



Interopérabilité des données environnementales

Présentation



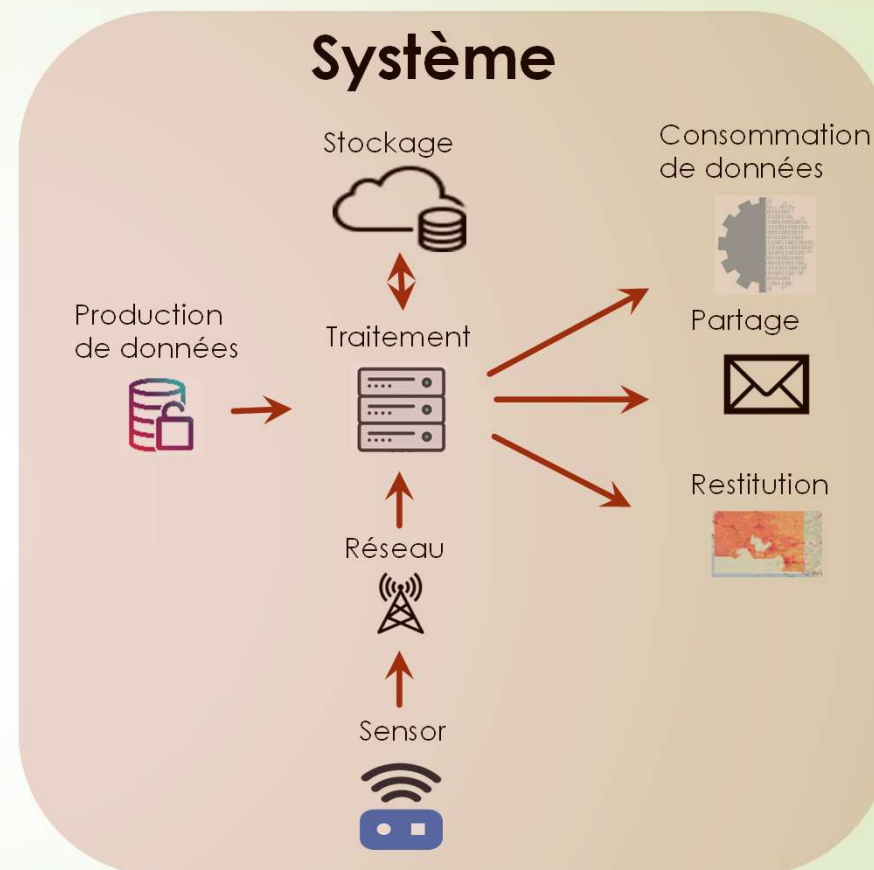
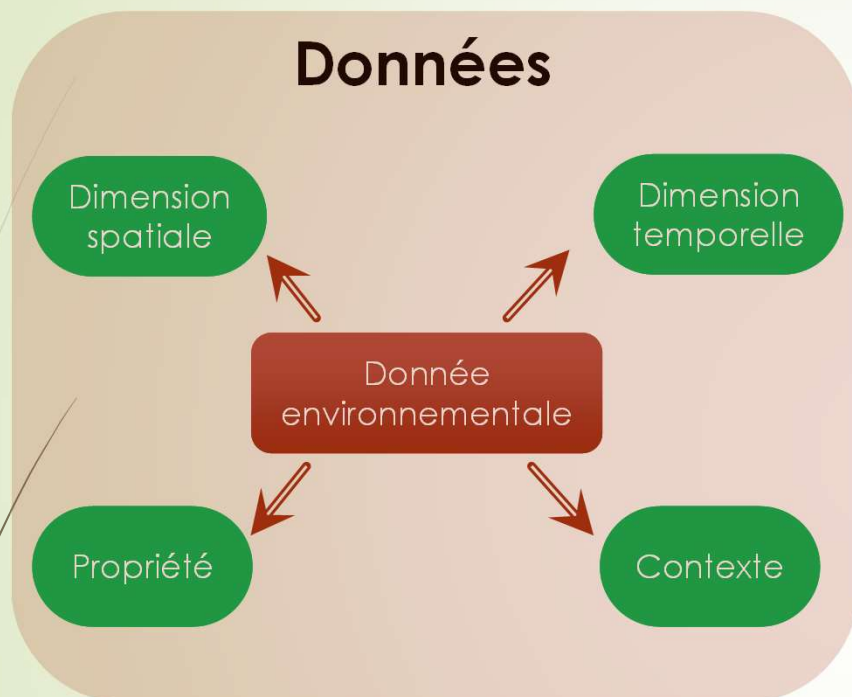
Sommaire

