14
CARTOGRAPHIE DES RUES ET PLACES LES PLUS POLLUEES EN PARTICULES FINES (MG/M3) DE LA CAPITALE (SOURCE: LE PARISIEN / AIRPARIF 2012)	14
CAPTURE D'ECRAN DE L'INTERFACE D'OPEN SOLAR MAP.....	16
UTILISATION DES DONNEES ENVIRONNEMENTALES SUR LE SITE ETALAB.....	21
EVOLUTION DU NOMBRE DE DEMANDE D'ACCES AUX DONNEES PUBLIQUES A LA CADA ET REPARTITION EN POURCENTAGES DES DEMANDES AUTOUR DE LA THEMATIQUE ENVIRONNEMENTALE (SOURCE - RAPPORTS D'ACTIVITES ANNUELS ENTRE CADA 2012 - 2015).....	23
EVOLUTION DU JOUR DE DEPASSEMENT ENTRE 1969 ET 2017	27
EVOLUTION DU NOMBRE DE PUBLICATION D'ETUDES SUR DES PROJETS DE SCIENCES CITOYENNES EN FONCTION DE LA THEMATIQUE ENTRE 1997 ET 2014 (SOURCE : FOLLETT, RIA; STREZOV, VLADIMIR – 2015).....	30
CONCEPT D'INFOLAB (SOURCE- FING)	30
EXEMPLES D'UNE STATION DE MESURE MOBILE A GAUCHE) ET D'UN APPAREIL PORTABLE FLOW A DROITE (SOURCES - GWADAIR.FR / PLUME LABS).....	34

Introduction:

Compte tenu de la dégradation de notre environnement et de leurs conséquences sur la santé (12 millions de décès /an dans le monde (1)), les impacts sanitaires liés à l'environnement sont de plus en plus inscrits dans le débat public. Les populations sont multi-exposées aux polluants de leur environnement (alimentation, pollution des milieux, travail...) et sont les premières concernées par ces problématiques sanitaires globales.

Parallèlement à cette problématique d'ordre environnementale, l'usage des nouvelles technologies de l'information et de la communication (NTIC) sont dans une croissance élevée (2). En 2015, une étude menée par le DOMO (entreprise américaine spécialisé dans les données) a annoncé une augmentation de 20 % des utilisateurs du web dans le monde entier pour atteindre les 3 milliards entre 2013 et 2015 (3). Ceci est corrélé en partie par l'évolution des pratiques comme la multiplication des interfaces (mobiles, ordinateurs, tablettes, objets connectés...) et des services (réseaux sociaux, géolocalisation, achats, mesures....) dans le domaine professionnel comme dans la sphère privée.

Ce système a modifié la manière dont les populations vivent au quotidien et sont en relation avec les autres personnes de leur communauté. A ce jour, 3,4 milliards de personnes ont un accès à internet (3). Tous ces utilisateurs génèrent un flux de données prodigieux de l'ordre de 4.4×10^{12} GB.

Compte tenu de ces volumes d'informations à traiter, issus de la production d'échanges de diverses sources (internet, objets connectés, ...) et du nombre d'utilisateurs, nous constatons que ces masses des données sont potentiellement un vivier inédit de connaissances dans le domaine de la santé publique et plus précisément en environnement-santé. Ainsi, c'est dans ce contexte globalisé, entre numérique et santé environnementale, que nous nous trouvons aujourd'hui. Un contexte régulé par un élément clé: la donnée. Atomes primaires de ces informations, les données sont indispensables et nous devons au préalable comprendre des concepts autour de celle-ci.

Pour définir la donnée, nous pouvons nous référer à la définition inscrite dans l'arrêté du 22 décembre 1981 sur l'enrichissement du vocabulaire informatique (4) . Cette dernière décrit la donnée comme la « représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement (en anglais : data) ». Cette définition induit

plusieurs notions importantes à assimiler pour saisir le fonctionnement des technologies qui les utilisent.

La donnée peut exister sous des formes multiples. Dans le contexte santé-environnement les données peuvent se manifester de la manière suivante: il peut s'agir de valeurs numériques, de caractères alphanumériques. Il peut s'agir aussi de valeurs mesurées par un appareil (températures, pression, concentration d'un polluant dans un milieu), ou alors d'informations recueillies dans une enquête ou une étude (nom de rue, informations personnelles, nombre d'habitants, coordonnées géolocalisées...). Elles peuvent être aussi des données textuelles ou des images issues de canaux divers comme les réseaux sociaux par exemple. Dans ce cas, les données sont dites " non structurées" et demandent des méthodes de traitements différentes. De manière classique, les données environnementales usent de la classification sous la forme de tableaux structurés.

Pourquoi les données sont-elles intéressantes? Elles vont avoir un rôle capital pour répondre à des hypothèses de travail, confirmer l'observation de phénomènes et faire avancer la démarche scientifique dans son ensemble. Pour comprendre leur utilité, nous pouvons citer Dave Tarrant, chercheur et formateur à l'Open Data Institute (5). Selon lui « la donnée est le pilier d'une pyramide en 3 parties » (Figure 1). De par sa nature, la donnée est une matière première pour émettre de l'information, le noyau primaire vers la connaissance.

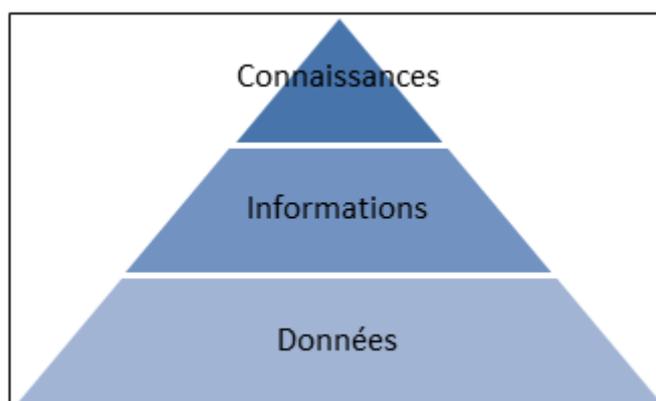


Figure 1: Pyramide de la donnée

Dans une démarche scientifique, comme le sont les sciences environnementales, la qualité et la quantité des données sont primordiales. Par ailleurs, en raison de leurs origines diverses, elles apportent une variabilité en termes de types de données. Le large champ de la santé environnementale (risques naturels et technologiques, qualité des milieux et exposition des populations, sciences humaines et sociales, démographie, épidémiologie,

toxicologie...) rend difficile l'identification précise du volume de données disponibles. Il est ainsi difficile d'évaluer la quantité totale que représente les données environnementales en France à ce jour. Néanmoins les volumes sont assez importants pour être exploités. C'est pourquoi, ces multiples faisceaux représentent un vivier exceptionnel d'informations. Les données sont dans un premier temps récoltées et stockées pour en extraire de l'information. Dès lors que l'information a été traitée, nous pouvons en déduire des résultats (causalité, corrélation par exemple) et en obtenir des connaissances plus approfondies sur un sujet.

Quand bien même ces deux facteurs (qualité et quantité) seraient réunis, cela ne serait pas suffisant. En effet, une donnée brute sans traitement n'apporte pas ou très peu de valeur ajoutée. Pour en extraire l'information qui nous est utile, il est nécessaire de la traiter, de l'analyser correctement. Progressivement des capteurs, outils de collectes et des sources de données se sont développés. Ce pool de sources de données est à la base des moyens de traitement utilisant les technologies liées à l'Open et au Big Data.

La multiplication des outils informatiques et de mesures ont engendré une augmentation majeure dans les volumes de données produits. Compte tenu des informations latentes issues de ces données, et en se basant sur cette méthode pyramidale, il paraît logique de vouloir se les approprier et de les traiter de la manière la plus pertinente. Dans ce contexte, deux outils informatiques existent: l'Open Data et le Big Data. Ces deux concepts s'articulent autour d'un objectif d'acquisition de la connaissance à partir de volumes de données importants.

L'émergence de l'Open Data remonte aux années 90-2000 au cœur de la communauté scientifique anglo-saxonne. Il s'agissait à l'époque d'un besoin de partage et de transparence des données en géophysique et en environnement (6) . Cette volonté de diffuser en libre accès des "biens communs" fait aussi écho dans l'informatique et dans le mouvement "open source". "Open source" que nous pouvons traduire par « code source ouvert », correspond à une philosophie de transparence et de partage. Pour y parvenir, les outils open source sont soumis à des licences qui permettent aux développeurs d'accéder, modifier et de réutiliser le code source à d'autres fins. L'open source autorise le développeur à accéder au cœur du programme et à tous ses rouages et à le réutiliser comme il le souhaite (copie partielle ou intégrale du code...). De ce fait, comme l'Open source, l'Open Data cherche à mettre en libre accès les données, s'inscrit dans cette logique d'ouverture et invite à la réutilisation. Dans son organisation, les données en libre accès sont disponibles sur des plateformes dédiées sur internet. Parmi elles, les données publiques sont celles qui

“ont été reçues ou produites par un organisme public et ce dans le cadre d'une mission de service public” (7).

Second outil de traitement que nous aborderons, la technologie Big Data est un outil permettant de collecter, traiter et d'extraire de l'information à un très grand volume de données. Elle se fonde sur le principe des “4 V” : volume, vélocité, variété, véracité. Ces termes, ensemble, caractérisent les données et leur traitement durant le processus. Le Big Data traite de très grands volumes de données (ordre de grandeur = Téra octet). Ceci s'explique en partie par la multiplication des appareils numériques et la croissance des capacités de stockage de ces derniers. Le Big Data se caractérise aussi par sa vitesse de traitement (parfois en temps réel). La variété et la véracité signifient que le traitement implique la possession de nombreuses variables qui dans la mesure du possible sont vérifiées et dont les informations sont correctes. Cette qualité de données réduit la marge d'erreur dans l'extraction de l'information.

Les données collectées sont issues des bases de données ou des réseaux (internet, réseaux sociaux...). L'extraction de ces données a pour vocation plusieurs dessins : tri, détection des signaux et observation des phénomènes et des tendances émergentes. Pour cela, les ingénieurs utilisent des algorithmes. Ce sont des programmes informatiques qui permettent de traiter des données selon plusieurs modèles mathématiques (par exemple les régressions). Ces flux d'informations sont collectés par des entreprises, des start-up et par toutes personnes ayant les capacités de le faire (par exemple les « Data Scientist », « Data Analyst »).

Nous savons à présent que l'utilisation des données sous forme ouverte ou en Big Data représente un formidable vivier potentiel d'informations. En effet, ces deux concepts (Open et Big Data) sont liés l'un avec l'autre et peuvent trouver des interactions par les sources des bases de données. Présentes en quantités importantes, pourrions-nous imaginer de saisir ces données pour améliorer la gestion des risques environnementaux ? Pour cela, elle ne peut se détourner du chemin de la connaissance. Dans cette optique, le gouvernement soutient son développement au travers de la “ loi pour une République numérique” (8). Elle vise à « favoriser la circulation des données et du savoir”. C'est pourquoi, elle promeut la création de services et la réutilisation des données à destination du citoyen. C'est ainsi que la France a atteint la 4ème place mondiale dans le classement de l'indicateur OKF (Open Knowledge Foundation) en 2016 avec un score de 70%. Cet audit évalue et classe les Etats selon les niveaux d'ouverture de leurs données. Sur les

sujets environnementaux, l'Open Data en matière de qualité air et de l'eau sur le territoire sont estimés à 85% et 70%.

Ce mémoire illustrera le fait que l'utilisation des Big et Open Data vont vraisemblablement construire un nouveau modèle riche et complexe de gestion des risques sanitaires et environnementaux à l'échelle individuelle. Face à ces nouveaux pouvoirs qui lui sont offert, le citoyen tient à jouer un rôle. En effet, des outils se développent utilisant des nouvelles technologies des données. Des solutions sont testées en faveur de l'environnement mais soulèvent un grand nombre d'interrogations. Ces nouveaux outils de gestion des risques incarnent-ils l'avenir ? Dans une époque de modernisation des méthodes de gestion, les "data" seraient-elles une technologie porteuse pour son développement ? Quelles thématiques santé-environnement sont concernées ? Comment les data peuvent-elles modifier sur notre territoire l'environnement en général et celui du citoyen en particulier?

Pour y répondre, nous allons dans un premier temps étudier le positionnement ainsi que le rôle du citoyen dans ce système dans le cadre français. Ensuite, nous analyserons les données "santé environnementales" disponibles aux citoyens et leurs usages. Nous nous concentrerons aussi sur les limites de leurs utilisations en matière de gestion des risques environnementaux. Enfin, nous conclurons sur les perspectives d'avenir de la convergence citoyenne et des données en faveur de la gestion des risques environnementaux.

Partie 1 : Accès et appropriation des données santé-environnementales: le citoyen utilisateur/producteur

Après avoir étudié les concepts du Big Data et de l'Open Data, focalisons-nous sur les données environnementales et leurs liens avec les citoyens.

Comment se positionne le citoyen dans l'écosystème Open /Big data? Quels sont ses forces, ses faiblesses? Quel rôle tient-il ? Pour répondre à toutes ces questions, nous devons nous interroger sur les acteurs qui composent ces deux systèmes.

I. Positionnement stratégique du citoyen dans le domaine des données

La structure que nous allons décrire se développe autour de trois parties prenantes. Reliés les uns aux autres, les acteurs qui la constituent sont ancrés sur l'Open Data et/ou le Big data. Parmi eux, nous comptons l'Etat et les organismes environnementaux, les entreprises et associatifs ainsi que les citoyens. Représentés sur l'Annexe A Schéma conceptuel Open Data / Big Data, nous allons présenter leurs places et leurs rôles.

1. Les acteurs autour du citoyen

a. Etat et organismes officiels

Les organisations actives sur la thématique des données ouvertes sont multiples et de natures variées. En France, le principal acteur est Etalab et la CNIL (Commission national informatique et libertés). La mission Etalab coordonne l'action de l'Etat pour « faciliter la réutilisation de données publiques » (9). Elle a mis en place une plate-forme unique compilant plus de 25000 jeux de données sur son site internet (www.data.gouv.fr). Etalab s'inscrit donc dans une démarche Open Data. Elle collabore avec des entreprises privées, les Agences rattachées du Ministère de l'environnement (Agence de l'eau, Ademe, Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL) ...), des start-ups et des établissements scolaires pour faire émerger des bases de données et des projets innovants utilisant les données publiques (que nous développerons ultérieurement). Toutes ces organisations diffusent en ligne des données.

b. Entreprises

Comme vu précédemment, les entreprises jouent un rôle concernant l'Open et le Big Data. En effet, les entreprises privées se positionnent dans le domaine des données comme

détenteurs, producteurs et potentiels réutilisateurs de données ouvertes. Deux catégories sont observables. Celles qui sont dans une démarche de services publics et celles dans une démarche commerciale. Pour les premières, parmi les figures de proue en France, nous pouvons citer Orange, et notamment la SNCF et la RATP pour la diffusion de données environnementales (émissions de dioxyde de carbone relatives aux liaisons TGV (10) (11)). La création de service et le partenariat avec d'autres acteurs (Gouvernement, start-up, établissements scolaires..) font partie de leurs démarches. Les autres se profilent sur une démarche d'Open Innovation ou Big Data davantage positionnés sur un axe commercial autour de la donnée produite par le citoyen en les récupérant et en la traitant.

c. Associations

Autres acteurs dans l'écosystème data, nous trouvons les collectifs et autres « Think tank » - traduction en français de « laboratoire d'idées » - (regroupant un ensemble de professionnels experts autour d'un même sujet). L'objectif de ces groupes est de faire évoluer le sujet d'expertise qui les concerne afin de faire émerger de nouvelles solutions. Elles ont aussi comme objectif de faire croître leur influence. Parmi ces collectifs, nous pouvons citer la Fondation Internet Nouvelle Génération (FING), Open Street Map (OSM), ou encore LiberTIC. Le rôle des associations est de faire le lien entre le citoyen et les données. Son fonctionnement est participatif, volontariste et cherche avant tout la prise en main des problématiques par le citoyen, pour le citoyen.

d. La Recherche

Les universités et le domaine de la Recherche sont impliqués dans l'étude des données environnementales par l'utilisation de la technologie de volume. Les applications sont variées : étude des polluants dans l'air par le programme SIGLES (Systèmes d'Information Géographique et Liens Environnement-Santé (12), des sols avec des groupements d'intérêt scientifique spécialisés dans le domaine de l'environnement, comme le GISSOL (13), l'évaluation de l'impact d'un projet d'urbanisme, par exemple. Compte tenu des données publiques et de celle produites par les centres de recherches scientifiques, les chercheurs détiennent un pool de données qui sous certaines conditions pourraient alimenter les bases de données publiques existantes. En ce sens, la Recherche représente une réserve non négligeable et un vivier de données pour l'innovation.

Comme vu précédemment, ces 2 concepts (Open et Big data) sont connexes et riches d'un potentiel informationnel à exploiter. Les quantités ne manquent pas et sont sans cesse en croissance en raison des nombreux canaux de production. Ces sources transitent

par des “noeuds” dont un nous intéresse plus particulièrement. Dans la partie suivante, nous étudierons de plus près le rôle du citoyen dans la gestion de données environnementales, et comment il est en mesure de les intégrer pour développer sa gestion du risque environnemental.

2. Placement du citoyen dans “l'écosystème data”

Il est intéressant de constater que ces deux concepts (données ouvertes et Big Data) sont liés et possèdent un dénominateur commun. Il s'agit du citoyen et de sa place dans cet éco-système numérique complexe.

Comme dans tout écosystème, il répond à des règles de fonctionnement et réagit avec les acteurs. Dans le but de comprendre ces interactions, référons-nous une nouvelle fois à l'Annexe A Schéma conceptuel Open Data / Big Data. Au centre de ce système, le citoyen peut être mis en relation avec l'Open Data. Dans ce cas, les données générées, produites par les organismes et/ou associations sont diffusées en Open Data sur les plateformes internet en libre accès ou sous conditions. Dès lors, ces données suivent deux trajectoires : elles sont utilisées par le citoyen directement ou elles peuvent être reprises par des personnes tierces compétentes, pour les modifier ou les réutiliser dans un autre but. Cela peut être pour de la visualisation de données, du data-journalisme, des applications à destination d'un public ou la Recherche. Les échanges pouvant aller dans les deux sens entre citoyens, organismes et tiers, le rapport de force se voit transversal. En effet, à la fois producteur volontaire et involontaire de données, le citoyen est un acteur central. Nous reviendrons sur ces réutilisations plus tard dans notre développement.

Par ailleurs, nous pouvons mettre en parallèle les rapports entre le citoyen et le Big Data, inscrits dans un système plus vertical, et plus unilatéral. Les données et méta-données générées par les utilisateurs des services sont collectées à travers les services qu'ils utilisent. Ces informations sont stockées et souvent échangées entre les firmes détentrices de ses jeux de données au sein du réseau. Les données subissent une analyse et des traitements algorithmiques selon les objectifs recherchés par leurs détenteurs (marketing, géomarketing, management,). Par conséquent, elles sont redirigées vers le citoyen pour alimenter des services.

II. En tant qu'individu / citoyen, j'ai accès à quoi ?

1. Qui peut avoir accès aux données et où?

L'accès aux données publiques axées santé et environnement se fait sur les plateformes internet. Les données qu'elles contiennent sont disponibles pour tout à chacun sous différents formats. Sous forme de carte interactive ou de base de données, des services d'extraction annexes sont aussi disponibles sur quelques plateformes. Dans ce cas, elles sont principalement sous les formats JSON (JavaScript Object Notation), CSV (Comma-separated Values) ou XLS (Excel). Ces formats se différencient en fonction leurs caractéristiques d'usages (en référence à l'échelle des 5 étoiles de Tim Berners-Lee (14)). Au plus la donnée est ouverte, libre et reliée à d'autres données mais aussi contextualisée (informations sur son auteur, raison de sa production...), au plus elle gagnera d'étoiles sur ce barème (Annexe C). Aussi, quelles données sont accessibles et où les trouve-t-on?