- Introduction
- Critère 1 : La finalité d'intérêt général
- Critère 2 : L'impact
- Critère 3 : La maturité et la structuration du projet
- Critère 4 : L'ouverture, la transparence et l'intégrité
- Critère 5 : Potentiel d'effet de levier de l'accompagnement par l'administration
- Synthèse
- *Annexe : Description du projet*
- *Annexe : Données environnementales*

Introduction

09/01/2022

Interopérabilité des données environnementales



Trois niveaux

- Technique « pouvoir communiquer » : format des données, protocoles (ex. Json, LoRa, Xarray, Tiff)
- Syntaxique « savoir communiquer » : mode de formatage et de codage des données (ex. Fichier, API, Payload)
- Sémantique « savoir se comprendre » : signification des données échangées (ex. modèle de données)

Données environnementales

Avis du conseil
national du numérique
– juillet 2020

« Faire des données
environnementales des
données d'intérêt général »

En deuxième lieu, l'interopérabilité et la qualité des données, ainsi que leurs structure, authenticité et intégrité, sont essentielles pour le partage de données d'intérêt général. Dès lors, le régime d'ouverture des données d'intérêt général implique la prise en considération de la lisibilité et l'interopérabilité des formats et des données. Celles-ci peuvent effectivement être disponibles dans un format uniforme et directement lisible et interopérable ou présentées dans des formats hétérogènes peu lisibles et donc difficilement interopérables. Les auditions du Conseil national du numérique ont confirmé la nécessité d'élaborer des standards de données afin que chaque producteur de données puisse émettre une donnée (ré)utilisable facilement. En effet, l'absence de standards rend souvent les données inexploitable ou nécessite un traitement supplémentaire ralentissant leur usage. Néanmoins, l'interopérabilité représente un coût, ce qui implique l'allocation de moyens financiers substantiels.

Ainsi, les membres du Conseil estiment que les données environnementales d'intérêt général devraient être mises à disposition « *dans des formats qui sont ouverts, lisibles par machine, accessibles, traçables et réutilisables, en les accompagnant de leurs métadonnées* » en recourant à des interfaces de programmation (API) appropriées et, le cas échéant, sous la forme d'un téléchargement de masse. Cela permettrait de définir clairement le périmètre des données accessibles, les méthodes d'interrogation de la base de données, etc¹⁴⁶.

MAPPING DATA PORTABILITY INITIATIVES, OPPORTUNITIES AND CHALLENGES

OECD DIGITAL
ECONOMY
PAPERS

December 2021

Interoperable specifications including standards and APIs

Lack of common standards and interoperability is one of the most frequently cited barriers to the implementation of data portability and for effective re-use of the data. Users, in practice, may face difficulties porting their (personal) data because most information systems are not interoperable and standards are a condition for interoperability. These standards can exist at various levels corresponding to the aforementioned interoperability requirements for data, protocol, information systems and digital services. Even use of commonly used machine-readable formats may not guarantee data interoperability. As highlighted in OECD (2019^[10]), and noted in previous sections, common formats may enable *syntactic* interoperability, i.e. the transfer of “data from a source system to a target system using data formats that can be decoded on the target system” (and thus accessibility). However, common formats do not guarantee *semantic* interoperability, defined as “transferring data to a target such that the meaning of the data model is understood”. Both syntactic and semantic interoperability are needed for the re-use of data. Regarding this potential gap, the (former) Article 29 Data Protection Working Party (2017^[87]) guidelines on data portability complement the requirement of machine-readable format with the intent to achieve interoperability. They explain that “[t]he most appropriate format will differ across sectors and adequate formats may already exist, but should always be chosen to achieve the purpose of being interpretable.”³⁵