Dans le but de déterminer les données disponibles, une étude sur le contenu des plateformes de base de données a été réalisée. La méthode d'investigation a été la suivante: dans un premier temps la liste des portails Open Data en France a été consultée sur le site "data.gouv.fr". Ensuite, une recherche plus large des portails a été menée sur internet et le site du Ministère de la transition écologique.

Sur le site "data.gouv.fr", puisqu'il n'existe pas de nomenclature précise pour répertorier et classer les données, les jeux de données se rapprochant des thématiques de gestion des risques santé-environnement ont été recherchés. La consultation de tous les jeux de données disponibles a été faite autour des mots clés suivant: "environnement, qualité de l'air, du sol, pollution, développement durable, santé...". Sur la totalité des données du site gouvernemental près de 1% des données englobent ces thématiques. En élargissant nos recherches à d'autres sources, nous pouvons avoir accès à toutes sortes d'informations.

Pour choisir quelle plate-forme peut satisfaire les besoins de l'utilisateur nous avons élaboré des Arbres de décision pour choisir sa plateforme de données (Annexe D). L'objectif de ce document est d'aiguiller les utilisateurs vers la base de données adéquate. Cela se décompose en 3 étapes. Premièrement, nous décidons de la thématique: l'eau (qualité des cours d'eau, cheminement des eaux superficielles...), le sol (qualité des sols, typologie des sous-sols), la qualité de l'air (pollution aux particules fines, oxydes d'azote...), les cosmétiques (liste des ingrédients par exemple). Ensuite, nous devons définir le type

d'utilisation de la plate-forme (en bleu): c'est à dire l'action que nous souhaitons entreprendre: la consultation (la plate-forme proposera une navigation simple des données), la consultation/production (l'utilisateur pourra s'informer et/ou contribuer à la construction d'une base de données, la consultation / extraction (l'utilisateur du site pourra télécharger les données en plus de les consulter) et enfin l'extraction. En fonction des besoins du citoyen, nous pouvons choisir une plate-forme d'utilisation ou un listing de portails. Ces derniers répertorient des bases de données concernant la thématique choisie. Enfin, l'utilisateur peut se diriger vers le site dédié (en jaune). Les cadres tout à droite caractérisent le site choisi comme plate-forme de cartographie/ de visualisation ou comme base de données. Par exemple, si nous souhaitons naviguer sur une plate-forme des données proposant de la cartographie, nous choisirons le site Géoportail. Si nous souhaitons consulter et télécharger des données environnementales de différents types, le site Open Data Store est alors à privilégier (Figure 2).

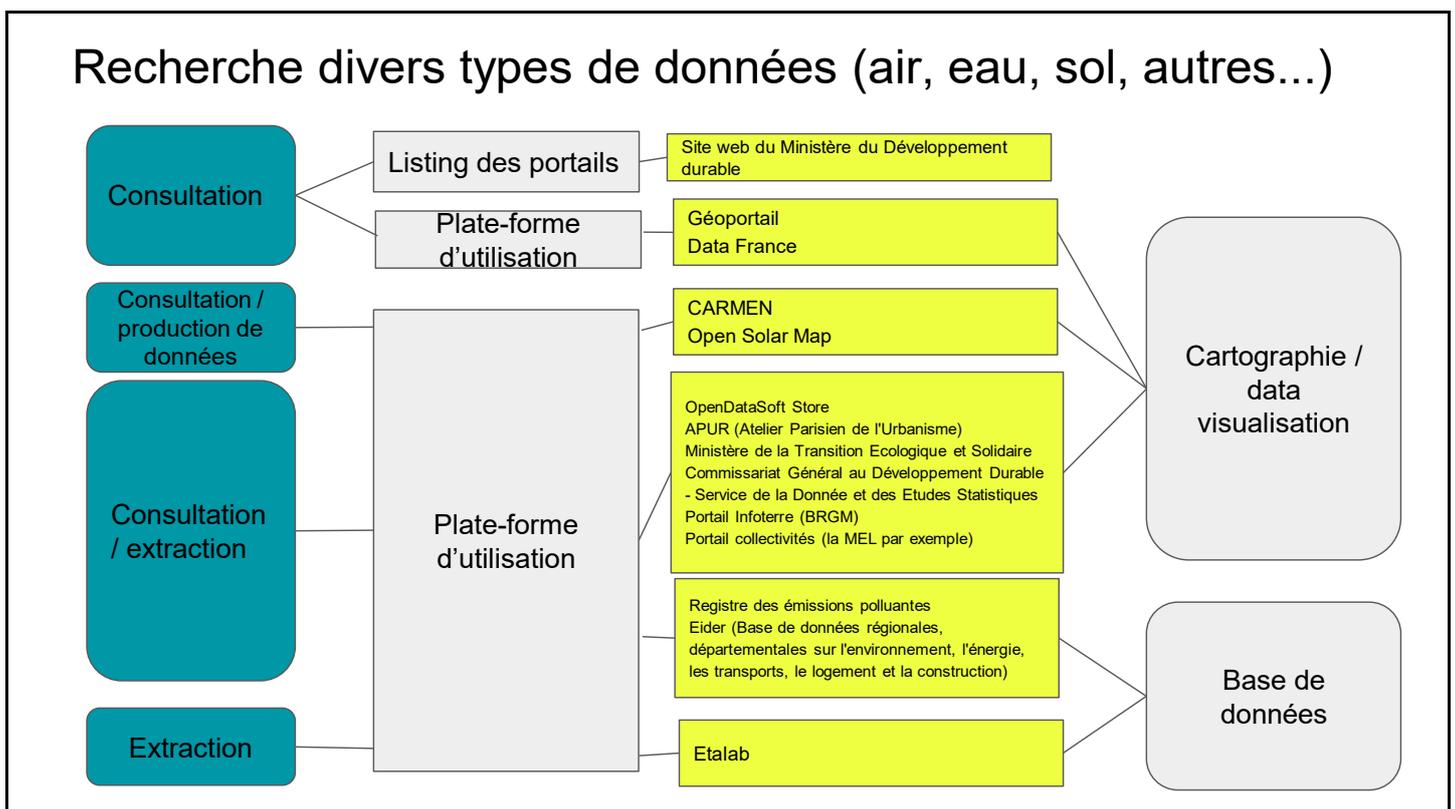


Figure 2 Exemple d'Arbre de décision - Recherche divers types de données environnementales

Afin d'accompagner le citoyen dans sa démarche, le Tableau des plateformes Open Data en santé-environnement (Annexe E) affiche une description des données disponibles sur près de 30 sites répertoriés.

2. Comment se procurer de nouvelles données?

Pour obtenir de nouveaux jeux de données, nous devons nous tourner vers la CADA (Commission d'accès aux documents administratifs). Il suffit de lui transmettre une lettre (15). Cet organisme « rend des avis sur les demandes de communication faites par des particuliers, des entreprises ou des associations aux Administrations » (7). Elle accompagne les organismes de l'Etat et leur donne des recommandations sur le caractère communicable des documents. L'ouverture de certaines données au public peut être soumise à des restrictions ou des autorisations sous conditions. La CADA ne donne pas suite aux demandes de documents relatifs à la sécurité du territoire, aux données personnelles et aux propriétés intellectuelles détenues par un tiers (15). Néanmoins, il peut y avoir des exceptions pour la diffusion et ce sous certaines conditions (anonymisation, droits d'auteur...).

III. Concrètement, quels usages par le citoyen des données santé-environnementales?

Dans le but d'obtenir des informations pertinentes, l'utilisateur des données se doit de penser en termes de projet. Quels investissements (temps, résultats attendus..) est-il prêt à faire ? Quels sont ses moyens et outils (capacité de traitement forte ou faible, seul ou en groupe) pour atteindre ces objectifs ? Quel est l'objet de sa recherche ? Suite à ces interrogations, il pourra choisir le mode de représentation qui lui permettra de faire ressortir un maximum d'information (cartographie, data-visualisation, services web, application mobile ...).

Il existe de multiples usages possibles des données publiques pour le citoyen. En effet, des mécanismes d'alertes en cas d'événements à risque pour la santé sont opérationnels. Nous pouvons citer les alertes ATMO en cas de pic de pollution par SMS/emails (16). Les outils actuels fournissent aussi un suivi temporel de la qualité de son environnement. A l'aide de graphiques nous pouvons visualiser l'évolution dans le temps de la qualité de l'eau, de l'air. Il existe d'autres exemples assez simples d'utilisation mais depuis les années 2000 et l'essor du numérique, c'est bien via la cartographie que le citoyen a accès à l'Open et Big Data en santé-environnement.

1. Cartographie un outil pour une gestion des risques santé-environnement? (Comment représenter mes données)

a. Origine du modèle

Historiquement, l'utilisation des données et des outils statistiques permettait de produire des études épidémiologiques. L'avantage de ces grands champs de données était de couvrir des échantillons importants pour obtenir des résultats représentatifs possédant des facteurs d'incertitude faible.

L'utilisation de données cartographiques est apparue bien longtemps avant l'avènement des outils informatiques. Son application la plus célèbre dans le spectre de la santé environnementale remonte au 19^{ème} siècle en Grande-Bretagne. Nous sommes dans le quartier de Soho dans la capitale britannique. John Snow, un médecin britannique, observe de nombreux cas de choléra. Dans son ouvrage « On the mode of the communication of cholera » paru en 1855, il recueille des informations cliniques et répertorie les lieux d'habitation des patients malades sur une carte (Figure 3). L'utilisation d'un plan a permis dans un premier temps, la localisation des victimes du choléra afin de définir un

périmètre d'exposition. Dans un second temps, de localiser une source potentielle de contamination. Par la suite, il décida pour tester une hypothèse de fermer le puit (source polluée) dans lequel les habitants malades avaient l'habitude d'y prendre l'eau.

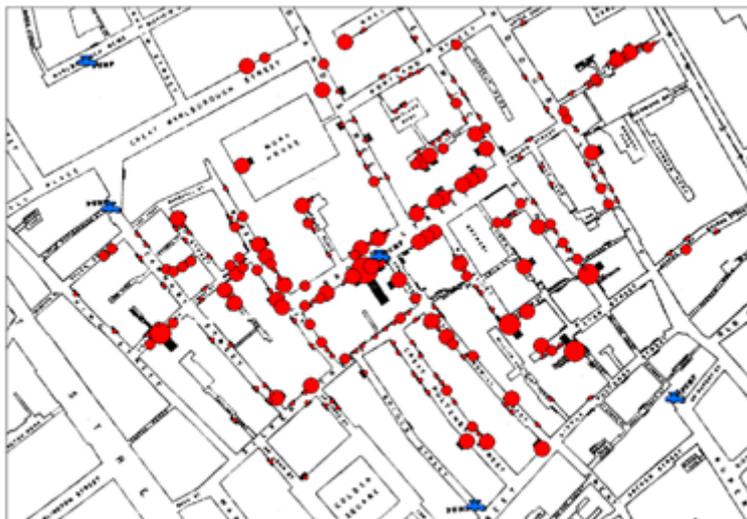


Figure 3 Cartographie des cas de choléra par le Docteur John Snow dans le quartier de Soho (1855)

Cette mise en quarantaine du point d'eau a fortement réduit les nouveaux cas de choléra. Par cet exemple, nous pouvons constater que les données (coordonnées exactes des patients, des points de captage d'eau, habitudes de vie) ainsi que le fond de carte (nom des rues, tracé des rues) ont permis d'avoir un impact positif sur la santé de cette population londonienne. Cet exemple traduit la méthode pouvant amener à une mise en place d'un système de gestion environnementale.

b. De la cartographie papier à la cartographie numérique

L'utilisation des cartes s'est développée avec l'émergence des outils informatiques et des logiciels SIG (Système d'Information Géographique). Le SIG est selon la définition de Marcel Bergeron «une discipline ayant pour objet la gestion des données à référence spatiale et qui fait appel aux sciences et aux technologies numériques d'acquisition de stockage de traitement et de diffusion » (17).

Concrètement, le SIG est un moyen permettant de modéliser des cartes superposant des couches successives de données d'origines et de natures variées (Figure 4) dans le but d'apporter une lecture transversale et complémentaires des informations. Ces strates de données peuvent être les réseaux (autoroutes, électriques..), les infrastructures (bâtiments, équipements scolaires par exemple), les éléments du biotope (eaux, surfaces végétales...). La compilation de ces données permet de produire des cartes complexes dont l'analyse et l'interprétation pourra offrir un angle nouveau d'explication des phénomènes.

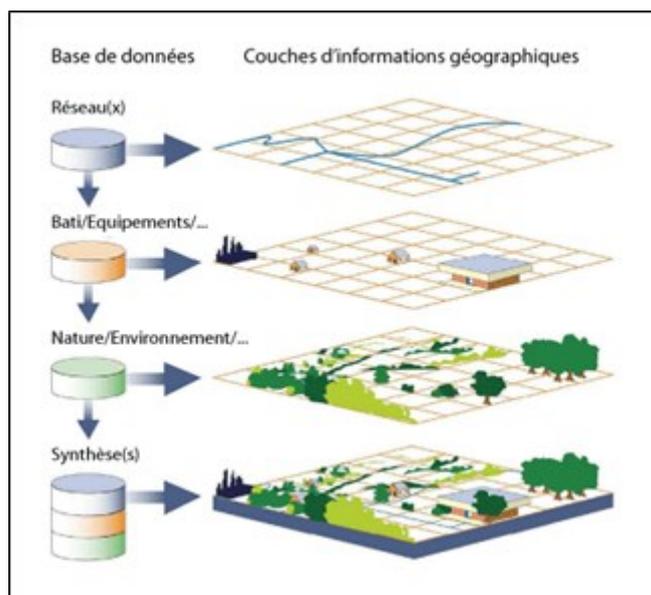


Figure 4 Schéma de superpositions des couches géographiques en SIG
(Source: Institut en conseil et d'études en développement durable - ICEDD - icedd.be)

Par son usage de bases de données privées comme publiques, cette discipline géographique s'inscrit dans une vision de numérique de la gestion des problématiques environnementales. En effet, cette technologie peut être utilisée pour établir une cartographie des différents niveaux de pollution liée à une ou plusieurs substances chimiques sur un territoire (

Figure 5) ou encore établir des liens entre une source nuisible pour l'environnement et les contaminations observées dans le milieu. L'étude des phénomènes environnementaux par le média cartographique offre une complémentarité d'information. Elle améliore la compréhension des événements. A ce titre, plusieurs initiatives existent déjà à ce jour. Nous pouvons par exemple mettre en avant différentes démarches dans ce domaine

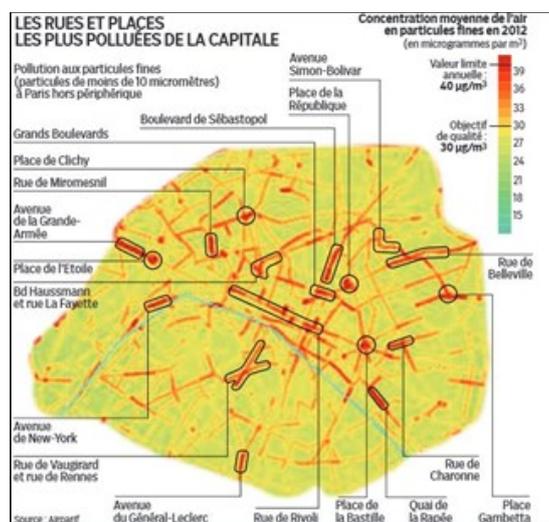


Figure 5 Cartographie des rues et places les plus polluées en particules fines (mg/m3) de la capitale (Source: Le Parisien / Airparif 2012)

2. Exemples d'outils de cartographie pour un usage du citoyen

a. Géoportail - outil de cartographie

Géoportail est une plate-forme cartographique développée par l'IGN (Institut National de l'Information Géographique et Forestière). Il s'agit d'un outil regroupant plusieurs fonds de carte. Il permet à l'utilisateur de rechercher, produire et partager des cartes. La base de données recouvre de multiples thématiques environnementales (occupations des sols, zones protégées, risques...) (18).

b. Les Assistants numériques - outils d'alerte

L'accessibilité aux informations utiles au quotidien se manifeste principalement par l'usage des appareils mobiles types smartphone ou tablette. C'est par ce canal que nous nous informons de plus en plus. La construction des analyses des données de stations fixes de surveillance de la qualité de l'air permettent de mettre en place des outils didactiques pour l'utilisateur. Intégrant des paramètres sur les mouvements, transports des polluants, ils développent des technologies qui ont la volonté de mettre en main un outil de décision personnalisé pour limiter / contrôler son exposition aux polluants atmosphériques. Pouvant alerter sur la qualité de l'air, ils pourront accompagner l'utilisateur sur les activités à privilégier, les itinéraires à emprunter, le moment pour aérer son habitat en fonction des niveaux de pollution. C'est par exemple le défi que s'est lancé Plume Labs avec son dispositif nomade appelé "Flow" (19) et l'appli "Plume Air report" ou Netatmo avec le "Healthy home coach" (20). Plume air report est une application mobile qui alerte et prédit les niveaux de pollution atmosphérique dans son agglomération. Flow et Healthy home coach sont deux dispositifs réalisant des évaluations des niveaux de pollution sur un secteur restreint et l'alerte de son utilisateur en fonction de la localisation. Ceci permet d'obtenir une cartographie de niveaux pollutions dans sa proximité.

L'utilisation de nouveaux outils de collecte de données par le citoyen permettrait d'ouvrir la recherche sur d'autres terrains d'exploration. En effet, ceci développerait deux approches pouvant être complémentaires dans l'étude des concentrations de polluants à l'échelle individuelle. De manière passive et à destination du public avec des applications comme Plume air report, ou l'assistant de Netatmo ; plutôt proactive et transversale avec l'utilisation de capteurs mobiles à l'instar de Flow, le second projet de Plume labs.

c. Des citoyens actifs dans une démarche de production:

Par le crowdsourcing:

Cette manière d'aborder la donnée place l'individu en tant que producteur de données brutes et récepteur de données traitées. D'autres possibilités d'agrémenter les bases de données lui sont offertes par des systèmes collaboratifs.

Il peut se tourner vers des plateformes de crowdsourcing et/ou de gamification. Le crowdsourcing est un système de recueil d'information en ligne dans lequel un groupe d'individus consolide ensemble une base de données. La gamification (ou ludification) est un moyen de rendre ludique, amusante une tâche complexe pour la rendre plus attractive et motiver son usage. Elle s'inspire des codes appliqués dans la réalisation de jeux vidéos. C'est un exemple de démarche qu'Open Solar Map (OSM) - un projet open source de cartographie de l'ensoleillement des toitures. A partir d'une photo satellite aléatoire d'un bâtiment, OSM demande au participant de son "jeu" d'indiquer l'orientation de la toiture parmi 3 choix (nord-sud, est-ouest, non applicable) (Figure 6). En collectant les orientations d'un panel d'utilisateur, sur plusieurs centaines voire de milliers de bâtiments, l'algorithme est capable de ressortir une évaluation cartographique du potentiel énergétique des toitures sur un territoire.

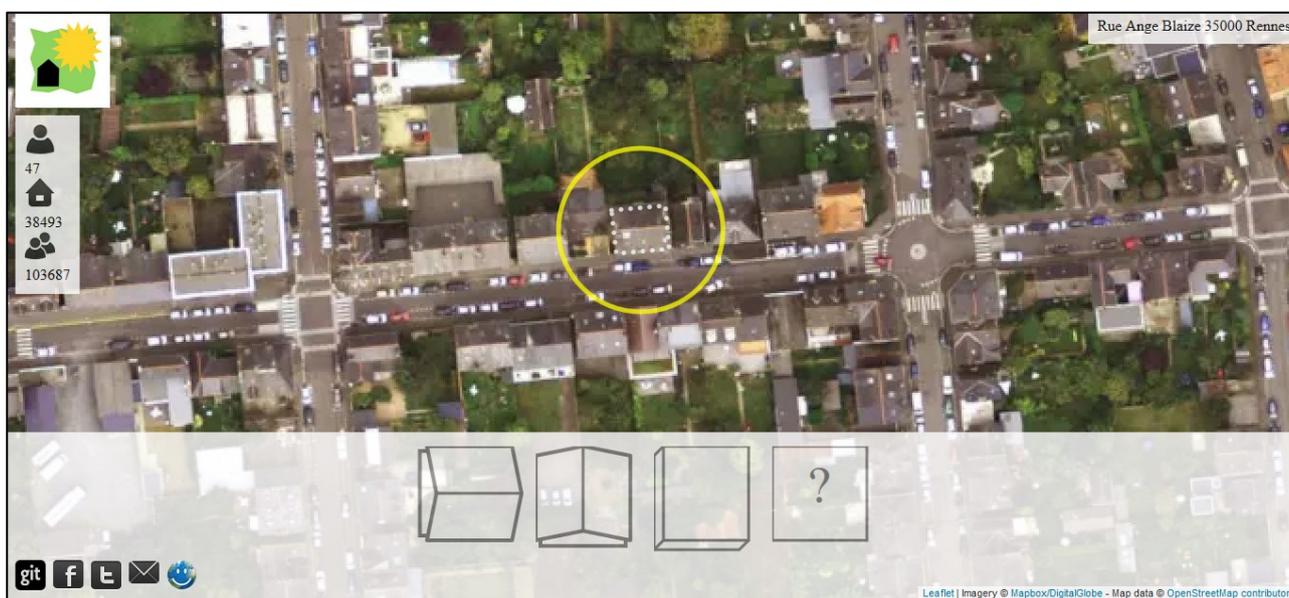


Figure 6 Capture d'écran de l'interface d'Open Solar Map

Son utilité finale sera de déterminer les habitats les plus propices à recevoir des panneaux solaires. A travers cet outil, nous découvrons que le citoyen peut jouer ici un rôle important en apportant une contribution directe dans la constitution de la base de données. Un des avantages de ce modèle est la réduction des ressources financières. Par ailleurs il

encourage la réalisation de l'action, la consolidation continue de la base de données. De plus, l'objectif de ce projet est aussi de mettre en avant une problématique de gestion environnementale. Revers de la médaille, tributaire de la qualité des photos et des choix réalisés par les contributeurs, ce modèle peut manquer de précision.

Hackathon

Autre moyen existant pour le citoyen et les données environnementales, le hackathon (ou codefest) est un modèle favorisant la création et l'innovation par l'utilisation de la donnée. Ces manifestations sont ouvertes pour tous et engagent les participants dans un concours. Contraction de *hack-* (que l'on peut traduire par "programmation bricolée" et de *-marathon*, l'objectif de ces évènements est de réunir des personnes (ingénieurs, développeurs, spécialistes ...) autour de la création de nouveaux outils numériques seul ou en équipe, et ce dans un laps de temps assez court (en général 2/3 jours). Ces projets sont par la suite évalués par un jury décidant des meilleurs. Non restreint aux problématiques environnementales, ce concept à l'avantage de faire croiser des compétences techniques (informatique) et des compétences spécialisées (sur le thème du hackathon : santé, environnement, énergie...).

Preuve que le hackathon est productif, début 2017, le Ministère de l'écologie a organisé un concours ayant pour objectif d'élire le meilleur projet de data-visualisation sur les pesticides dans les eaux souterraines (21). Près de 50 candidatures ont été proposées et 9 projets ont été validés par le jury pour la délibération finale. Parmi eux, Hydro-viz (22), 2ème lauréat de ce hackathon. Ce site internet permet aux citoyens de consulter les niveaux de pollutions des eaux souterraines aux pesticides en fonction de plusieurs paramètres. Ils varient selon la période de temps (2007 - 2014), la région géographique ou de la nappe phréatique. Cet outil interactif montre les répartitions des polluants en fonction de leurs familles, fonctions et sur de leur dangerosité. Cela offre la possibilité de visualiser simplement le niveau de nocivité de ses produits utilisés dans son quotidien et en faire des comparatifs. Enfin autre plate-forme interactive, data-pesticides.fr (23) affiche les niveaux de concentrations sous différent formats (par station, par département) et laisse le choix à son utilisateur de sélectionner le(s) substance(s) dont il souhaite connaître les informations.

La force de ces nouveaux outils de data-visualisation est dans leur capacité à apporter du sens à la donnée. En la remettant dans son contexte, cela apporte une perspective supplémentaire et permet une meilleure objectivité de l'information extraite. Elle guide, informe et fait s'approprier la thématique à des publics (producteur ou utilisateur du service)

qui peut-être ne sont pas sensibilisés au sujet en question. La donnée peut jouer un rôle social dans les interactions humaines qu'elle génère dans ce contexte.

Enfin, la visualisation des données environnementales sous cette forme (concrète) constitue un puissant outil de communication. En effet, elle sert à développer la diffusion de messages en santé environnement (en terme de gestion des risques) à destination du grand public, mais aussi pour faire comprendre les intérêts et les enjeux de la production, de la conservation et de l'ouverture des données.

IV. Limites de l'usage des données santé-environnementales par le citoyen

L'ouverture des données environnementales, sous ses aspects ouverts, progressistes et innovants a néanmoins quelques inconvénients. Il s'agit des incertitudes de la donnée, des questions d'accessibilité, d'usages et des compétences, et nous allons les analyser.

1. Etat des lieux des données environnementales - Absence de certaines données

Le choix de diffusion de données en relation avec les thématiques santé-environnementales est un processus complexe et réfléchi. A ce jour, toutes les données ne sont pas mises en ligne. En effet, même si la démarche est considérable de la part du gouvernement (25 000 jeux de données sur le site data.gouv.fr toutes thématiques confondues), certains jeux ne sont pas mis en ligne et pourraient faire l'objet d'une ouverture (ex: données sur les intoxications au monoxyde de carbone, consommation de pesticides des communes).

En Février 2016, des journalistes de France 2 de l'émission "Envoyé spécial" révèlent une carte des consommations de produits phytosanitaires sur le territoire métropolitain (24). Grand succès médiatique, le programme a réuni 3,3 millions de spectateurs (25). La particularité de cette carte provient de son origine. En effet, elle a été construite à partir d'une base de données confidentielles. Contacté par téléphone, un reporter a expliqué ne pouvoir diffuser les données brutes afin de ne pas trahir l'identité de sa source. Par ailleurs, un argument de secret commercial a aussi été avancé pour restreindre sa diffusion. A travers cet exemple, nous mettons en avant les contraintes possibles pour leurs mises en ligne.

Ces lacunes peuvent être expliquées de plusieurs manières par des contraintes de sécurité, de confidentialité voir par l'absence de données. En effet, la diffusion des données se réalise sous conditions de certains critères (26) : le caractère sensible (liés au caractère anonyme des données et aux risques relatifs à la sûreté), aux intérêts pour des réutilisations et la volonté politique. En effet, pour ce dernier point, nous pouvons comprendre que la diffusion de données diminuant l'image d'une commune (exemple un site contaminé) soit un facteur restrictif pour son attractivité.

Par leurs intérêts, nous constatons aussi que des entreprises souhaitent créer de nouveaux outils dans le domaine numérique via l'axe environnemental (application mobile,

partenariats, innovation...). C'est une manière de développer leurs activités et leur image. Ceci transparaît dans une étude BVA de 2011 qui annonce que les entreprises sont pour 40% d'entre elles favorables à partager des données environnementales. En effet, les acteurs privés considèrent l'ouverture des données comme une opportunité dès lors qu'elle s'inscrit dans une stratégie d'innovation (27).

2. Incertitudes intrinsèques de la donnée

Le second facteur d'incertitude lié directement à la donnée est sa fiabilité, sa qualité. En l'absence de normalisation des données pour la construction des bases de données, ceci engendre irréversiblement des manquements de mise en forme (quels formats, quelle est la qualité de ces dernières). Ceci ne veut pas dire néanmoins qu'elles sont inexploitable. En effet, grand nombre des bases de données disponibles sont issues d'organisations gouvernementales ou accréditées (ADEME, Agence de l'eau...) et gage de qualité. Pour finir, nous devons songer à la fréquence de mise à jour de données. Ce paramètre en tant qu'indicateur qualité doit aussi être pris en compte.

Dans ce cadre d'ouverture massive de la donnée et face à cet appel à la contribution du citoyen, il est légitime de s'interroger sur son impact vis à vis de la qualité de son action. Quelles sont les incertitudes liées au contributeur (source) / réutilisateur?

Dans le cas du crowdsourcing par exemple, pour élargir son audience il est important de mettre en place des systèmes de contrôle pour s'assurer de la qualité d'information apportée et de simplifier au maximum la démarche afin d'encourager les participations. En effet, il sera plus aisé de mobiliser des grandes populations de contributeurs par une demande simple, reproductible et ludique que l'inverse.

3. Accessibilité, usages et compétences

Face à toutes ces données, un utilisateur peu averti peut s'y perdre. Pour parvenir à produire par exemple une réutilisation des données, il lui faut détenir des compétences spécifiques (programmation, statistiques, utilisation de tableur...). Ce besoin de maîtrise des logiciels abouti à une restriction d'exploitation des données en raison de la complexité des logiciels. Et cela réduit donc le nombre de personnes pouvant être motivées par leur utilisation. Les profils sont a priori plutôt complexes et rares à trouver. Ce besoin de pluralité en ce qui concerne les compétences peut cependant se compenser par une association de profils différents mais surtout complémentaires. Ce qui explique la nécessité de rencontre et de regroupement autour de projets collectifs comme les hackathons par exemple.

Cependant, pour palier à ce manque de compétences des citoyens, des spécialistes développent de plus en plus d'outils interactifs permettant de visualiser et d'interpréter ces données sans connaissances a priori.

En étudiant le chiffrage des réutilisations des données environnementales sur le site Etalab (Mai 2017), nous constatons que les données sont très peu réutilisées dans leur ensemble. Sur les 297 jeux environnementaux, 273 jeux n'ont jamais été réutilisés (et déclarer sur le site), et seulement 22 l'ont peu été (les deux autres concernent des thématiques qui ne rentrent pas dans notre sujet). Ceci démontre plusieurs choses. La première que les données environnementales diffusées ne sont peut-être pas assez exploitables et ce pour plusieurs raisons (compréhension de la donnée environnementale - à quoi elle correspond concrètement, spécificité territoriale de la donnée, absence de moyen pour une réutilisation, difficulté d'utilisation comme vu précédemment ...).

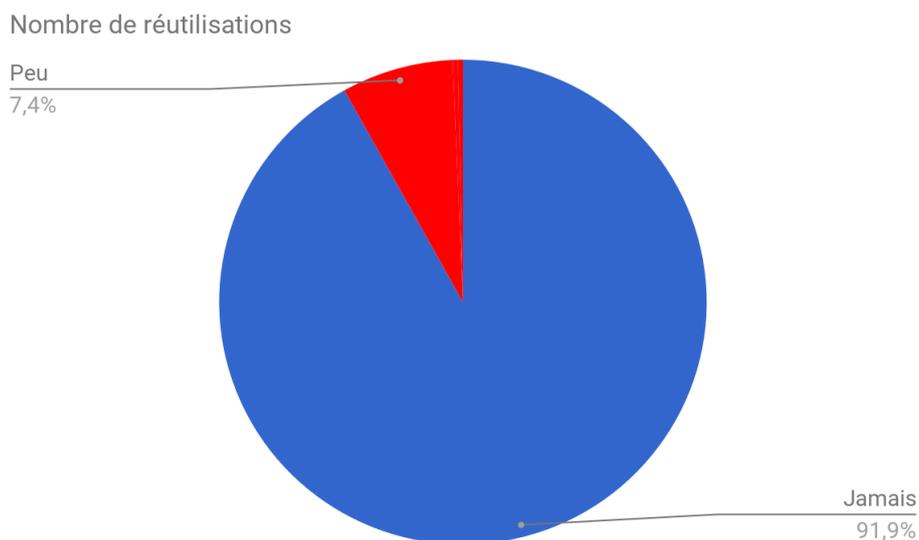


Figure 7 Utilisation des données environnementales sur le site Etalab

La faible proportion de réutilisation peut aussi trouver son explication dans la production institutionnalisée de la donnée (28). Pour reprendre la pensée d'Antoine Courmont (Doctorant à Science Po), produites dans un contexte par et à l'usage de la structure spécifique, " les données restent marquées de leur empreinte institutionnelle (dans le format et la structure) ce qui restreint le champ d'utilisation possible de ces données. Par ailleurs, les données sont plus facilement accessibles et complètes sur les plates-formes spécialisées.

Conclusion Partie 1

Les données disponibles sur les différentes plateformes ainsi que les potentiels usages qui peuvent en être faits préfigurent des avancées pertinentes en termes de gestion des risques santé-environnementaux. Prévention, gestion personnalisée des expositions, analyse de phénomènes (avec plus de précision dans les applications)... Les acteurs sont présents, les applications concrètes multiples et cependant, il reste encore beaucoup de choses à développer dans ce domaine.

Encore trop peu connu par le grand public, l' "objet data" pour rentabiliser sa production (au sens d'usage) aura besoin d'une solide compréhension de son utilisation et ce, dans sa globalité. C'est ce qu'on appelle la Data-literacy. Le rôle du citoyen malgré les implications actuelles doit encore trouver sa place. Cette compréhension nécessitera en parallèle un accompagnement aux connaissances des enjeux Santé environnement. Son ancrage dans l'écosystème data dépendra des outils qui pourront marquer son action sur son environnement.

Par ailleurs, puisque les enjeux environnementaux sont liés les uns autres, l'usage des données collectées sur un sujet peut amener à des données utiles pour d'autres problématiques. C'est dans cet esprit d'ouverture, de croisement d'informations entre les organisations et les citoyens que des transformations des modes de gestion s'engageront. Ceci passera aussi par une plus grande utilisation des objets connectés dans l'écosystème des données.

Enfin, comme tous modèles, la convergence de l'utilisation de données environnementales pour et par des citoyens nécessite des améliorations à plusieurs niveaux. Dans un premier temps, l'exactitude de la donnée doit rester garantie tout au long du processus de conception de l'outil pour assurer la qualité du service. Deuxièmement, un débat approfondi entre les parties prenantes devra mettre sur la table les sujets, problématiques que chaque acteur souhaite voir diffuser, et discuter des conditions d'usage des données. Enfin, en raison de la complexité de la manipulation des données brutes, des compétences spécifiques nécessaires et des productions réalisées à ce jour, le rôle du citoyen vis-à-vis des données environnementales sera précieux mais reste encore mal dessiné car pose de nombreuses questions quant à son impact.

PARTIE 2: Perspectives d'avenir par le prisme du citoyen: impacts multisectorielles des datas SE

Nous l'avons vu précédemment, les données représentent un potentiel encore trop peu exploité aujourd'hui. Les centaines de jeux disponibles sur les réseaux nécessitent la mise en place de collaboration entre les acteurs pour fédérer des actions utiles à la gestion des risques et soulèvent de nombreuses questions. Comment pouvons-nous entrevoir l'avenir du mouvement data concernant les données environnementales? Quels en seront les impacts politiques et économiques? De quelles manières ces changements vont-ils se répercuter sur le champ social? Le développement des méthodes de traitement des données aura-t-il un impact positif sur l'environnement? Enfin, comme le cadre législatif va-t-il évoluer?

I. Politique

1. Transparence et empowerment: un mouvement politique émergent

A ce jour, les données environnementales sont assez marginales dans le domaine public mais des démarches sont possibles pour inverser la tendance. Nous pouvons déjà observer une volonté de transparence en légère croissance. En effet, même si la Commission d'Accès aux Documents Administratifs a mesuré une baisse des demandes de 1% entre 2012 et 2014 pour la thématique Environnement / Développement durable / transport et plus globalement sur toutes les thématiques, nous observons une remontée du pourcentage de demandes en 2015 (Figure 1 Figure 8). Parallèlement, cette évolution trouve son sens dans ce que nous pouvons appeler "l'empowerment".

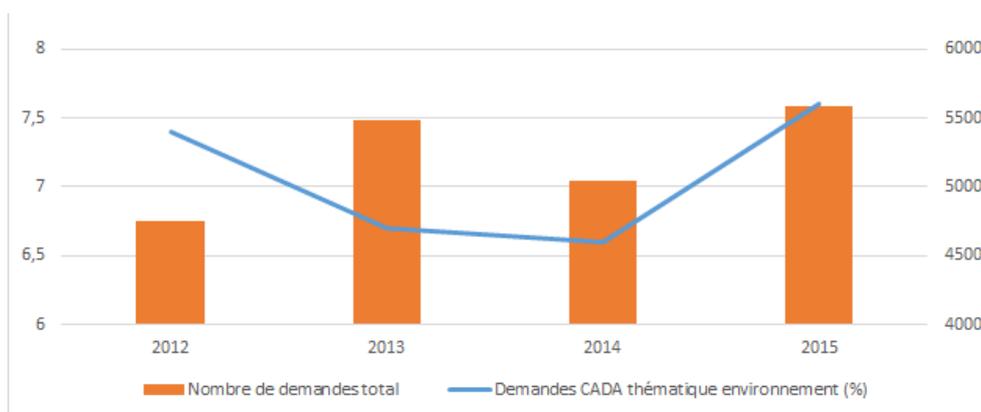


Figure 8 Evolution du nombre de demande d'accès aux données publiques à la CADA et répartition en pourcentages des demandes autour de la thématique environnementale (Source - Rapports d'activités annuels entre CADA 2012 - 2015)

L'empowerment est un mouvement cherchant à donner plus de pouvoir aux populations par la mise en place d'initiatives et par la collaboration entre des organisations. Il s'agit de l'acquisition d'une partie du pouvoir décisionnel par les populations locales. Cela offre la possibilité de mettre sur le débat publique des enjeux préoccupant pour les populations.

Dans son article - "L'empowerment, un concept pour la France ?" (29) - Carole Dane défend l'idée qu'en France, "le travail avec les habitants est souvent conçu en terme d'animation et de projets, alors que les processus d'empowerment et de développement social vont au-delà". Dans cette logique, une démarche de gestion des risques plus intégrée par le citoyen ouvrirait certaines opportunités de voir la population se saisir des problématiques autour de l'environnement (pollution atmosphérique, qualité des eaux, santé...). Cette volonté de réappropriation pourra se mettre en place dans le cadre de la Loi pour une République numérique. Nouveau levier d'action, elle donnera des outils pour sensibiliser et alerter sur des thématiques environnementales.

Ceci se manifesterait par exemple par l'augmentation plus forte des demandes de diffusion de données environnementales. Et plus précisément par des demandes spécifiques relatives à des problématiques de santé publique (ex: pesticides, cosmétiques, polluants émergents...). Plus des données seront mises en ligne, plus des démarches santé-environnementales les utilisant pourront être développées. Face à cela, des actions ont été mise en place, encore aujourd'hui par le Ministère de l'écologie. En juin 2016, une plateforme de data - visualisation a été mise en ligne par exemple (30). A travers cet exemple nous constatons la nécessité de créer des passerelles entre les institutions productrices de données et les citoyens.

Nous pouvons imaginer une utilisation des données par le citoyen comme moyen de pouvoir demander des comptes aux politiques si les démarches engagées n'ont pas donné les résultats attendus. En effet, la diffusion des données et la création de système numérique donne la possibilité aux citoyens de construire des indicateurs, des outils de contrôle vis-à-vis des actions promises par les collectivités. Cela nécessite des mises à jour plus fréquentes et complètes des informations. Cette démarche pousserait les citoyens à faire pression sur des problématiques de gestion des risques.