Responses³⁶ from the Online Platform Survey underline the lack of standards and technical inconsistencies as major challenges in importing data from other online platforms. In particular, a major challenge was to ensure compatibility between data models. In this context, respondents highlighted the challenge of identifying common standards for data while balancing the need for new and innovative (as well as an increasing variety of) data formats with data interoperable formats and standards. Another related challenge was to provide data that are both machine-readable and intelligible to end-users, while managing the growing variety of technological, business and legal requirements.

Respondents noted that data formats were often specific to use cases even within a single category of data. In this regard, building common, open-source data models and data format standards was highlighted as a possible solution. These common standards should also guarantee privacy protection in data

Liens

- Faire des données environnementales des données d'intérêt général (rapport CNNum juillet 2020)
- Pour une politique publique de la donnée (rapport Bothorel décembre 2020)
- Proposal for a European Interoperability Framework for Smart Cities and Communities (may 2021)
- OECD Mapping data portability initiatives, opportunities and challenges (December 2021)

Critère 1 : La finalité d'intérêt général

1 - Finalité d'intérêt général – problème identifié

L'interopérabilité des données environnementales est une préoccupation d'intérêt général et s'inscrit dans une politique publique prioritaire et stratégique.

L'absence d'interopérabilité est constatée à plusieurs niveaux :

- *Les capteurs et systèmes d'acquisition de données privés et publics utilisent des formats de données propriétaires non standard (nécessite l'usage d'un « décodeur »).*
- *Les formats d'échange et de stockage sont standardisés au niveau technique (ex. format JSON) mais pas au niveau sémantique (nécessite également l'usage d'un décodeur pour chaque format).*
- *Il n'existe pas de format d'échange défini pour partager des données à la fois spatiales, temporelles et multi-propriétés*
 - *Ex. le format Geolson ne sait pas traiter les données temporelles,*
 - *Ex. les fichiers CSV sont adaptés pour la dimension 2 (un tableau) mais pas pour la dimension 3*

Exemples de conséquences :

- **On ne sait pas mettre en place un système de « dosimétrie » qui dialogue avec des capteurs présents à la fois à la maison, sur un lieu de travail, dans les transports, dans les lieux publics**
- **On ne sait pas mettre à disposition de capteurs « plug and play »**
- **On ne sait pas fournir en open-data des données complètes (ex. un niveau de pollution défini par un ensemble de concentrations de polluants mesurées sur une échelle de temps et sur une répartition spatiale,)**

1- Finalité d'intérêt général – solution proposée

La solution proposée est de :

- **construire un standard opérationnel et applicable pour les données environnementales (à partir des standards existants)**
 - Définies par trois composantes : physique, spatiale, temporelle
- **Fournir des bibliothèques de connecteurs open-source**
- **Fournir des « guidelines » de déploiement et de prescription**

Exemples d'usages possibles :

- **Mettre en place dans les lieux publics des capteurs accessibles à tous depuis un mobile**
- **Permettre à chacun, sans développement et en open-source, de connecter un capteur, de stocker les données et de les partager**
- **Fournir une open-information compréhensible (open-data de plus haut niveau qu'aujourd'hui)**

Critère 2 : L'impact

08/01/2022

11

Confidential C

2 – L'impact

Nombre de personnes touchées :

L'impact est global et concerne toutes les personnes qui produisent, échangent et consultent des données environnementales

Capacité à traiter une cause racine :

L'absence de standard est la principale cause racine (mais pas la seule) à l'absence d'interopérabilité

Diffusable à large échelle

Par définition, un standard est diffusable à large échelle

Réplicable

L'approche multi-dimensionnelle retenue (cf principes définis dans le standard ISO telle qu'elle est définie par le standard ISO-19156 « observation and measurement ») est applicable à d'autres catégories de données

Critère 3 : La maturité et la structuration du projet

3 – La maturité

Projet ayant fait ses preuves :

Plusieurs expérimentations aussi bien sur des capteurs que sur des échanges de données ont été effectués (cf travaux réalisés par « A lab in the air »)

Projet ayant déjà commencé à travailler :

Le coeur du standard est déjà disponible et partagé (présentation du projet, spécification du format textuel, spécification du format binaire).

Projet ayant déjà mis un premier produit en production:

Un premier niveau de convergence de standard a été déployé en septembre 2021 : standard Bluetooth : extension de l'ESS (Environmental Sensing Service)

Les outils associés (connecteurs) sont déjà disponibles en open-source dans deux langages (python, C++).

Projet porté par une équipe solide :

Le projet est porté par 3 personnes :

- Un expert technique ayant une forte compétence en échange de données et pilotage de projet
- Un expert innovation ayant une forte compétence en lancement de projets et structures d'innovation
- Un expert informatique ayant une forte compétence en architecture logiciell

Critère 4 : L'ouverture, la transparence et l'intégrité

4 – Ouverture, transparence, intégrité

Initiative ouverte et transparente:

Un standard ne peut être défini, construit, mis en œuvre et déployé sans approche ouverte et transparente.

Les premiers travaux réalisés sont totalement ouverts (partage sur la plateforme Github) et transparents (les principes sous-jacents sont expliqués et partagés).

Respect du cadre légal en vigueur

La mise en œuvre du projet s'appuiera sur une structure adaptée. Plusieurs solutions sont possibles :

- Structure dédiée de type SCIC
- Intégration à des structures d'innovation ou de projets déjà existantes (France / Europe)
- Adossé à des structures spécialisées (ex. NumFOCUS)

Critère 5 : Potentiel d'effet de levier de l'accompagnement par l'administration

5 – Effet de levier

Aide demandée à l'administration:

Le soutien des services publics est indispensable pour la réussite de ce type de projet qui rentre difficilement dans une logique de start'up ou de structure privée.

Leviers identifiés

Mise en réseau avec les lieux et communautés d'innovation

Ce levier est important pour permettre de valider les concepts et choix effectués, finaliser la définition de la solution et construire un premier cercle de soutien et de support.

Promotion et valorisation du projet

L'état est à la fois le principal prescripteur de normes et standards et son principal bénéficiaire. La promotion et la valorisation du projet par les services publics est un gage de réussite du projet.

Solutions de financement et aide au montage juridique

Les dispositifs existants ne sont pas explicitement adaptés pour ce projet. Une aide de l'administration pour structurer ce projet aussi bien dans sa phase de conception que dans sa phase de déploiement est importante.

Distribution de la solution et terrains d'expérimentation

L'expérimentation est nécessaire pour valider la solution et ses apports. L'état est à même de fournir des cadres d'expérimentation pertinents pour ce type de projet.

Les services publics ont également un rôle de prescripteur qui facilite dans un second temps la distribution de la solution.

Synthèse

09/01/2022

19

Confidential C

Synthèse

Le sujet abordé est une préoccupation d'intérêt général et s'inscrit dans une politique publique prioritaire.

Le projet répond à des besoins clairement identifiés.

Il « coche » l'ensemble des 5 critères de sélection définis

Liens proposés :