Enfin, s'il fallait encore douter que la thématique environnementale soit un enjeu politique, nous pouvons observer que cette question nourrit toujours les débats et inquiète (31). En effet, à l'heure où les Etats-Unis se sont retirés de l'accord de Paris pour le climat et que son organisation gouvernementale (USEPA) voit ses dotations diminuées par l'administration Trump, d'autres manœuvres sont en cours. Parmi celles qui nous intéresse,

un travail d'archivage et de contrôle des données environnementales américaines par l'Environmental Data & Governance Initiative est en cours (32). Les acteurs de l'EDGI sont des citoyens qui se sont alliés pour s'opposer en prévention à la perte de ces informations sur le champ public. Cet exemple montre que l'appropriation de "l'objet data" par le citoyen est une forme d'empowerment, une manière nouvelle de marquer la gestion environnementale de son empreinte.

2. Intérêts politiques pour un développement de ces techniques de gestion sanitaire et environnementale

L'application de moyens de gestion numériques afin de d'obtenir une réduction des problématiques environnementales peut offrir des possibilités de développement pour une meilleure gestion appuyée. Elle pourrait s'inscrire dans une politique d'innovation pour garantir les intérêts publics. En effet, compte tenu de la puissance des technologies Big Data (en termes de précision, de modèles prédictifs par exemple) et de l'Open Data (avec une mutualisation des sources -crowdsourcing, hackathon), nous tenons ici une force pour accélérer les avancés scientifiques dans le domaine santé-environnementale. Ces progressions permettront peut-être de mettre en avant les problématiques sanitaires et environnementales.

Les bénéfices de telles actions seraient multiples. Tout d'abord la valorisation des données environnementales mises en libre accès et de l'utilité de leurs utilisations. De plus, cela rehaussera le pays dans la volonté de développement via de l'innovation. Open data et Big data sont des technologies prometteuses tout comme les objectifs politiques du Plan Nationale Santé Environnement 3 sont ambitieux sur la santé et l'environnement (33).

Par ailleurs, tous ces efforts pourraient représenter un avantage concurrentiel en termes d'image et de rayonnement sur la qualité de sa gestion environnementale. En effet, la France se trouve 10ème dans le classement général de l'Indicateur de Performance Environnemental (34) derrière des pays figurant comme modèles tels l'Islande 2ème ou la Finlande 1ère.

Pour cela, l'appui des Instances gouvernementales aux projets de développement des modes de gestion sera essentiel. Par exemple, il existe des projets d'Open Sciences. Certains sont menés notamment par la Commission européenne (35). Ils visent à faire collaborer les chercheurs entre eux et à développer la recherche scientifique dans différents

domaines dont les sciences environnementales. D'autres solutions sont mises en places pour inclure le citoyen dans cette démarche.

Cependant, les décisions politiques n'avancent pas au même rythme que les avancées sociétales, lesquels politiques plutôt réfractaires au changement. En comparaison des données d'ordre économique en marketing, et leurs usages dans les Big Data notamment, ne pouvons-nous pas observer un retard, peut-être même un manque d'ambition des gouvernements sur notre territoire? La réticence ne viendrait-elle pas d'une prise de conscience du manque à gagner et des enjeux économiques sous-jacents? L'impulsion viendra donc du citoyen et de sa capacité à s'organiser autour de la donnée environnementale.

II. Economique

A l'heure où les ressources naturelles sont de plus en plus vite épuisées (Jour de dépassement de plus en plus précoce)

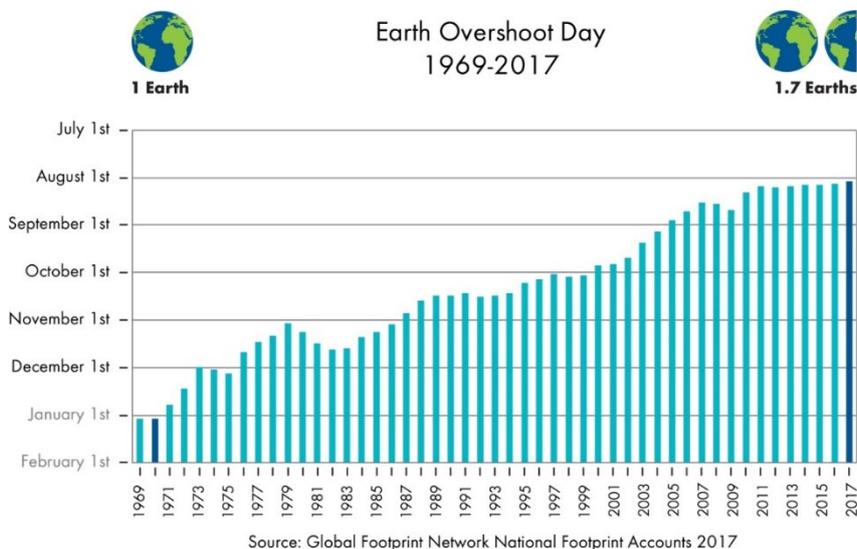


Figure 9 Evolution du Jour de dépassement entre 1969 et 2017

Les ressources naturelles représentent une richesse pour les pays les détenant mais aussi un coût. L'utilisation des données environnementales provenant de sources multiples (organisations, citoyens) et leur traitement par des systèmes numériques favorisent dans certains cas l'efficacité de leurs consommations. Ceci pourra apporter un cadre de vie de qualité supérieur, facteur d'attractivité de plus en plus recherché. Le développement des réutilisations de données environnementales offrirait un avantage concurrentiel, un gain en attractivité.

Par ailleurs, la protection de l'environnement représente des enjeux économiques notamment en termes d'image et d'investissements. En effet, nous savons par exemple que la pollution de l'air est responsable de 48 000 décès en France (36) et pèse financièrement. La pollution de l'air intérieur seule concentre un coût annuel de près de 20 milliards d'euros (37). Face à cela, sous réserve que des démarches soient mises en application, le gain global sanitaire en France pourrait être de 11 milliards d'euros d'ici 2030 (8), si des mesures de réductions sont atteintes. L'utilisation de la donnée peut aider à remplir cet objectif.

Concernant le citoyen plus directement, l'ouverture de ces données pourrait provoquer un effet favorable à la création de nouveaux métiers liés au numérique et à la gestion environnementale. Programmeurs, Data Scientist, Administrateurs de données sont autant de professions qui évolueront vers des domaines en relation avec l'environnement et sa gestion. Cependant, l'engouement ne semble pas encore se manifester complètement car

seulement 30% des entreprises interrogées dans cette étude étaient prêtes à investir dans l'Open Data. Leurs craintes sont principalement liées aux risques perçus par l'utilisation de cette technologie (image de marque, contrôle des réutilisations...) (27) (9) .

Nous l'avons vu précédemment, la démocratisation des outils mobiles notamment présente des avantages sur le plan du développement des données environnementales. Cette production individualisée et partagée de données relatant la qualité de l'environnement, l'exposition ou même la santé peut constituer une réelle opportunité pour des études santé-environnement à grande échelle (type cohortes). Cette évolution dans les pratiques pour mener une campagne de mesure à grande échelle et ce de manière très précise (spatio-temporel) coûte actuellement très cher pour un organisme. Peut-être que nous avons ici une opportunité de réduire les coûts ou de proposer des études avec des budgets plus réduits.

Plus ouvertes à un usage Big Data, les grandes firmes américaines jouent un rôle déterminant. Dans le domaine environnemental, Google est inscrit dans une démarche "verte" de réduction de ses impacts sur l'environnement. Dans cette optique, les ingénieurs de la multinationale ont intégré des capteurs de méthane sur la flotte de leur véhicule Google Street View pour mesurer les émissions des véhicules en mouvement (38). Ce projet a pour objectif de construire des cartographies rue par rue des zones étudiées. L'entreprise cherche peut-être ici à développer sa technologie dans une optique commerciale.

III. Social

1. Une prise de conscience des problématiques environnementales via les données par le citoyen

a. Prise de conscience du potentiel des données: acculturation.

Pivot central de l'écosystème des données, le citoyen est impacté par les Data dans ses sphères sociale et privée. Nous l'avons décrit précédemment, de grands flux de données sont échangés par et à travers lui, c'est un nœud d'information, un maillon dans la chaîne des données. Parmi les aspects de sa vie en mutation et notre relation aux données environnementales, nous pouvons prévoir l'émergence d'une prise de conscience citoyenne.

La prise de conscience des citoyens sera un processus long. L'assimilation des problématiques relatives à la gestion environnementale par le citoyen devra évoluer par un processus d'acculturation croisée (en combinant culture de la donnée et connaissances environnementales). Comment ce processus va-t-il se mettre en place et pourquoi?

b. Vers un développement participatif et des politiques locales

Le processus d'acculturation se réalisera par la rencontre des différents acteurs de l'écosystème, l'échange d'informations et l'établissement de nouveaux partenariats entre les porteurs de projets et les acteurs souhaitant les soutenir. Ce rapprochement tendra vers un développement participatif dont il faudra en créer les espaces. Ils se distinguent par la capacité de faire converger des personnes aux intérêts communs favorables à une nouvelle manière d'améliorer la gestion des risques et la mise en projets de leurs idées. Le nombre de démarches est à la hausse et le sujet semble se développer d'année en année dans la littérature (et ce dans divers domaines incluant l'environnement) depuis 2010 (39). La Figure 10 Evolution du nombre de publication d'études sur des projets de sciences citoyennes en fonction de la thématique entre 1997 et 2014) indique une augmentation des initiatives Open sciences environnementales. Cependant, le graphique de droite précise qu'elles ne reprennent pas des projets de ré-utilisations des données (data re-use en anglais).

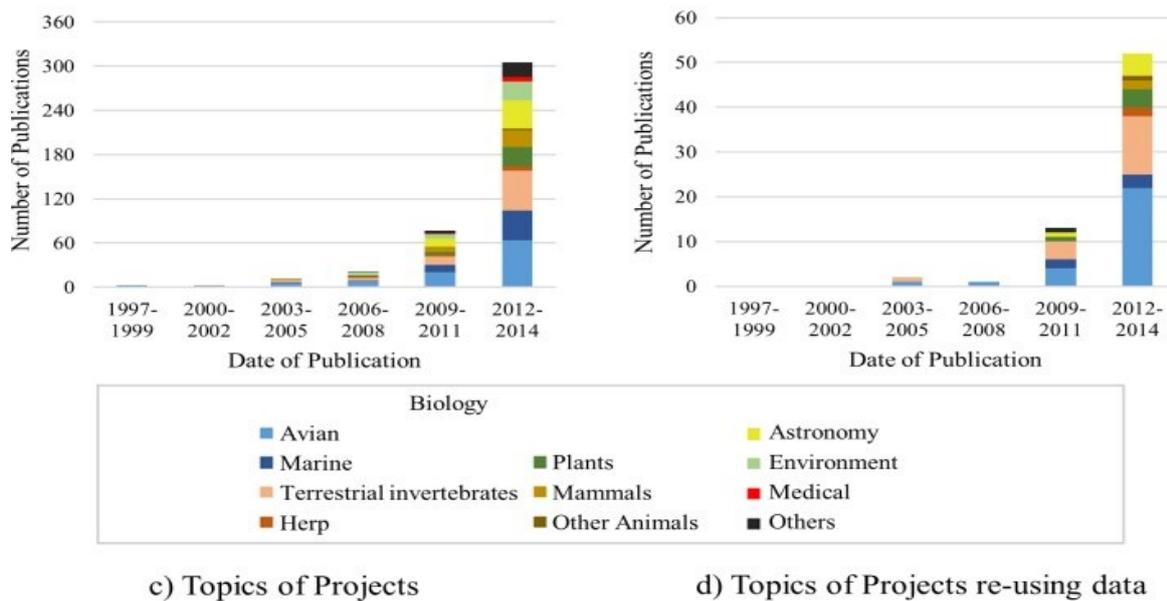


Figure 10 Evolution du nombre de publication d'études sur des projets de sciences citoyennes en fonction de la thématique entre 1997 et 2014 (Source : Follett, Ria; Strezov, Vladimir – 2015)

Exemple de processus d'acculturation : Infolab

Parmi les projets d'acculturation au monde des données, l'initiative "Infolab" de la FING fait figure de précurseur. Ce programme lancé en 2013 est un projet visant à promouvoir l'utilisation des données par tous les acteurs de la société (Figure 11). En allant communiquer directement avec eux, ils les sensibilisent, les font collaborer et les accompagnent dans leurs projets en mettant en valeur les données. Certains de leurs dispositifs s'articulent autour de la création de groupes de travail pluridisciplinaires dans lesquels se déroulent débats et sessions de sensibilisation. Cet exemple démontre que la coopération est primordiale et annonce la direction vers laquelle se dirige la gestion environnementale : une transition où numérique et environnement convergeront vers une gestion participative.

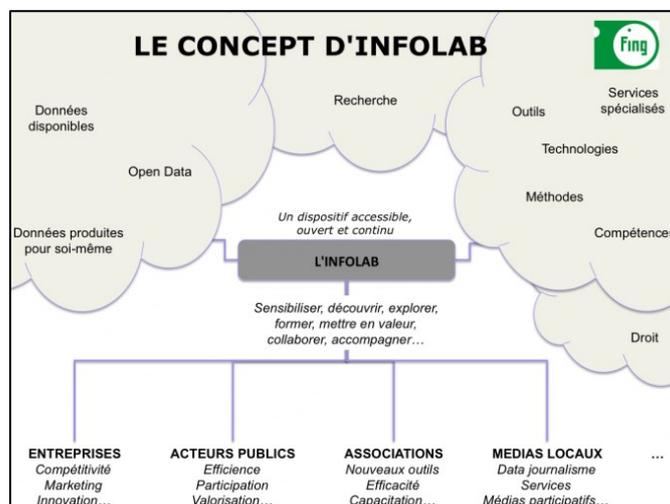


Figure 11 Concept d'Infolab (Source- FING)

Au sein du programme Infolab, se trouve l'initiative Transition²: C'est un exemple de projet pro-environnementale ancré dans le numérique. Il a pour objectif de "décloisonner les communautés de l'écologie et du numérique" (40). Axé sur plusieurs thématiques, cette plateforme collaborative met en libre accès près de 530 contributions produites par les membres ou extérieurs du dispositif. Nous pouvons y retrouver des boîtes à outils, des projets. La participation est libre et les projets de participer à des conférences ouvertes à tous.

c. Intégration des méthodes de gestion des données vers un public plus large : en marche vers la production de données?

L'objectif de développement de la gestion des risques par les données environnementales nécessite une ouverture des concepts et des connaissances au grand public. Comment le mettre en place? Comme vu précédemment, l'évolution de la gestion des risques environnementaux passe par l'acquisition des enjeux autour de la donnée (mode de production, utilisation...) ainsi que de nouvelles compétences numériques. Néanmoins, il paraît assez difficile de manipuler les données. Selon Daniel Kaplan, interviewé pour son rôle de co-fondateur de la FING (41), "la complexité de la technicité engendre un écartement du citoyen". Pour inverser cet état, il faut "passer par la médiation car les gens ne connaissent pas directement les données car peu sont au courant des portails" nous explique M. Kaplan. Ce qui explique la nécessité de mettre en œuvre une mission de communication vers le grand public en ce qui concerne cette thématique (exposition, sensibilisation, atelier...). Néanmoins, la participation d'individus sur ces projets de "science citoyenne" paraît être "prêt à être développé" sous réserve de moyens adéquats à leurs réussites (outils innovants de gestion de bases données, d'analyses statistiques) (42).

La manipulation de la donnée environnementale comme matière de réflexion peut représenter une "clé d'entrée", un nouvel angle pour intéresser les jeunes générations aux problématiques santé-environnementales. En effet, l'utilisation des données environnementales peut être un bon moyen de sensibiliser et transmettre des connaissances. A partir d'un certain niveau (lycée/université) il est envisageable de faire construire des cartes ou des data-visualisations (exemple en Annexe F) à l'aide d'outils pour rendre l'apprentissage plus stimulant pour les élèves et les enseignants" (43). Il s'agit encore d'une manière de développer la culture de la donnée environnementale, d'autant plus importante en terme de pédagogie que les décisions de gestion des risques

environnementaux dans notre société se fondent “à partir des données et où nous vivons dans un pays où cette variable n’est pas bien intégrée” (41).

2. Migration vers un suivi personnalisé des individus: le quantified-self ou l’individu objet data.

a. Concept

Le quantified-self (traduction de mesure de soi) est un concept qui couvre à la fois les méthodes et les outils permettant de faire des mesures de variables relatives à son mode de vie (44). Très utilisé pour des applications santé afin de mesurer le nombre de pas journalier, les prises alimentaires, sa fréquence cardiaque par exemple, le quantified-self possède des caractéristiques intéressantes pouvant être transposées vers la gestion des risques Santé Environnementale. La mesure de variables individuelles offre la possibilité de mettre en place des indicateurs personnalisés. Là où la gestion des risques actuelle affiche des informations à grandes échelles (pays / région / ville), le quantified self fonctionne sur des données propres à chacun (position GPS, rendez-vous, rythme cardiaque...). L’avenir sera peut-être de porter des capteurs mesurant les niveaux de pollution à proximité (exemple de Plume labs). Ces informations couplées à d’autres données non individuelles permettraient par exemple de cartographier d’une autre manière des zones géographiques et de nouveaux espaces (intérieur des habitations, rues, quartiers...). En effet, beaucoup d’applications de collecte de données sous ce mode utilisent la géolocalisation. De ce fait, l’application de cette approche révèle un impact potentiel sur la gestion des risques en santé-environnement en termes de production et de traitement des données.

b. Impact sur la gestion des risques en santé environnement

La participation individuelle de la collecte des données permettra de constituer des groupes. D’un point de vue expérimental, cela peut être intéressant pour réaliser des comparaisons entre ces derniers. Les différences de comportements des utilisateurs d’outils de quantification pourront faire l’objet d’études d’expositions (pollution atmosphérique notamment). A leur manière, ces groupes peuvent constituer des “citoyens capteurs” (45) un peu de la même manière des espèces sentinelles. En effet, l’étude des risques de nos jours prend en compte des variables d’exposition “moyennes” ou “normées”. L’apport de données individualisées correspondra mieux aux habitudes de vie de ce dernier (type de lieu de vie (urbain / rural) ; proportion de temps en extérieur ou non, temps passé dans les transports...).

En effet, nous ne sommes pas tous égaux face à l'exposition aux polluants. Cette configuration d'étude apporterait d'autres connaissances sur les circonstances expositions (classifications des publics (enfants, femmes/ hommes, milieu social), comportements, activités). L'un des avantages de l'utilisation de données et de pouvoir réaliser des croisements de bases de données. Ceci sera très utile pour évaluer des niveaux d'exposition à des polluants (entre les données ouvertes, privées). Cette application pourrait se situer pour les agents nocifs dans différents milieux et secteurs (air, eau, alimentation, cosmétiques...). En poussant le principe d'évaluation des expositions, nous pourrions exploiter les algorithmes pour construire des modèles prédictifs.

L'entrée des données dans le champ des pratiques citoyennes sera un changement dans la manière d'aborder la gestion des risques. Prenons l'exemple de la pollution de l'air intérieur. Sachant que premièrement nous passons près de 70% de notre temps en intérieur (logement, travail, activités ...) (46) et que deuxièmement l'air intérieur est souvent plus contaminé que l'air extérieur, la connaissance de l'exposition sur ce point paraît intéressante. La data peut en ce sens être un vecteur de développement. Par exemple, les applications Open food fact et Open beauty fact_(47) (48) proposent d'obtenir des informations sur les produits disponibles dans le commerce. Pour le premier sur des produits alimentaires, pour le second sur des cosmétiques. Une partie des informations ont été directement collectées par des citoyens pour garnir la base de données.

Dans un contexte où le citoyen sera à la fois producteur et utilisateur des données, nous pouvons imaginer une adaptation de la gestion des risques dans son quotidien. L'intégration des informations et des messages de prévention dans la gestion de ses activités se manifestera à travers des changements de comportement répondant à une volonté de réduction des risques. Aussi, la problématique de la gestion de données personnelles est un sujet important que nous développerons plus loin.

IV. Technique

1. Vers de nouvelles manières de produire de la donnée environnementale

Parmi les innovations techniques qui sont déjà en cours et/ou qui verront le jour dans les années à venir, le développement des outils portables (applications, moniteurs...) permettront de produire plus de données environnementales par le citoyen. Ce changement de mode de production sera le fruit de plusieurs sources technologiques. La première et la plus notable est la miniaturisation des dispositifs de mesures. En comparaison avec des stations de mesure de polluants atmosphériques (Figure 12), ces nouveaux outils seront transportables, et offriront une grande mobilité à leurs utilisateurs.

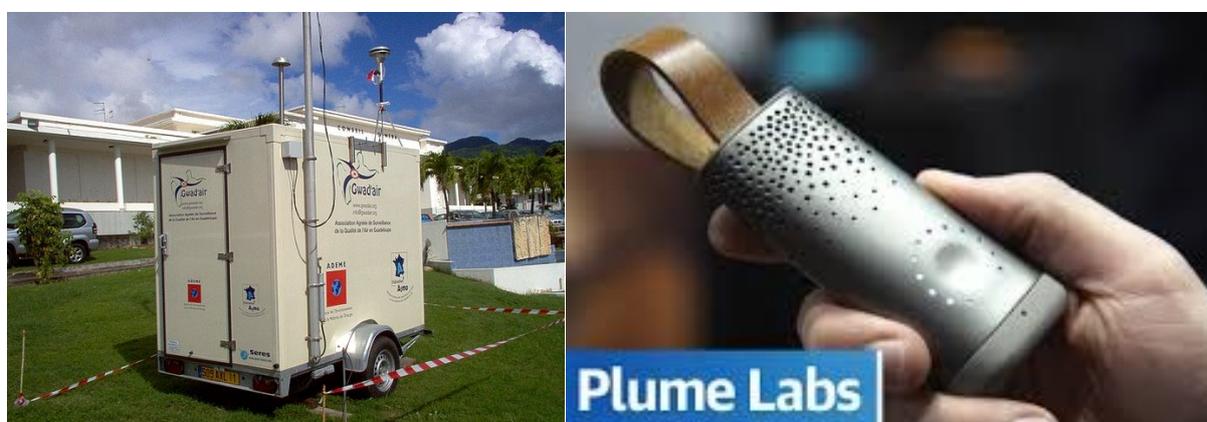


Figure 12 Exemples d'une Station de mesure mobile à gauche) et d'un appareil portable Flow à droite (Sources - gwadair.fr / Plume labs)

La miniaturisation de ces outils est notamment la conséquence des avancés techniques, de l'externalisation des calculs vers les Data Centers et l'utilisation du cloud. En effet, Cette décentralisation opère ainsi la délégation des tâches complexes vers d'autres installations (serveurs informatiques) et laisse l'interface être gérée par les capacités de la machine de l'utilisateur (car moins puissantes).

Autre progrès que nous pouvons entrevoir, l'augmentation du nombre de substances, de polluants pouvant être mesurés par ces dispositifs. Déjà des sociétés proposent de quantifier des radioéléments, l'électromagnétisme, la qualité de l'eau et la qualité de l'air (49). Reste que les performances des appareils existants à l'heure actuelle sont assez contredites. La mesure des polluants "classiques" n'est pas encore optimale.

Pouvons-nous alors anticiper les conséquences de tels changements? Toujours sur le plan technique, la multiplication de moyens mobiles pose la question de son rôle actif ou passif de l'utilisateur dans sa connexion avec l'objet de mesure. Durant les phases de prélèvement, l'utilisateur se retrouverait dans un mode passif si elles se font de manière

autonome ou de manières actives si elles sont enclenchées par le citoyen. Nous le voyons bien ici, les responsabilités du citoyen peuvent être multiples et motivées par une décision personnelle.

Les réponses techniques devront répondre à des besoins simples en termes d'utilisation et de compréhension pour minimiser les manipulations de données complexes. L'objectif principal de ces évolutions techniques sera d'accompagner l'utilisateur pour l'alerter, le conseiller pour mieux se protéger dans son quotidien. Enfin, une étude BVA indique que ces innovations se situent sur un terrain favorable, car si 81% des "Français sont soucieux de leur santé et de celle de leur proche", "ils se montrent assez réceptifs à un potentiel apport des nouvelles technologies pour améliorer leur quotidien : plus des $\frac{3}{4}$ d'entre eux se déclarent intéressés par le recours à des objets connectés (dont ceux de santé à 48%)" (50).

2. Citoyen producteur de données Non Structurées: quelles opportunités en santé environnement?

Autre source à exploiter, les données non structurées sont un vivier d'information. Les données non-structurées sont toutes celles ne pouvant pas être inscrites dans des bases de données classiques. En font parties, les données textuelles et les images par exemple. Comment ces données (qui vont provenir de ressources variées) vont pouvoir converger vers un usage bénéfique pour la gestion santé environnementale?

Pour cela, nous pouvons reprendre l'exemple d'analyse que Google avait réalisé pour prédire des pics de grippe en étudiant les résultats des recherches web. Le Google Flu est un service qui prend comme matière de traitement les requêtes du moteur de recherche. Dans un contexte environnemental, nous pourrions imaginer une analyse des requêtes portant autour des problématiques santé-environnement. Ceci pourrait donner vers une ouverture d'espaces dédiés à la communication et à la discussion (chat, forum,...). Nous utiliserons ainsi les données des citoyens comme matériau d'étude.

V. Écologique

1. Impact sur la gestion des risques en santé environnement

L'entrée des données dans le champ des pratiques citoyennes sera un changement dans la manière d'aborder la gestion des risques. L'utilisation des données permet l'établissement d'un état des lieux continu de l'environnement. Illustré par des indicateurs, cela génère des outils de suivi des améliorations. A titre d'exemple, l'indicateur environnementale (EPI) de l'Université de Yale (34) sur le volet "impacts sur la santé de l'exposition environnementale", définit un score d'exposition. Il prend en compte les facteurs de potabilité de l'eau, de l'assainissement, de la pollution de l'air intérieur, aux particules et à l'ozone. Aujourd'hui la France se situe à la 37ème place avec un score de 88%. Ce genre d'indicateur peut ouvrir à la construction d'autres indicateurs nationaux et aussi individuels pour mesurer son exposition.

La mesure d'indicateur simple et compréhensible pour chacun offre une visibilité des problématiques environnementales dans le quotidien avec les outils numériques. Les nombreuses applications des outils de données santé-environnement dégageront des bénéfices pour l'accessibilité, la pédagogie et l'information en faveur de la santé publique. Ce sont les ingrédients nécessaires à l'amélioration de la connaissance du citoyen aux risques sanitaires. Peut-être que ce rapport numérique à l'environnement lui donnera un intérêt supérieur dans les années à venir pour mieux le respecter.

2. Impact environnemental de l'usage des données

Revers de la médaille, l'utilisation de données et donc de ressources informatiques n'est pas sans conséquences. C'est un fait, les Data Centers Big et Open Data consomment de l'énergie pour deux raisons. Premièrement pour leur exploitation et deuxièmement pour le fonctionnement de leurs unités de refroidissement. En effet, les serveurs produisent de la chaleur et son accumulation peut endommager le matériel. Aussi, il est important de faire chuter la température de ces locaux.

A l'heure où la réduction des consommations d'énergie est recherchée, il est assez difficile d'évaluer les quantités représentatives de l'usage des données environnementales. Cependant, une croissance globale des consommations énergétiques liés à l'exploitation des serveurs est prévue (51) . Dans le but d'intégrer toujours plus le citoyen, certaines initiatives visant à les réduire sont intéressantes comme celles qui souhaitent récupérer la

chaleur produite par les serveurs pour chauffer ses espaces (logement, établissements publics, entreprises...) (52). Ceci permettrait d'inscrire l'utilisateur de ces serveurs-radiateurs dans un cercle vertueux. Plusieurs autres exemples existent, comme celui de la société Stinergy qui en partenariat avec l'Université de Lyon-3 à développer un système similaire pour alimenter les ballons d'eau chaude d'un gymnase limitrophe (53). Cette dernière solution couvrirait 65% des besoins en alimentation en eaux chaudes (54). L'évaluation de l'empreinte carbone et l'efficacité de ces nouvelles formes de chauffages paraît néanmoins intéressante à réaliser.

Autre contrainte, qui dit données, dit interfaces de collecte. Si la demande d'appareils de quantified-self est à la hausse depuis quelques années, il est envisageable que cette croissance concerne aussi les outils de mesures environnementales personnalisés. Ceci produirait *in fine* plus de déchets, plus de consommations d'énergie.

VI. Législatif

1. Données environnementales: éthique et responsabilités pour les utilisateurs

La législation actuelle de référence pour les données fait suite la direction décrite lors de la Convention d'Aarhus (1998). Elle est légiférée par le texte de Loi 2016 - 1321 dit "loi pour une République numérique » présentée par Mme Axelle LEMAIRE, secrétaire d'État chargée du numérique sous le mandat de François Hollande. Ce texte a parmi ces objectifs de : " favoriser la circulation des données et du savoir" (55), c'est-à-dire de renforcer et élargir l'ouverture des données publiques ; créer un service public de la donnée ; introduire la notion de données d'intérêt général, pour permettre leur réutilisation par tous et enfin développer l'économie du savoir et de la connaissance. ».

Ce texte fait référence à la Directive européenne INSPIRE (56). Cette directive demande aux acteurs publics d'harmoniser la collecte et la diffusion de données environnementales d'ordre géographiques. Elle demande aussi de fournir les moyens pour réaliser des cartographies pour les citoyens de l'Union Européenne.

Face à l'apparition de ces nouvelles pratiques, des questions éthiques se posent. Quelles sont les responsabilités qui incombent aux utilisateurs de données? Si le citoyen souhaite réutiliser les données environnementales, il doit en respecter les clauses. L'article 10 de la loi CADA autorise la réutilisation de toutes les données, sans altérer la nature ou le sens de ces dernières (15).

2. Possession et protection des données (personnelles et non personnelles)

La protection des données est assujettie à la Directive 95/46/CE de la Commission Européenne et la loi Informatique et Liberté de 1978 par la CNIL.

La CNIL est l'instance de protection et de contrôle des données personnelles. Elle a pour mission l'information, l'accompagnement des particuliers et des entreprises en matière de données personnelles. Elle assure aussi un rôle d'agent de contrôle et de sanction. Enfin, c'est un acteur incontournable sur les thématiques liées aux innovations et en termes de veille technologique concernant les données. Le non-respect des règles de la CNIL est passible de sanctions pénales d'emprisonnement et pouvant atteindre des amendes de 300 000 € (57).

Aujourd'hui la réglementation des données publiques impose une anonymisation des données publiques. Elle a pour objectif d'empêcher l'identification des individus. Ceci concerne donc aussi les données environnementales. En revanche, la question devient plus intéressante quand il s'agit des données récoltées par le Big Data. En effet, quelles sont les garanties en termes de protection des informations?

Nous l'avons vu, les entreprises Big Data communiquent avec les citoyens et les entités Open Data qui fournissent des informations. La protection de l'anonymat des populations peut se voir compromise par le croisement des bases de données. Selon un rapport du Sénat, dans une mission d'information, le risque est "supérieur en raison des entreprises privées qui traitent des données massives" (58).

Rappelons qu'un géant du web comme Google s'est saisi du sujet des données environnementales en cartographiant la pollution atmosphérique à travers les rues de certaines agglomérations (59). De ce fait, nous pouvons projeter que les données ouvertes, des données produites par des entreprises et des données privées (personnelles) soient compilées et potentiellement revendues à d'autres acteurs. Ces derniers pourraient être des assurances qui à terme calculeront un risque environnemental et sanitaire pour ajuster leurs prix. D'ailleurs l'exploitation de certaines des données de santé (nombre de pas, fréquence cardiaque, poids...) sont déjà en cours chez certaines assurances. En d'autres termes, il s'agirait d'une approche individualisée de l'étude risques de l'assuré prenant compte des données environnementales. Dans une société connectée (montres, voitures, domotique...), l'analyse de ces informations environnementales pourrait changer les rapports entre les citoyens et leurs assurances.

Dans notre réflexion, les progrès techniques et leurs applications personnalisées posent la question du quantified-self et de la protection des données. En effet, ce sujet illustre de manière évidente que des données personnelles peuvent-être transformées pour un suivi individuel. Très liées à ce sujet, les données de santé sont aussi à considérer si l'on se pose dans la question de la gestion des impacts environnementaux. En effet, les utilisateurs des services quantified-self transfèrent des informations sanitaires à ces services. Ces échanges sont réglementés dans la loi concernant les données de santé. Une réglementation plus forte devra empêcher le croisement des données permettant d'obtenir des informations identifiant le citoyen.

Aussi, quelles sont les garanties que ces données ne soient pas détournées pour d'autres usages? Nous ne développerons pas davantage les aspects de sécurité. Cependant, il est intéressant d'observer que des piratages informatiques ont lieu sur le

territoire touchant tous types de structures (industries, Etats, hôpitaux,..). Des données de santé piratées par des personnes malveillantes sont par ailleurs disponibles sur les Dark net (60)(32) (partie du réseau internet non référencée et accessible uniquement à l'aide de logiciels spécifiques d'anonymisation contenant du contenu illégal). Les organisations développant des données environnementales (et sanitaires) ne seront pas épargnées par cette menace.

Conclusion Partie 2

Les données entraînent et vont entraîner des modifications dans nos sociétés à plusieurs niveaux. Les influences des mouvements Open Data en politique en sont un exemple. La gestion des risques et l'usage des nouvelles technologies utilisant les data en sont une autre. La tendance qui se profile annonce une gestion environnementale d'un côté plus centré vers les utilisateurs. Elle se focalisera sur une information adaptée à l'utilisateur en fonction de ses besoins propres (habitudes de vie, déplacement, activités...). Nous nous dirigeons donc vers une gestion personnalisée des risques environnementaux.

Cette transformation devrait être accompagnée par une multiplication des sources de données par l'intermédiaire des organisations, des citoyens eux-mêmes par l'intermédiaire d'outils connectés. L'accessibilité plus simple aux données et leur traitement encouragent la compréhension des problématiques environnementales et l'éducation des populations à ces sujets riches et complexes.

De natures malléables, les données pourront se greffer à d'autres disciplines (sociales, économiques..), faire des ponts pour améliorer la recherche, s'intégrer plus en profondeur dans notre société et découvrir de nouvelles applications notamment dans le domaine médical (61). Néanmoins, leurs utilisations devront faire l'objet d'un cadre permettant de faire respecter les droits individuels comme la confidentialité des données ou la sécurité (dans le cas de l'usage de données personnelles).

CONCLUSION GENERALE DU MÉMOIRE

A l'instar de tous les phénomènes ayant eu un impact retentissant sur nos habitudes quotidiennes, les avancées majeures de nos sociétés ont traversé plusieurs phases. D'abord "ridicules, puis dangereuses, et enfin évidentes", ces adjectifs se sont appliqués à de nombreuses révolutions comme le dit Idriss Aberkane dans sa conférence sur l'économie de la connaissance (62). L'apport mutuel d'informations prôné par ce concept est au cœur du processus Open Data. Le Big Data lui, moins ouvert à un rapport de force égalitaire, reste néanmoins un outil puissant de traitement et d'analyse pour le champ environnemental.

Compte tenu des nombreuses bases de données existantes (en ligne et privées), les thématiques environnementales les plus importantes d'un point de vue de la santé publique sont concernées. Pollutions atmosphériques, et de l'air intérieur, contaminations des eaux souterraines sont des exemples majeurs de sujets pouvant être traités grâce et pour le bien-être des citoyens. Quoi qu'il en soit, il demeure important de conserver cette matière première, les bases de données pour leurs usages futures (recherche, analyse prédictive notamment).

Face aux problématiques environnementales connues aujourd'hui et celles émergentes, les populations ont la nécessité de se prendre en main et de mettre en place des moyens de se protéger. Pour citer Didier Gourc *et al.* (63), "il est préférable qu'une prise de conscience intervienne le plus tôt possible. Il s'agit alors de sortir de l'impasse de l'inaction" - Ceci se traduit par un besoin de s'organiser, de réunir et de produire des nouveaux outils car les données sont diversifiées et sans cesse renouvelées. La compréhension des enjeux des données environnementales est capitale. C'est par un travail d'information du public que nous impliquerons plus favorablement chaque individu.

Les applications des données environnementales en gestion des risques vont dans les années à venir mettre en avant de nouvelles interrogations et problématiques dans notre conception de la gestion des risques. Vu sous l'aspect grandissant des enjeux des données dans le quotidien, leurs utilisations ont été décriées et le sont relativement encore aujourd'hui. Quelles seront les nouvelles données à diffuser en Open Data, quel avenir quant à l'utilisation des données personnelles, comment le quantified-self impactera l'utilisateur dans son rapport avec l'environnement? Ces interrogations soulèveront les bases d'une nouvelle façon de concevoir la gestion environnementale.

Pour reprendre le concept d' "homo data sapiens" de Simon Chignard (6), nous nous situons à une époque dans laquelle l'homme est défini par les données qu'il produit. Nous nous dirigeons vraisemblablement vers une société plus connectée à son environnement. Urbanisme, circulation automobile, politiques environnementales, consommations, impacts sanitaires et environnementaux (liés aux changements climatiques, polluants émergents...) toutes ces thématiques auront leurs volets de gestion modifiée par la donnée. L'innovation à travers le traitement des données environnementales changera la gestion des risques en particulier sur notre conscience à l'exposition aux polluants.

En effet, dans la perspective où il sera sans doute possible dans les prochaines années de compiler, croiser et de traiter toutes sortes de données (individuelles produites volontairement ou non: environnementales, de santé, identité...) nous pourrions imaginer la possibilité de produire, de définir des "profils environnementaux d'exposition" individualisés, au plus proche des citoyens. En ce sens, les données environnementales seront un pool de ressources pour l'amélioration d'une gestion des risques plus centré sur l'individu, ses habitudes de vie, son quotidien.

Bibliographie

1. Organisation Mondiale de la Santé (OMS). L'insalubrité de l'environnement provoque 12,6 millions de décès par an. [En ligne] 15 Mars 2016. [Citation : 9 Septembre 2017.] <http://www.who.int/mediacentre/news/releases/2016/deaths-attributable-to-unhealthy-environment>.
2. EDITORIAL TEAM. The exponential growth of data. INSIDE BIG DATA . [En ligne] 17 Février 2017. [Citation : 9 Septembre 2017.] <https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>.
3. DOMO. Data never sleeps 3.0. [En ligne] <https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/>.
4. Arrêté du 22 Décembre 1981 sur l'enrichissement du vocabulaire informatique. Site internet du Bulletin de la Bibliothèque de France. [En ligne] 22 Décembre 1981. <http://bbf.enssib.fr/consulter/bbf-1982-06-0355-009>.
5. Open Data Institute - European Commission. E-learning. European data portal. [Online] <http://www.europeandataportal.eu/elearning/fr/module1/#/id/co-01>.
6. Chignard, Simon. Comprendre l'ouverture des données publiques. s.l. : FYP Editions, 2012.
7. Commission d'accès aux documents administratifs - CADA. La CADA. Site web de la CADA. [En ligne] [Citation : 9 Septembre 2017.] www.cada.fr.
8. Assemblée Nationale. Projet de Loi pour une République numérique. [En ligne] 2015 Décembre 2015. <http://www.assemblee-nationale.fr/14/projets/pl3318.asp>.
9. Etalab. Qui sommes-nous? Le Blog d'Etalab. [En ligne] [Citation : 9 Septembre 2017.] www.etalab.gouv.fr/qui-sommes-nous.
10. Société Nationale des Chemins de fer Français -SNCF. Plate-forme Open Data SNCF. [En ligne] [Citation : 9 Septembre 2017.] <https://data.sncf.com/>.
11. Régie Autonome des Transports Parisiens - RATP. Plate-forme Open Data RATP. [En ligne] [Citation : 9 Septembre 2017.] <https://data.ratp.fr>.
12. FACULTÉ DE PHARMACIE DE LILLE LABORATOIRE DES SCIENCES VÉGÉTALES ET FONGIQUES. Le programme SIGLES : Systèmes d'Information Géographique et Liens Environnement-Santé. [En ligne] [Citation : 9 Septembre 2017.] <http://www.sigles-sante-environnement.fr/>.
13. Gis Sol. Présentation. [En ligne] [Citation : 9 Septembre 2017.] <https://www.gissol.fr/le-gis>.
14. Information sur le programme Open Data 5 étoiles de Tim Berners-Lee. Open Data 5 étoiles. [En ligne] [Citation : 9 Septembre 2017.] <http://5stardata.info/fr/>.
15. BORDEAUX MÉTROPOLE. Open Data facile: Guide de demande de données publiques auprès des collectivités. [En ligne] Janvier 2015.

http://www.opendatafrance.net/wp-content/uploads/2012/02/guide_de_demande_open_data.pdf.