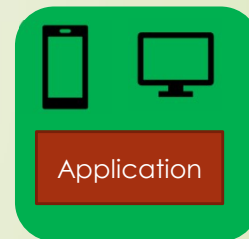
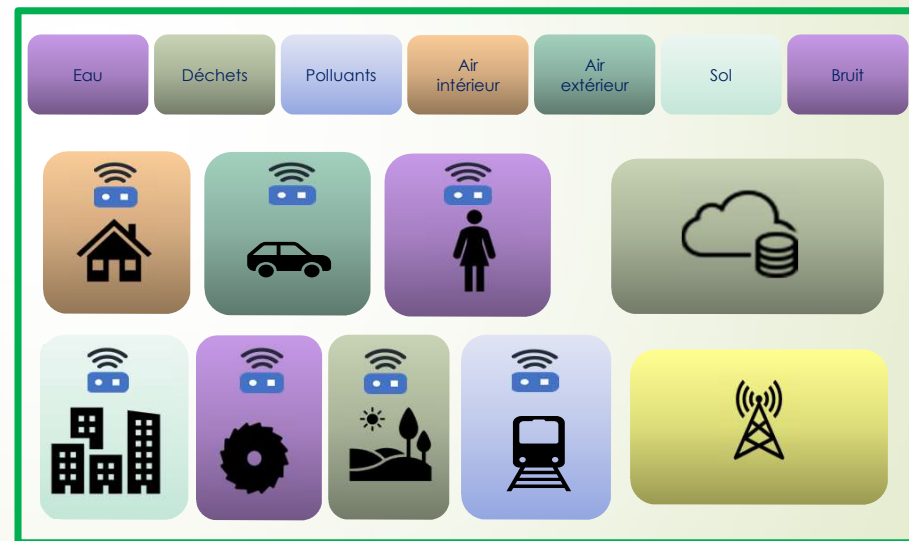
- [Présentation du projet ES \(Environmental Sensing\)](#)
- [Connecteurs ES open-source \(version python\)](#)
- [Intégration des données air dans le profil Bluetooth « environmental sensing »](#)

Annexe : description du Projet

Objectifs

Rendre les données environnementales interoperables

- Fournir des services et outils pour faciliter l'acquisition, le partage, le traitement et la mise à disposition des données et informations environnementales ...



Solutions segmentées

- Propriétaire
- Interface
- Verticale

Informations

- Partielle
- Peu accessible
- Non adaptée

Outils disparates

- Sensor
- Réseaux
- plateforme IoT
- BDD, SIG

Capteur banalisé

- Accès public
- Accès restreint
- Lieux publics / privé

Multi-réseaux

- Etendu - LPWAN
- Local - LAN
- Personnel - PAN

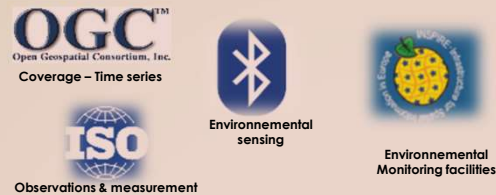
Partage-Accès standard

- Multi-variables
- Structure BDD
- API

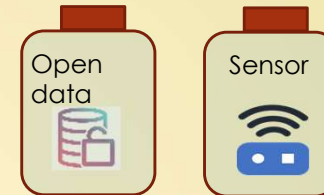
Connecteurs et Standards

...en s'appuyant sur des standards reconnus et convergés qui banalisent ces services et outils (ex. casques audio Bluetooth)

Multi-standards



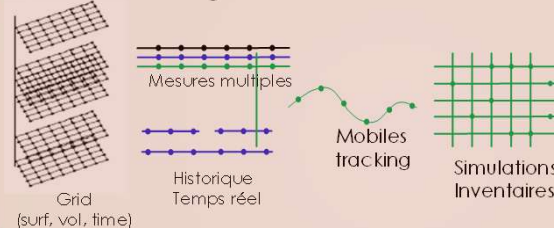
Acquisition



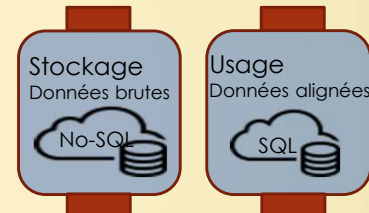
Réseau



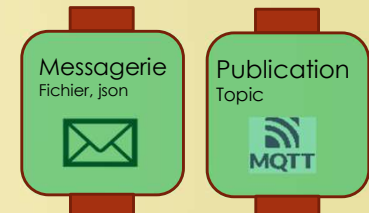
Multi cas d'usage



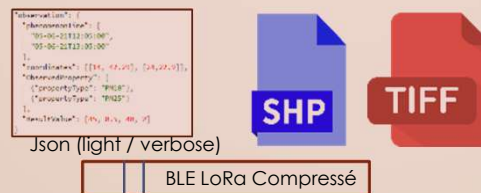
Stockage



Partage



Multi-formats



Texte, binaire, fichier

Traitement



Ouvert

Extensible, paramétrable, ajustable

Clients

Fournisseurs de plateformes et de services
Fournisseurs de logiciel de conception, d'étude et d'exploitation
Gestionnaires de flottes de capteurs
Intégrateurs de systèmes de mesure

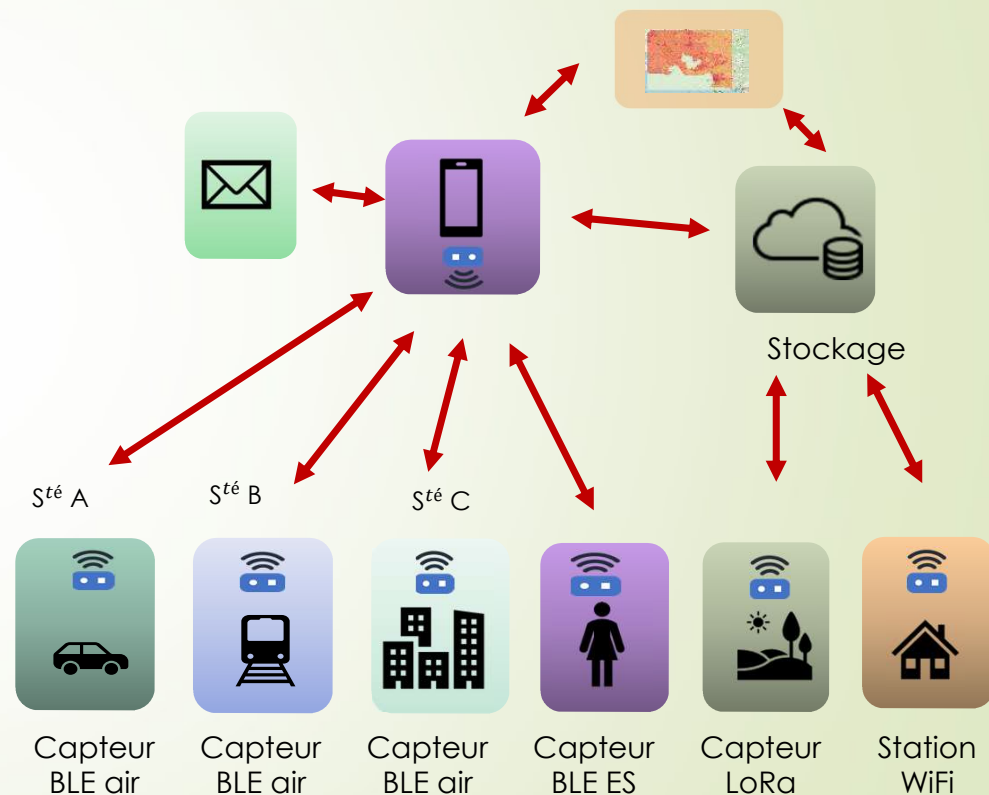
Exemple : « dosimétrie - réseau citoyen »

- Les lieux publics, les véhicules sont équipés de capteur BLE de mesure d'un polluant. L'application de mon téléphone se connecte aux capteurs que je croise et enregistre les valeurs mesurées.
- A la maison j'ai également ma station de mesure personnelle.
- En fin de journée, je sais consulter mon exposition aux polluants (y compris lors de mes trajets extérieurs).
- Je les raccorde à la base de données de mon choix (famille, quartier...)
 - Base de données accessible à tous
 - Ou base de données réservée à mon réseau
- Les données sont partagées et je peux les visualiser sous différents formats (graphiques, cartes...)
- Je peux transférer tout ou partie de la base vers une autre base.

Les capteurs sont standards et en mode affecté ou non affecté (pas d'appariement, compatibles toutes marques).