16. Atmo Hauts-de-France. Abonnez-vous aux alertes. Site internet d'ATMO Hauts-de-France. [En ligne] [Citation : 13 Septembre 2017.] <http://www.atmo-hdf.fr/abonnez-vous-aux-alertes.html>.

17. PORNON, Henri. SIG - La dimension géographique du Système d'Information. Paris : Dunod, 2011.

18. Institut national de l'information géographique et forestière - IGN. Géoportail. [En ligne] <https://www.geoportail.gouv.fr>.

19. Plume Labs. Flow, by Plume: the first smart air quality tracker. [En ligne] 2017. [Citation : 13 Septembre 2017.] <https://plumelabs.com/en/>.

20. Netatmo. Healthy Home Coach | Capteur de qualité de l'air. Site internet de Netatmo. [En ligne] 2017. [Citation : 14 Septembre 2017.] <https://www.netatmo.com//fr-FR/product/aircare/homecoach>.

21. Ministère de la Transition écologique et solidaire. Le concours de data-visualisation sur les pesticides dans les eaux souterraines. [En ligne] 7 Décembre 2016. [Citation : 14 Septembre 2017.] <http://www.ecologique-solidaire.gouv.fr/concours-data-visualisation-sur-pesticides-dans-eaux-souterraines>.

22. PARIS, Julien et MELKI, Florian. HydroViz - beta 0.2. [En ligne] 2016. [Citation : 14 Septembre 2017.] <http://www.hydroviz.fr/>.

23. Découvrez la carte des pesticides en eaux souterraines en France. Data pesticides. [En ligne] 2016. [Citation : 14 Septembre 2017.] <https://www.data-pesticides.fr>.

24. Cash Investigation - Produits chimiques : nos enfants en danger. France 2, 2016.

25. Lopes, Benjamin. Audiences TV du jeudi 2 février 2017 : Envoyé spécial à haut niveau. [En ligne] 3 Février 2016. <http://www.toutelatele.com/audiences-tv-du-jeudi-2-fevrier-2017-section-de-recherches-leader-sur-tf1-envoye-special-a-haut-niveau-tmc-puissante-88285>.

26. Valentin, Jérémie. Les données environnementales : un cas particulier dans la mise en place des données publiques en Open Data ? Netcom, 27-1/2 | 2013 Les données environnementales en libre accès. 01 Septembre 2013.

27. Bluenove, BVA. Open Data : quels enjeux et opportunités pour l'entreprise ? 2011.

28. Courmont, Antoine. Open Data et développement économique - Comment assurer la solidité des données ouvertes ? Les Rencontres Décryptageo. [En ligne] 10 Juin 2015. [Citation : 9 Septembre 2017.] <http://lesrencontres.decryptageo.fr/wp-content/uploads/sites/2/2015/04/R2015-reutilisation-Courmont.pdf>.

29. Dane, Carole. L'empowerment, un concept pour la France ? [éd.] ERES. Vie sociale. 20 Septembre 2007, Vol. N° 2, pp. 59-72.

30. Statistiques SDES. [En ligne] [Citation : 26 Septembre 2017.] <https://data.statistiques.developpement-durable.gouv.fr/pages/home2/>.

31. Commissariat Général Au Développement Durable. Chiffres et statistiques - Opinions et pratiques environnementales des Français en 2015. 2016.
32. Site internet de l' Environmental Data & Governance Initiative - EDGI. [En ligne] <https://envirodatagov.org/>.
33. Ministère des Solidarités et de la Santé. Plan Nationale Santé Environnement 3 (2015-2019).
34. Université de Yale. Environmental Performance Index. [En ligne] <http://epi.yale.edu/country-rankings>.
35. Commission européenne. Open Science Monitor. [En ligne] [Citation : 9 Septembre 2017.] <http://ec.europa.eu/research/openscience/index.cfm?pg=home§ion=monitor>.
36. Santé publique France . Impacts sanitaires de la pollution de l'air en France : nouvelles données et perspectives. [En ligne] 21 Juin 2016. [Citation : 09 Septembre 2017.] <http://www.santepubliquefrance.fr/Accueil-Presses/Tous-les-communiqués/Impacts-sanitaires-de-la-pollution-de-l-air-en-France-nouvelles-donnees-et-perspectives>.
37. Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail. Première étude en France sur l'estimation du coût de la pollution de l'air intérieur. [En ligne] <https://www.anses.fr/fr/content/premi%C3%A8re-%C3%A9tude-en-france-sur-l%E2%80%99estimation-du-co%C3%BBt-de-la-pollution-de-l%E2%80%99air-int%C3%A9rieur>.
38. Google. Mapping the invisible: Street View cars add air pollution sensors. Google environnement projects. [En ligne] [Citation : 09 Septembre 2017.] <https://environment.google/projects/airview>.
39. Follett, Ria et Strezov, Vladimir. An Analysis of Citizen Science Based Research: Usage and Publication Patterns. PLoS One. 2013, Vol. 10, 11.
40. Présentation de Transitions². [En ligne] 2017. <http://www.transitions2.net/p/a-propos>.
41. Kaplan, Daniel. Interview de Daniel Kaplan (Co-Fondateur de la FING). [interv.] Rayane Hadji. 28 Juillet 2017.
42. Bonney, Rick, et al. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. 2009.
43. Geospatial Data Based Environment in INSPIRE4Youth. Charvat, Karel, et al. 2017, Procedia Comput. Sci., Vol. 104.
44. Commission Informatique et Libertés - CNIL. Quantified self, m-santé : le corps est-il un nouvel objet connecté ? [En ligne] <https://www.cnil.fr/fr/quantified-self-m-sante-le-corps-est-il-un-nouvel-objet-connecte>.
45. Citoyens capteurs. [En ligne] <http://www.citoyenscapteurs.net/2014/03/31/citizenair-io/>.
46. Observatoire la Qualité de l'Air Intérieur - OQAI. Description du budget espace temps et estimation de l'exposition de la population française dans son logement. 2009.
47. Open Food Facts. [En ligne] <https://fr.openfoodfacts.org>.

48. Open beauty facts. [En ligne] <https://fr.openbeautyfacts.org/>.
49. 10 environmental sensors that go along with you. Treehugger. [Online] 5 Mars 2013. [Cited: 9 Septembre 2017.] <https://www.treehugger.com/clean-technology/environmental-sensors.html>.
50. BVA. Sondage BVA - Les français et leurs préoccupations de la vie quotidienne. 2016.
51. Mills, Mark P. The cloud begins with coal - Big Data, big networks, big infrastructure, and big power : an overview of the electricity use by the global digital ecosystem. 2013.
52. Qarnot. Qarnot - The first computing heater for smart buildings. [En ligne] <https://www.qarnot.com>.
53. Université Jean Moulin Lyon 3 . La chaudière numérique Lyon 3 : les premières photos de l'installation . [En ligne] 2014. <http://www.univ-lyon3.fr/la-chaudiere-numerique-lyon-3-les-premieres-photos-de-l-installation-886637.kjsp?RH=1405931635005>.
54. Le Monde. Des pionniers veulent chauffer les villes avec nos ordinateurs. Journal Le Monde. [En ligne] 7 Janvier 2016. http://www.lemonde.fr/smart-cities/article/2016/01/07/ces-pionniers-qui-veulent-chauffer-les-villes-avec-nos-ordinateurs_4843208_4811534.html.
55. Gouvernement. Pour une République numérique. Site officiel du Gouvernement. [En ligne] <http://www.gouvernement.fr/action/pour-une-republique-numerique>.
56. Commission Européenne. EUR-Lex - I28195 - EN - EUR-Lex. [En ligne] <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=LEGISSUM%3AI28195>.
57. Commission Nationale Informatique et Libertés - CNIL. Les sanctions pénales. Site internet de la CNIL. [En ligne] <https://www.cnil.fr/fr/les-sanctions-penales>.
58. Sénat. La protection des données personnelles dans l'open data : une exigence et une opportunité. [En ligne] 2014. <https://www.senat.fr/rap/r13-469/r13-4694.html#toc68>.
59. Environmental Defense Fund. Mapping air pollution with new mobile sensors. [En ligne] <https://www.edf.org/airqualitymaps>.
60. Goodman, Ben. Health data security. [En ligne] http://www.ehcca.com/presentations/HIPAA24/goodman_2.pdf.
61. NIH Common Fund. Big Data to Knowledge . [En ligne] <https://datascience.nih.gov/bd2k>.
62. Aberkane, Idriss. L'économie de la connaissance : un nouvel eldorado ? [En ligne] 27 Mars 2015. [Citation : 18 Septembre 2017.] <http://le-cera.com/leconomie-de-la-connaissance-un-nouvel-eldorado/>.
63. Gourc, Didier, Vacher, Béatrice et Pingaud, Hervé. Manager les risques en projets : de la prise de conscience à la mise en confiance. Communication et organisation. 2001, Vol. 20.

Table des Matières

REMERCIEMENTS	A
SOMMAIRE	B
GLOSSAIRE.....	C
TABLE DES ILLUSTRATIONS	E
INTRODUCTION:	1
PARTIE 1 : ACCES ET APPROPRIATION DES DONNEES SANTE-ENVIRONNEMENTALES: LE CITOYEN UTILISATEUR/PRODUCTEUR	6
I. POSITIONNEMENT STRATEGIQUE DU CITOYEN DANS LE DOMAINE DES DONNEES.....	6
1. LES ACTEURS AUTOUR DU CITOYEN	6
a. <i>Etat et organismes officiels</i>	6
b. <i>Entreprises</i>	6
c. <i>Associations</i>	7
d. <i>La Recherche</i>	7
2. PLACEMENT DU CITOYEN DANS "L'ECOSYSTEME DATA"	8
II. EN TANT QU'INDIVIDU / CITOYEN, J'AI ACCES A QUOI ?	9
1. QUI PEUT AVOIR ACCES AUX DONNEES ET OU?	9
2. COMMENT SE PROCURER DE NOUVELLES DONNEES?	11
III. CONCRETEMENT, QUELS USAGES PAR LE CITOYEN DES DONNEES SANTE-ENVIRONNEMENTALES?	12
1. CARTOGRAPHIE UN OUTIL POUR UNE GESTION DES RISQUES SANTE-ENVIRONNEMENT? (COMMENT REPRESENTER MES DONNEES)	12
a. <i>Origine du modèle</i>	12
b. <i>De la cartographie papier à la cartographie numérique</i>	13
2. EXEMPLES D'OUTILS DE CARTOGRAPHIE POUR UN USAGE DU CITOYEN.....	15
a. <i>Géoportail - outil de cartographie</i>	15
b. <i>Les Assistants numériques - outils d'alerte</i>	15
c. <i>Des citoyens actifs dans une démarche de production:</i>	16
Par le crowdsourcing:.....	16
Hackathon.....	17
IV. LIMITES DE L'USAGE DES DONNEES SANTE-ENVIRONNEMENTALES PAR LE CITOYEN	19
1. ETAT DES LIEUX DES DONNEES ENVIRONNEMENTALES - ABSENCE DE CERTAINES DONNEES.....	19
2. INCERTITUDES INTRINSEQUES DE LA DONNEE	20
3. ACCESSIBILITE, USAGES ET COMPETENCES	20
CONCLUSION PARTIE 1	22
PARTIE 2: PERSPECTIVES D'AVENIR PAR LE PRISME DU CITOYEN: IMPACTS MULTISECTORIELLES DES DATAS SE	23
I. POLITIQUE.....	23
1. TRANSPARENCE ET EMPOWERMENT: UN MOUVEMENT POLITIQUE EMERGENT.....	23
2. INTERETS POLITIQUES POUR UN DEVELOPPEMENT DE CES TECHNIQUES DE GESTION SANITAIRE ET ENVIRONNEMENTALE.....	25
II. ECONOMIQUE	27
III. SOCIAL.....	29
1. UNE PRISE DE CONSCIENCE DES PROBLEMATIQUES ENVIRONNEMENTALES VIA LES DONNEES PAR LE CITOYEN.....	29

a.	<i>Prise de conscience du potentiel des données: acculturation</i>	29
b.	<i>Vers un développement participatif et des politiques locales</i>	29
	Exemple de processus d'acculturation : Infolab	30
c.	<i>Intégration des méthodes de gestion des données vers un public plus large : en marche vers la production de données?</i>	31
2.	MIGRATION VERS UN SUIVI PERSONNALISE DES INDIVIDUS: LE QUANTIFIED-SELF OU L'INDIVIDU OBJET DATA. ..	32
a.	<i>Concept</i>	32
b.	<i>Impact sur la gestion des risques en santé environnement</i>	32
IV.	TECHNIQUE	34
1.	VERS DE NOUVELLES MANIERES DE PRODUIRE DE LA DONNEE ENVIRONNEMENTALE	34
2.	CITOYEN PRODUCTEUR DE DONNEES NON STRUCTUREES: QUELLES OPPORTUNITES EN SANTE ENVIRONNEMENT?	35
V.	ÉCOLOGIQUE	36
1.	IMPACT SUR LA GESTION DES RISQUES EN SANTE ENVIRONNEMENT	36
2.	IMPACT ENVIRONNEMENTAL DE L'USAGE DES DONNEES.....	36
VI.	LEGISLATIF	38
1.	DONNEES ENVIRONNEMENTALES: ETHIQUE ET RESPONSABILITES POUR LES UTILISATEURS	38
2.	POSSESSION ET PROTECTION DES DONNEES (PERSONNELLES ET NON PERSONNELLES).....	38
	CONCLUSION PARTIE 2	41
	CONCLUSION GENERALE DU MÉMOIRE	42
	BIBLIOGRAPHIE	44
	TABLE DES MATIERES	48
	TABLE DES ANNEXES	I
	RESUME / SUMMARY	XIX

Rayane Hadji
Master 2 – Santé publique
Evaluation et gestion des risques en santé-environnement.
2015/2016

Annexes

Open Data – Big Data : Quel rôle pour le citoyen dans la gestion
sanitaire et environnementale des risques ?



Date de la soutenance : 2016

Directeur de mémoire : Maître de Conférences Associé, Université Lille 2

Membres externes du Jury:

Faculté d'Ingénierie de Management de la Santé - ILIS

Université Lille 2 – Droit et Santé

Adresse :

42, Rue Ambroise Paré

59200 - LOOS

Téléphone :

03.20.62.37.37



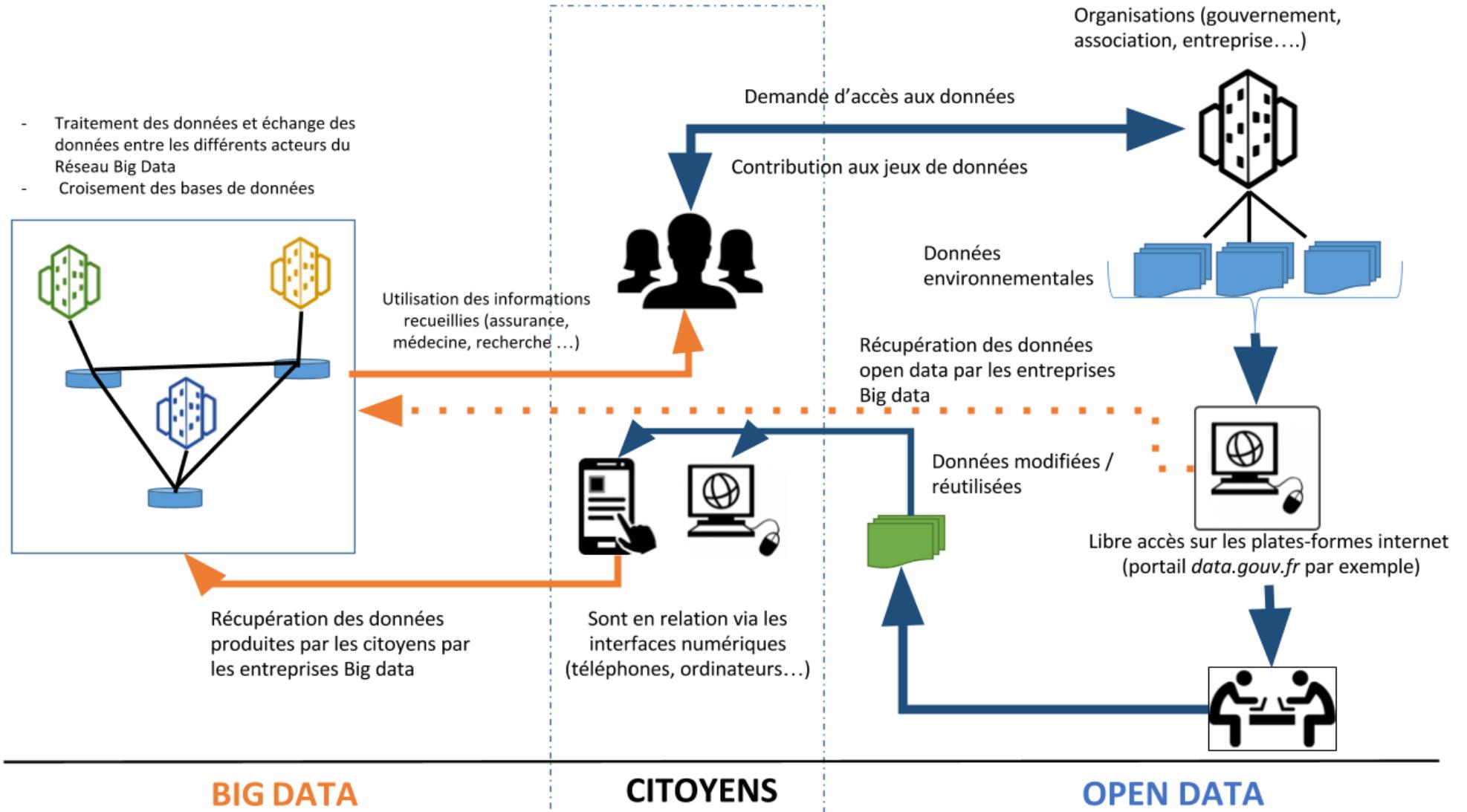
Université Lille 2
Droit et Santé



Table des annexes

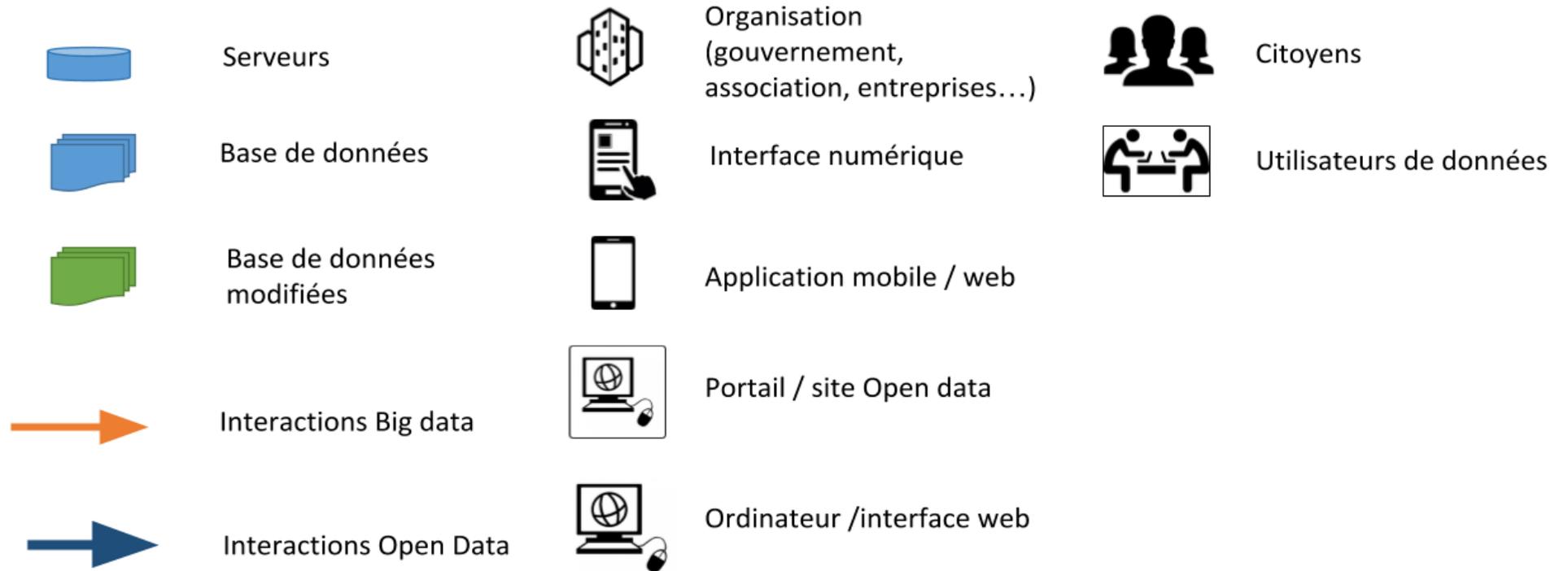
ANNEXE A SCHÉMA CONCEPTUEL OPEN DATA / BIG DATA.....	I
ANNEXE B LE ROLE CENTRAL DU CITOYEN DANS LE DOMAINE DES DATA (LEGENDE)	II
ANNEXE C NIVEAUX D'OUVERTURE DES DONNEES PAR TIM BERNERS-LEE	III
ANNEXE D ARBRES DE DECISION POUR CHOISIR SA PLATEFORME DE DONNEES.....	IV
ANNEXE E TABLEAU DES PLATEFORMES OPEN DATA EN SANTE- ENVIRONNEMENT	XIII
ANNEXE F DATA-VISUALISATION DES EMPREINTES CARBONES MONDIALES – SOURCE PINTEREST	XVIII

Annexe A Schéma conceptuel Open Data / Big Data

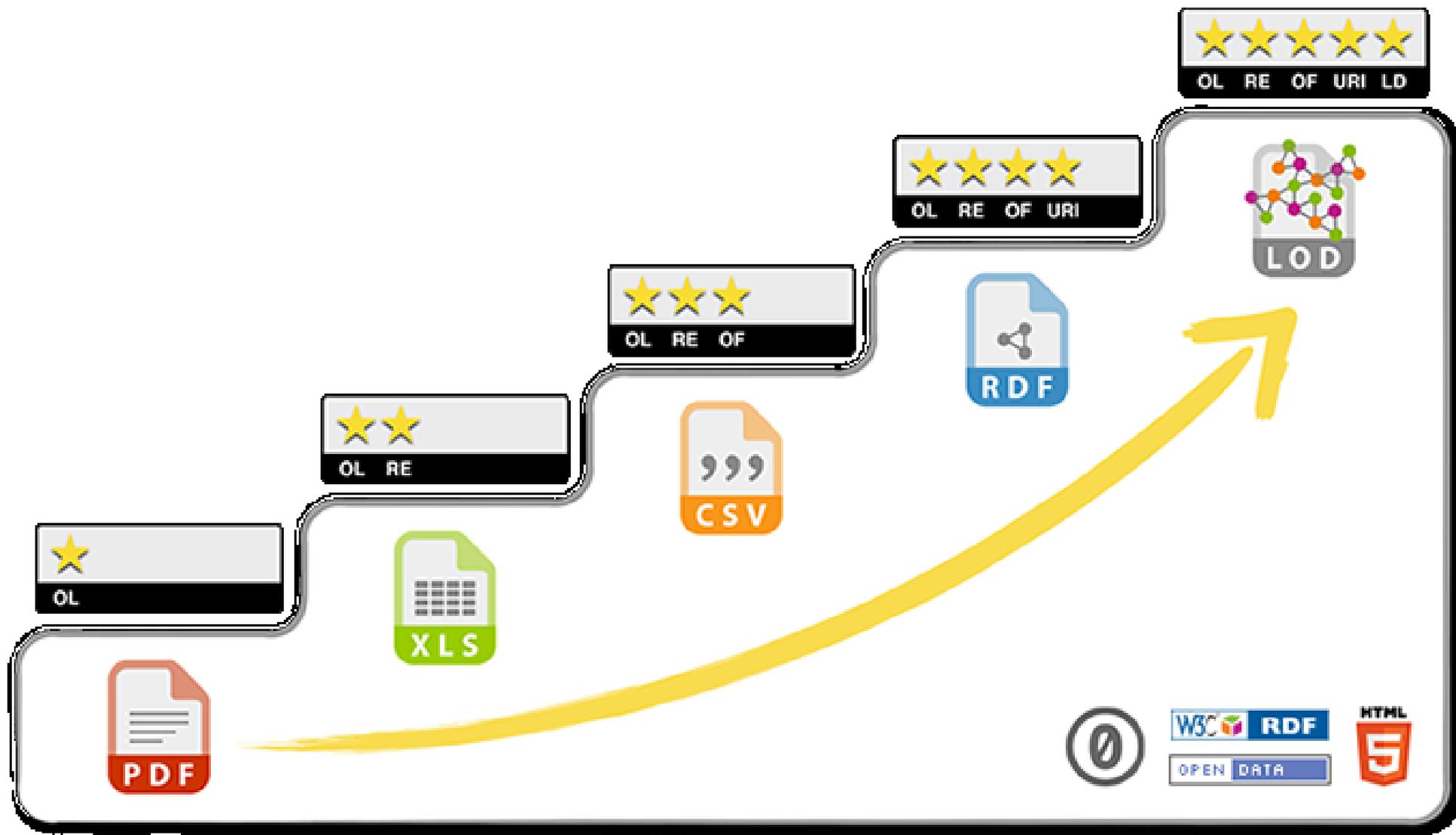


Annexe B Le rôle central du citoyen dans le domaine des data (Légende)

Légende:



Annexe C Niveaux d'ouverture des données par Tim Berners-Lee



Arbres de décisions

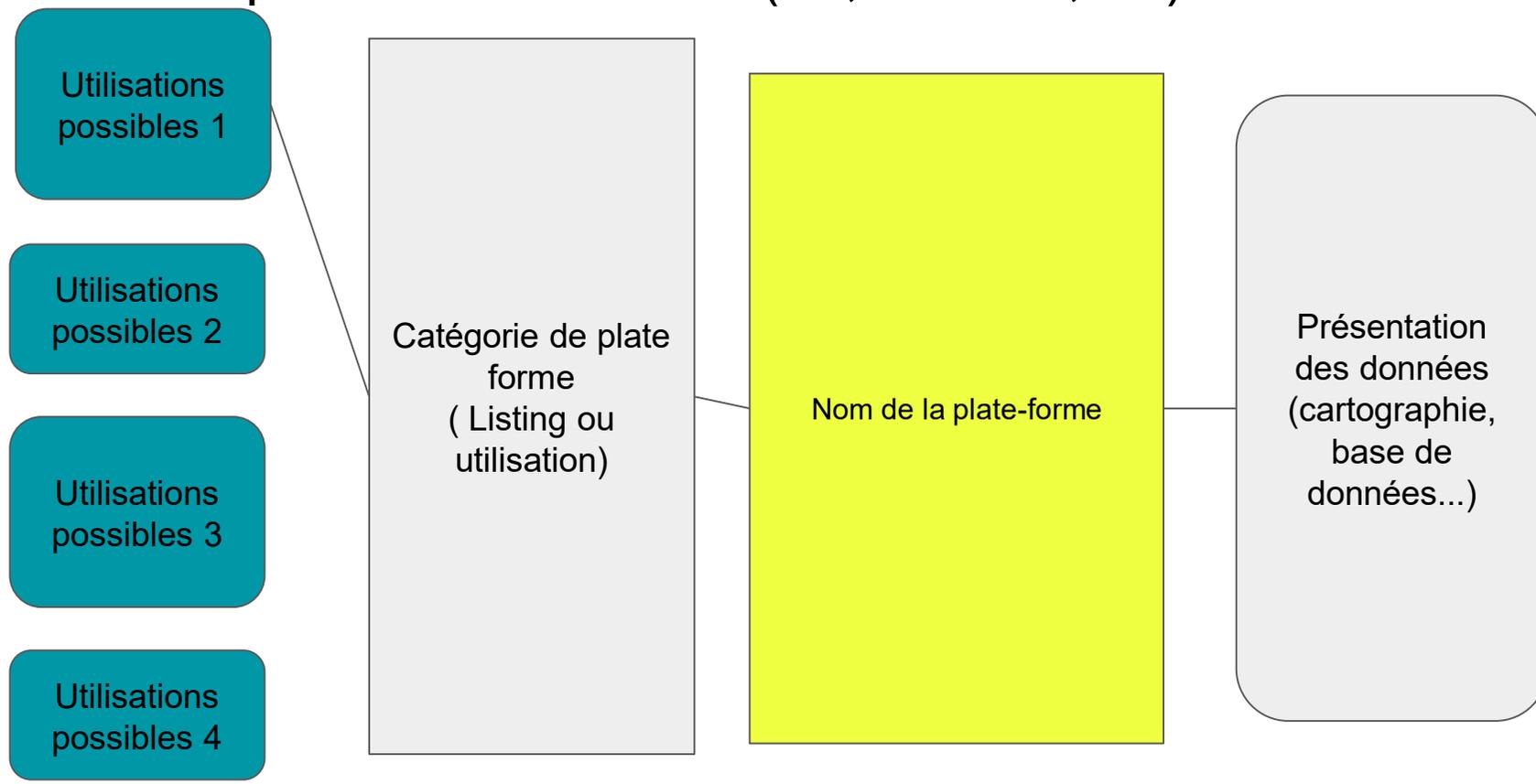
Comment accéder aux données selon mes besoins ?

Sommaire

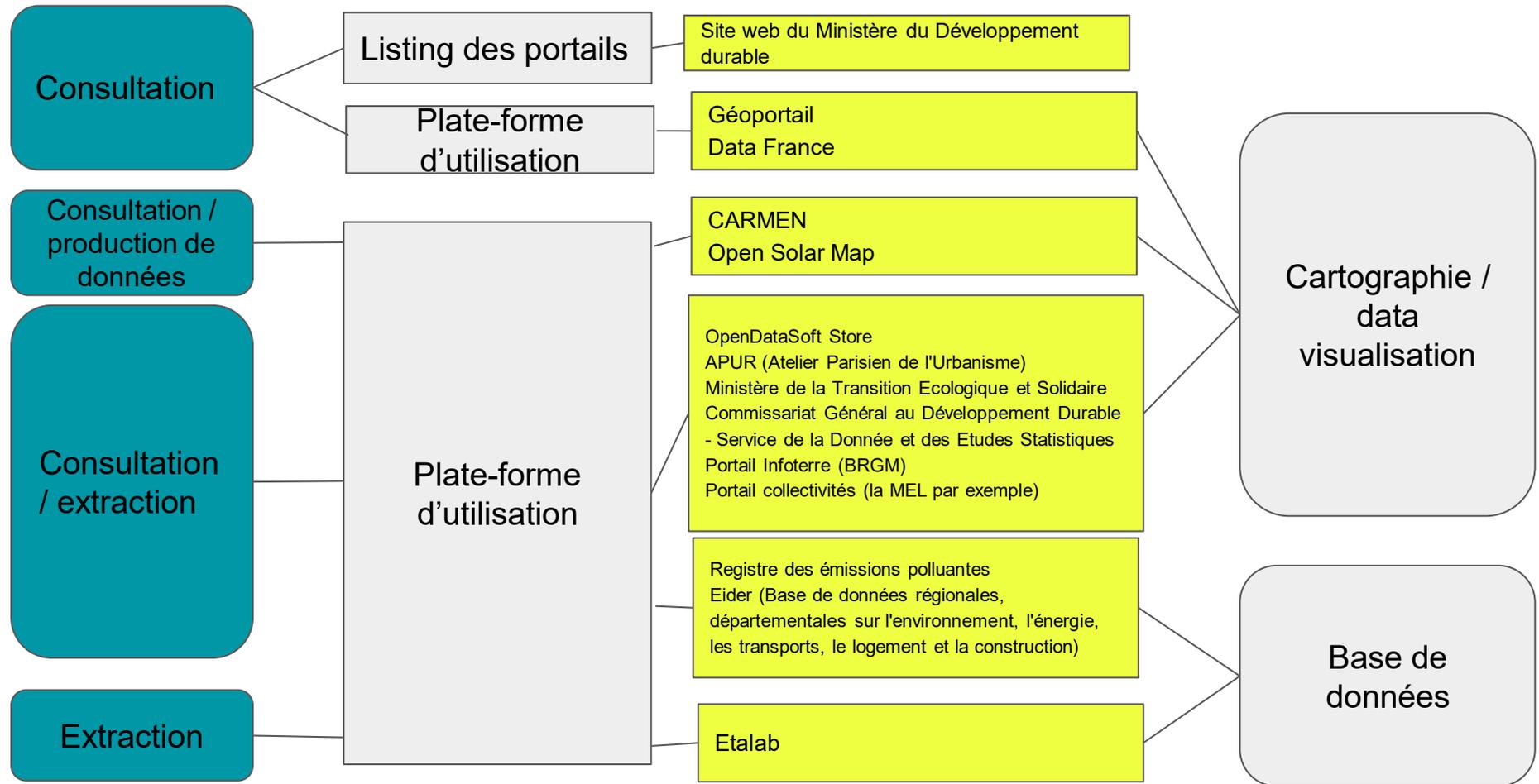
- Comment fonctionne les schémas?
- Schéma de décisions
 - Recherches de divers types de données
 - Recherche de données sur la qualité de l'air
 - Recherche de données sur la qualité des sols
 - Recherche de données sur la qualité des eaux
 - Recherche de données sur les cosmétiques
 - Recherche de données sur les polluants

Comment fonctionne ces arbres?

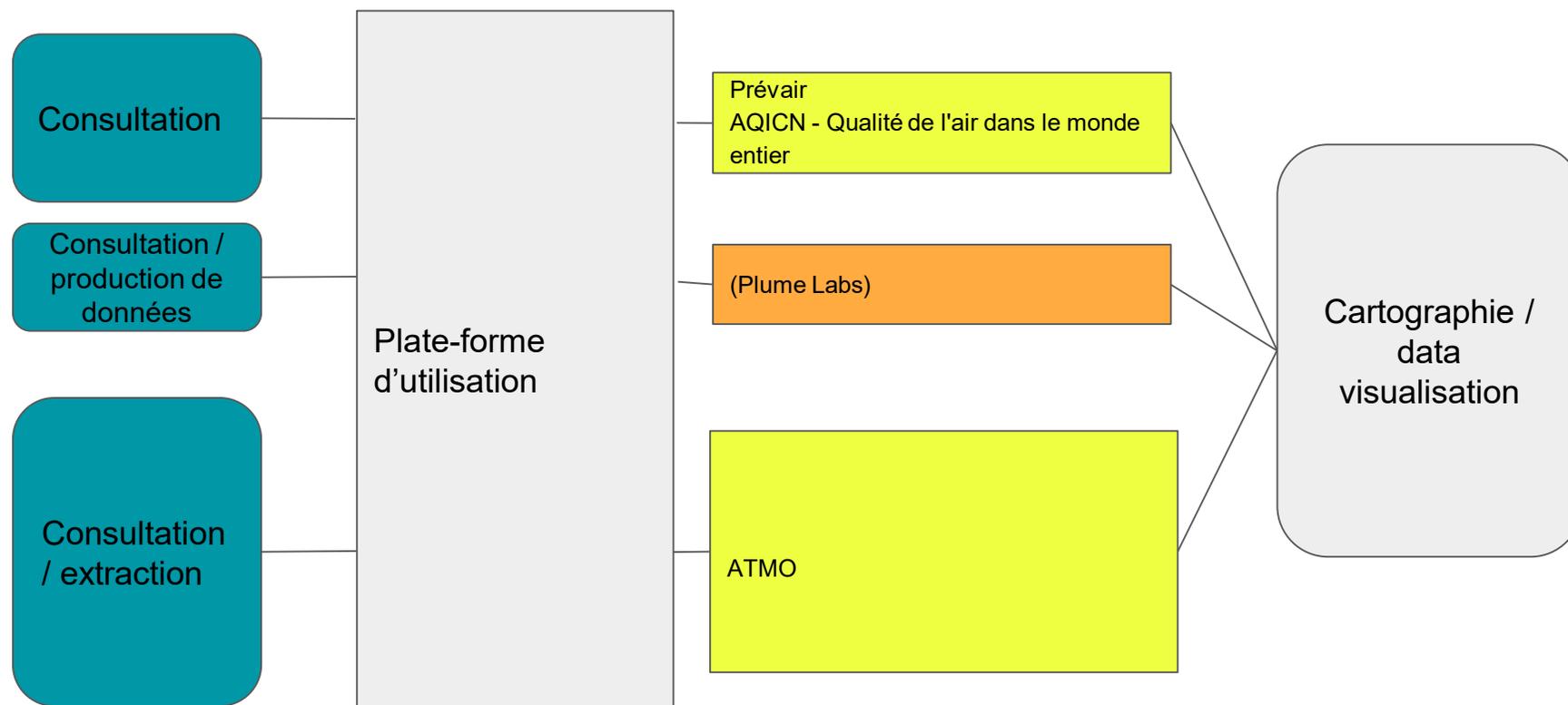
Thématique de la recherche (air, eau sol,)



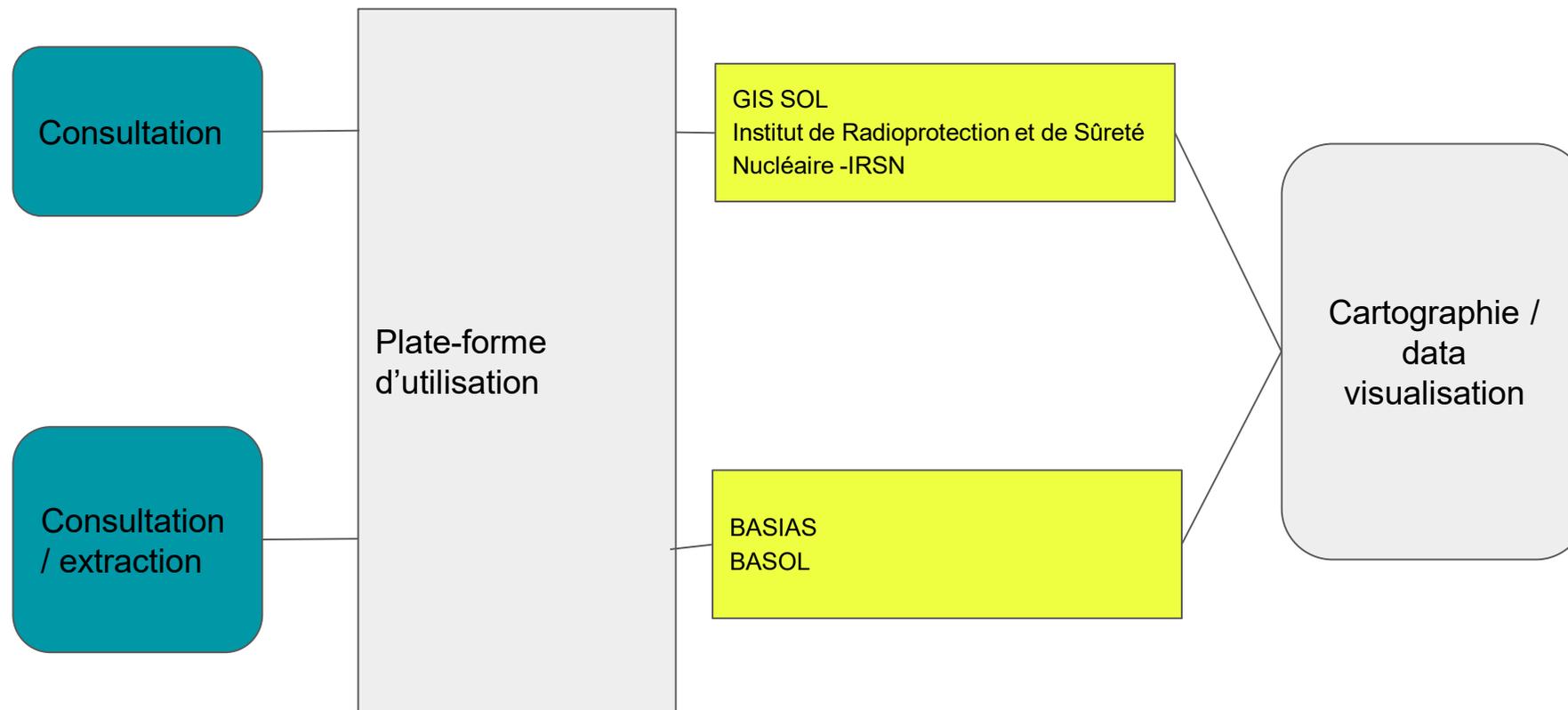
Recherche divers types de données (air, eau, sol, autres...)