Les serveurs Cloud sont banalisés (données et API standards)

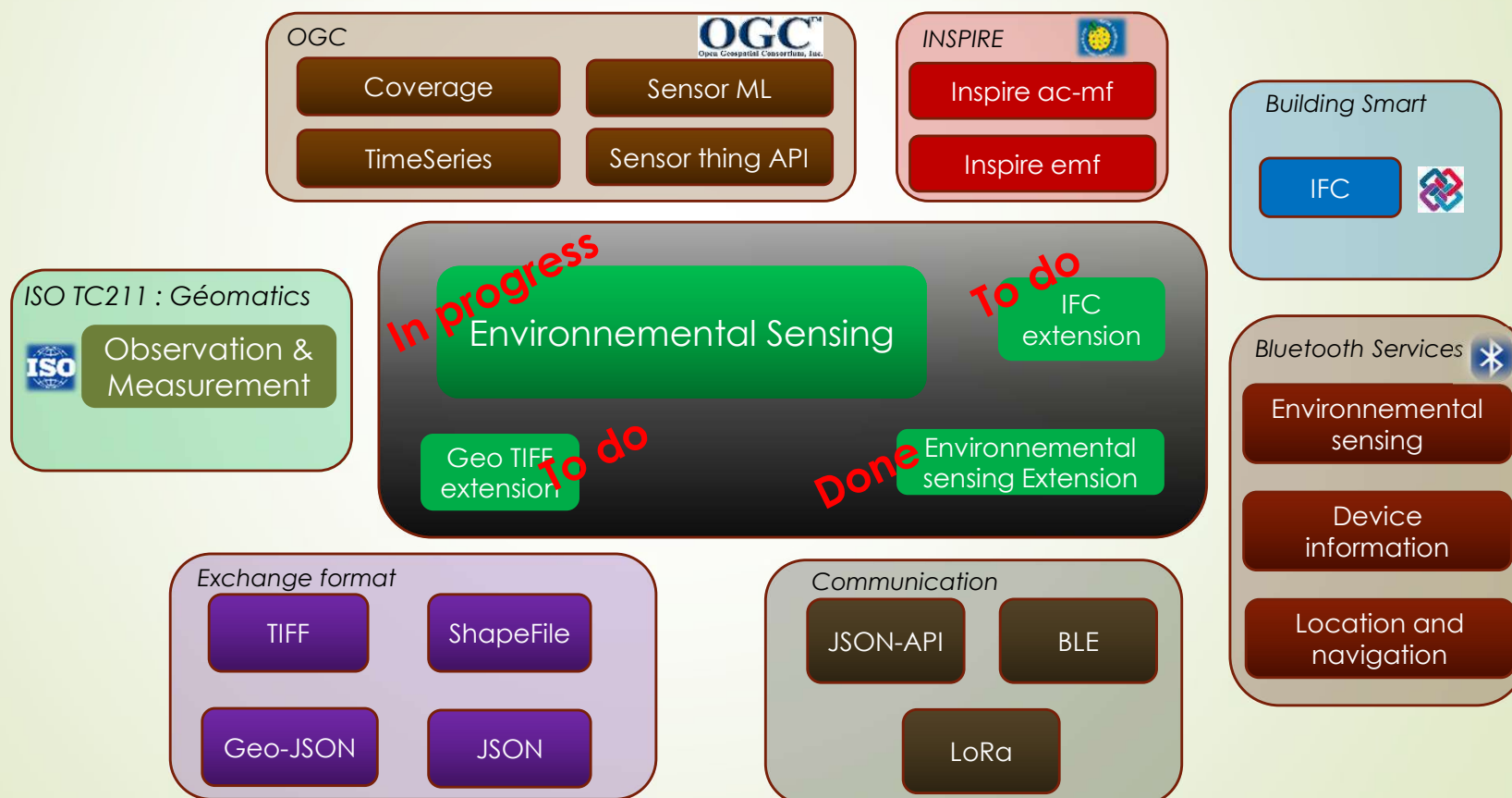
Les données récoltées par l'application de mon téléphone peuvent être stockées sur tout type de serveur ou envoyées à un ami (format d'échange standard).



Référentiel normes et standards

25

- Nombreux travaux de standardisation / structuration
- Domaines disjoints



Annexe : Données environnementales

Avis du conseil national du numérique – juillet 2020

« Faire des données environnementales des données d'intérêt général »

Dès lors, les données environnementales peuvent être définies de manière extensive comme toute donnée, par nature ou par destination, relative à l'environnement, à son état et/ou à ses flux d'interaction. Cette définition présente l'avantage de pouvoir qualifier plusieurs types de données – à l'instar des données agricoles, des données naturalistes ou des données relatives à la mobilité – comme des données environnementales.

1.2 Des usages multiples au service de l'intérêt général

Comme précisé dans l'un des objectifs de la feuille de route sur le numérique et l'environnement⁵⁰, les données environnementales pourraient, et devraient, être mobilisées au service de la transition écologique et solidaire et des objectifs de développement durable.

Avis du conseil national du numérique – juillet 2020

« Faire des données environnementales des données d'intérêt général »

Le Conseil considère que le partage de données environnementales est essentiel afin de fournir des informations utiles aux décideurs, citoyennes et citoyens et aux scientifiques⁵¹. Le partage répond aussi à des besoins économiques. Les finalités de l'ouverture de données (publiques ou privées) seraient dès lors les suivantes :

- Des finalités relatives à la conduite de politiques publiques et l'information citoyenne grâce au partage de données privées vers le Gouvernement ou le partage de données entre les pouvoirs publics. À cet égard, le groupe d'experts sur le partage de données privées vers les gouvernements a pu constater que le potentiel d'utilisation des données et de leurs résultats au profit de la société reste inexploité⁵². Le partage des données entre les pouvoirs publics est tout aussi nécessaire, car il peut également contribuer à l'amélioration des politiques et des services publics, et aussi à la réduction de la charge administrative à travers le principe du « dites-le-nous une fois »⁵³.
- Des finalités d'ordre économique à travers le partage de données entre acteurs privés ou le partage de données du Gouvernement vers les acteurs privés, telles que l'innovation, la recherche, le développement de nouveaux services, l'essor de l'intelligence artificielle ou encore le développement de la concurrence.

Avis du conseil national du numérique – juillet 2020

« Faire des données environnementales des données d'intérêt général »

II. La consolidation du statut des données environnementales d'intérêt général

Les données environnementales d'intérêt général peuvent donc être définies largement de manière à englober les données privées et publiques dont l'ouverture se justifie pour répondre à un motif d'intérêt général relatif à la transition écologique et solidaire. Cela pourrait permettre de transcender la distinction, entre données environnementales publiques et données environnementales privées.

Le Conseil envisage la question juridique des données environnementales d'intérêt général comme un processus de création d'un statut (1) afin de prendre en considération les différentes pratiques observées sur le terrain, combinant ainsi différentes approches qui peuvent être complémentaires et/ou cumulatives.

Le processus de création des données environnementales d'intérêt général se traduit par un régime conciliant maîtrise et ouverture pour les données environnementales d'intérêt général (2).

European Interoperability Framework for Smart Cities and Communities (EIF4SCC)

Recommendation #10

Use open standards and open technical specifications when developing local data platform and services, include multiple access and assistance channels to ensure that users can choose the option that best addresses their needs and/or preferences.

Recommendation #16

Establish, publish and maintain API design framework(s) to facilitate the automation of data sharing and data access for cities and communities to enable the development of (new) services and solutions.

Recommendation #24

Create more horizontal services towards local data platforms, to overcome silos within different domains, by encouraging collaboration and engagement among inhabitants, business, visitors, organisations and city/community administrators.

Recommendation #26

Use commonly agreed open standards and open technical specifications for achieving interoperability of data, systems, and services, to support cities/ communities and suppliers during the design, development and implementation phase of new services or reengineer existing ones (to avoid “reinventing the wheel”).

Recommendation #27

Use and raise awareness of the benefits of open standards and open technical specifications amongst SCC service providers, during the design, development and implementation phase.