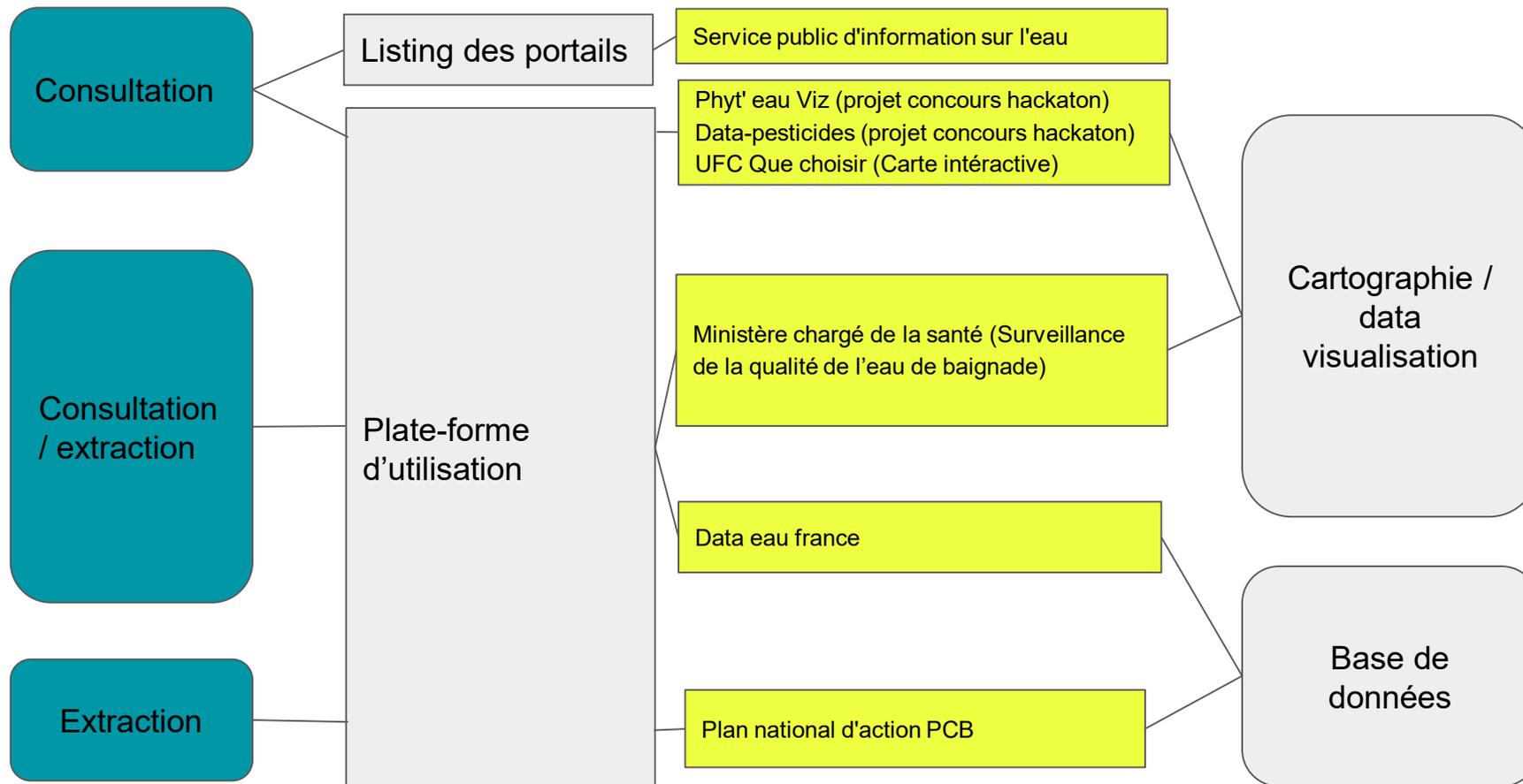
Recherche de données sur la qualité de l'air



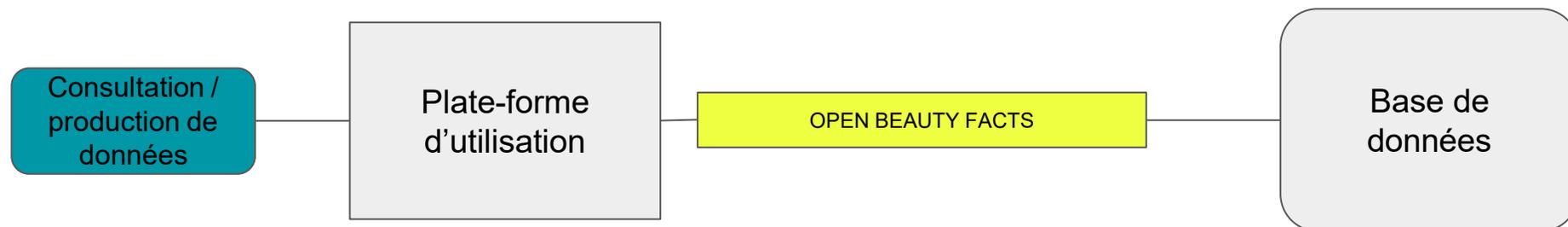
Recherche de données sur la qualité des sols



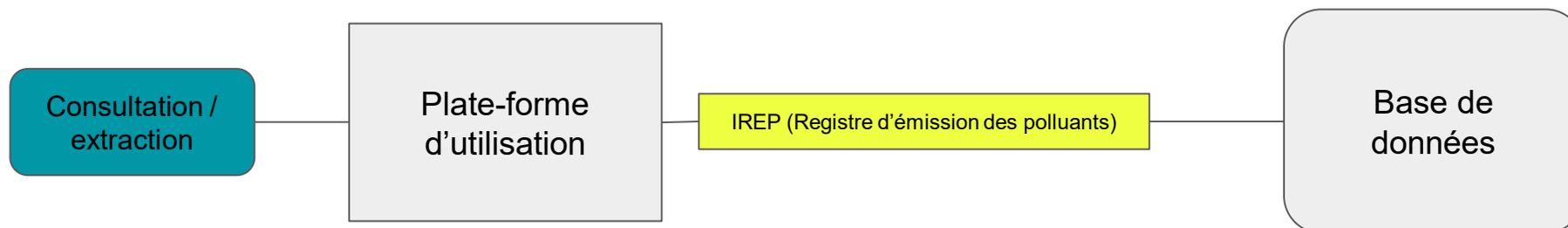
Recherche de données sur la qualité des eaux



Recherche de données sur les cosmétiques



Recherche de données sur les polluants



Annexe E Tableau des plateformes Open Data en santé-environnement

Plateforme	Type d'acteur	Thématique	Catégories	Utilisations	Description	Format	Adresse web
Open beauty facts	Association	Cosmétiques	Plate-forme d'utilisation	Production / consultation	Caractéristiques du produit Dénomination générique Quantité Conditionnement Pays de vente Liste des Ingrédients Analyse des familles d'ingrédients (avec / sans produits...) Additifs Recherche par polluants	Site internet CSV	https://fr.openbeautyfacts.org
Etalab	Gouvernement	Divers types de données	Plate-forme d'utilisation	Extraction	Plusieurs jeux de données disponibles	CSV	https://www.data.gouv.fr
OpenDataSoft Store	Entreprise	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Plusieurs bases de données disponibles sur des thématiques variées: Qualité des milieux (air, eau), Quantités de déchets produits, espaces verts.... Données consultable sur le site (tableaux, carte, analyse)	CSV, XLS	https://data.opendatasoft.com
Géoportail	Gouvernement	Divers types de données	Plate-forme d'utilisation	Consultation	Cartes interactives	Carte interactive	https://www.geoportail.gouv.fr/
Data France	Association	Divers types de données	Plate-forme d'utilisation	Consultation	Visualisation de données ouvertes Indicateurs qualité sur : éducation, santé, transports,	Carte interactive	https://datafrance.info/

Portail Infoterre (BRGM)	Gouvernement	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Carte interactive Accès aux couches de la Banque du Sous-Sol (BSS), aux cartes géologiques, ainsi qu'aux autres données relatives aux thématiques du BRGM (dont rapports)	Carte + CSV + PDF	http://infoterre.brgm.fr/viewer/MainTileForward.do
Portail collectivités (la MEL par exemple)	Gouvernement	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Données locales sous forme de carte et base de données	Carte + CSV	https://opendata.lillemetropole.fr/page/home/
CARMEN	Gouvernement	Divers types de données	Plate-forme d'utilisation	Production	Editeur de cartes	Carte interactive	http://carmen.naturefrance.fr/
Open Solar Map	Organisation	Divers types de données	Plate-forme d'utilisation	Production	Editeur de cartes	Carte interactive	http://openstreetmap.org
Site web du Ministère du Développement durable	Gouvernement	Divers types de données	Portail - listing	Consultation	Listing des bases de données	Liens HTML	http://www.centre.developpement-durable.gouv.fr/donnees-bases-de-donnees-et-cartes-a1009.html#sommaire_10
APUR (Atelier Parisien de l'Urbanisme)	Association	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Carte interactive et bases de donnée sur l' Ile de France de l'Atelier Parisien d'Urbanisme	shapefile	http://cassini.apur.opendata.arcgis.com/
Ministère de la Transition Ecologique et Solidaire Commissariat Général au Développement Durable - Service de la Donnée et des Etudes Statistiques	Gouvernement	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Plateforme de data-visualisation	site internet	https://data.statistiques.developpement-durable.gouv.fr/pages/home2/

Eider (Base de données régionales, départementales sur l'environnement, l'énergie, les transports, le logement et la construction)	Gouvernement	Divers types de données	Plate-forme d'utilisation	Consultation / extraction	Tableaux, série longues et rapports personnalisés	CSV, XLS	http://www.stats.environnement.developpement-durable.gouv.fr/Eider/tablees.do
Registre des émissions polluantes	Gouvernement	Divers types de données Polluants air, eau, sol	Plate-forme d'utilisation	Consultation / extraction	Registre des polluants émis sur le territoire Registre des sites de Production et traitement des substances dangereuses	CSV	http://www.georisques.gouv.fr/dossiers/irep-registre-des-emissions-polluantes
Data eau france	Gouvernement	Eau	Plate-forme d'utilisation	Consultation / extraction	Base de données		http://www.data.eaufrance.fr
Plan national d'action PCB	Gouvernement	Eau	Plate-forme d'utilisation	Extraction	Données pollution des eaux aux PCB, dioxines, mercure par bassin	XLS	http://www.pollutions.eaufrance.fr/PCB/donnees.html

Data-pesticides (projet concours)	Gouvernement	Eau	Plate-forme d'utilisation	Consultation	Carte interactive de la répartition géographique des pesticides dans les eaux souterraines et l'évolution de leur concentration dans le temps (approche France entière ou départementale). Mise à disposition d'informations sur les caractéristiques des familles de pesticides détectés.	Carte interactive	www.data-pesticides.fr
Phyt' eau Viz (projet concours)	Gouvernement	Eau	Plate-forme d'utilisation	Consultation	Visualisation des mesures de concentration des pesticides dans les multiples masses d'eau souterraine de France métropolitaine.	Carte interactive	https://pedagueau.shinya.pps.io/phyteauviz/
UFC Que choisir	Entreprise	Eau	Plate-forme d'utilisation	Consultation	Carte interactive de la qualité de l'eau en France	Carte interactive	https://www.quechoisir.org/carte-interactive-qualite-eau-n21241/
Ministère chargé de la santé	Gouvernement	Eau	Plate-forme d'utilisation	Consultation / extraction	Indicateurs qualité des eaux de baignade	Site internet	http://baignades.sante.gouv.fr/baignades/navigMap.do
Service public d'information sur l'eau	Gouvernement	Eau	Portail - listing	Consultation	Listing des bases de données sur l'eau (qualité eaux de surfaces, eaux souterraines...)	Site internet	http://www.eaufrance.fr/donnees/accéder-aux-donnees
IREP (Registres des émissions de polluants)	Gouvernement	Polluants	Plate-forme d'utilisation	Consultation / extraction	Principaux rejets et transferts de polluants dans l'eau, l'air, les déchets déclarés par certains établissements; à savoir: Industriels, STEP de plus de 100 00 eq hab, certains élevages Données quantitatives, annuelles, cumulatives et comparatives en fonction du secteur d'activité et du polluant	CSV	http://www.georisques.gouv.fr/registre-des-emissions-polluantes-irep/presentation-0

Prévoir	Organisation	Qualité de l'air	Plate-forme d'utilisation	Consultation	Carte interactive pollution air (PM 10 et 2,5 + NO2 + Ozone + (CO2 et SO2 - mesures des stations ATMO)	Carte	http://www2.prevoir.org/
Qualité de l'air dans le monde entier	Entreprise	Qualité de l'air	Plate-forme d'utilisation	Consultation	Indice de qualité de l'Air dans le monde entier en temps réel Basé sur les mesure PM10 2.5 , ozone, NO2, SO2	Carte interactive	http://aqicn.org/here/fr/
ATMO	Organisation	Qualité de l'air	Plate-forme d'utilisation	Consultation / extraction	Carte interactive pollution air (PM 10 et 2,5 + NO2 + Ozone + (CO2 et SO2 - mesures des stations ATMO)	Carte + CSV	http://www.atmo-france.org/fr/index.php?/200804119/carte-des-aasqa/id-menu-222.html
Systèmes d'information géographique et liens environnement santé - SIGLES	Recherche	Santé Environnement Sociodémographie Inégalités territoriales	Plate-forme d'utilisation	Consultation	Atlas de cartes Rapports / publications	Carte	http://www.sigles-sante-environnement.fr/
Basias	Gouvernement	Sol	Plate-forme d'utilisation	Consultation / extraction	BASIAS Inventaire historique des sites industriels et activités de service	CSV	http://basias.brgm.fr/
BASOL	Gouvernement	Sol	Plate-forme d'utilisation	Consultation / extraction	Base de données Basol sur les sites et sols pollués ou potentiellement pollués appelant une action des pouvoirs publics à titre préventif ou curatif.	Site internet	http://basol.developpement-durable.gouv.fr/
GIS SOL	Organisation	Sol	Plate-forme d'utilisation	Consultation	Carte interactive Base de données	Carte interactive	http://www.gjssol.fr/donnees/webservices
Institut de Radioprotection et de Sécurité Nucléaire -IRSN	Organisation	Sol	Plate-forme d'utilisation	Consultation	Cartographie des sols contenant du radon	Carte interactive	http://www.irsn.fr/FR/connaisances/Environnement/expertises-radioactivite-naturelle/radon/Pages/4-cartographie-potentiel-radon-commune.aspx#.WcPveuL-TCs

Résumé / Summary

Rayane HADJI Master 2 Santé Publique - Evaluation et gestion des risques en santé-environnement.

L'ouverture des données publiques (Open Data) et le développement du Big Data ont amorcé un mouvement en faveur de l'utilisation des données environnementales. Dans cet écosystème numérique, le citoyen s'y trouve au centre et tient un rôle dans l'amélioration de la gestion des risques sanitaires et environnementaux. L'objectif de cette étude est de déterminer le rôle du citoyen dans l'univers Open et Big Data. Autour de lui, l'Etat et ses organismes, les entreprises, les associations et la Recherche participent à la production et la diffusion de données environnementales sur la pollution atmosphérique par exemple. Ces données sont accessibles à travers plus de 30 plateformes en ligne, et se caractérisent par des bases de données, des cartes interactives ou encore des systèmes d'alertes. La contribution citoyenne existe aussi par des initiatives collectives telles le crowdsourcing et le hackathon. Malgré les incertitudes relatives aux données (qualité, accessibilité et usages), l'utilisation des données environnementales par et pour le citoyen présente un potentiel à divers niveaux. Les perspectives sont multiples. D'un point de vue politique, un mouvement d'empowerment se développe pour obtenir plus de données. Ceci permettra aux citoyens un meilleur suivi des évolutions de la qualité de l'environnement. L'innovation dans les méthodes de gestion permettra un gain économie. Sur le plan social, l'appropriation de la culture de la donnée est une porte d'entrée vers une prise de conscience collective des problématiques environnementales. Enfin les progrès de miniaturisation développent la portabilité des outils de quantified-self. Ces derniers chercheront à déterminer les expositions individuelles à grande échelle (nombres et lieux). C'est pourquoi, la frontière entre données environnementales et sanitaires et données privées est de plus en plus ténue.

Mots-clés : open big data citoyen santé environnement

Rayane HADJI Master 2 Public Health – Health and environmental risks assement and management.

The widening of public data (Open Data) and the development of « Big Data » have started a motion towards the use of environmental data. The citizens are at the center of this digital ecosystem and play a role regarding the improvement of environmental and public health risk management. This study's purpose is to determine the citizen's role in those Open & Big Data universes. Alongside him, the State's institutions, the businesses, the charities, and researchers participate to the production and spread of environmental data, in the air pollution domain for example. This data is available through over 30 online platforms characterized as databases, interactive maps or alert systems. The civic contribution also exist through collective initiatives such as the crowdsourcing or the hacktathon. Despite relative doubt regarding the data (quality, access and utilization), the use of environmental data by and for the citizens shows potential at different levels. Perspectives are numerous. From a political standpoint, an empowering movement develops itself to obtain more data. This will allow better monitoring of air pollution changes for citizens. Innovation in management methods will lead to an economic benefit. On a social ground, data's cultural appropriation is a gateway towards a global rise of awareness of environmental issues. At last, miniaturizing breakthroughs improve quantified-self tools' portability. Those aim at defining individual exposure on a larger scale (number and sites). This is why the border between environmental and public health data and private data becomes even thinner.

Key-words: open big data citizen health environment