

Proposition de mandat pour un groupe de travail « Qualification des données géographiques » de la Commission RMO du CNIG

Version du 28/11/18

PRÉAMBULE

Le Cerema et ses partenaires travaillent depuis plusieurs années sur la question de la qualification des données géographiques et de ses enjeux. Au niveau méthodologique, un décryptage des normes ISO, piloté par le Cerema, a donné lieu début janvier 2018 à la publication d'une série de fiches¹. Pour le CRIGE-PACA, le contrôle de la qualité des données est un élément essentiel de sa plateforme, et il a organisé en juin 2017 à Aix-en-Provence une journée technique² permettant d'identifier les enjeux et les principaux acteurs.

Afin de poursuivre cette dynamique, le CRIGE-PACA et le Cerema ont organisé un séminaire, du 6 au 8 février 2018, à Lyon, sur le thème : « Quels outils et organisations pour qualifier les données géographiques ? ». Ce séminaire était à vocation opérationnelle, sur la mise en pratique de la qualification des données géographiques. En regroupant une quinzaine de structures, publiques comme privées, qui ont pu travailler sur la qualité des données, le principe était de constituer un groupe pour à la fois favoriser et mettre en cohérence les initiatives de chacun. L'objectif affiché était celui de la construction d'un plan d'actions.

A l'occasion de ce séminaire est née l'idée d'un groupe de travail national sous l'égide du CNIG, pour amplifier la dynamique amorcée, et généraliser l'appropriation de la qualification des données géographiques, particulièrement les données ouvertes.

1 – ÉLÉMENTS DE CONTEXTE RÉGLEMENTAIRE

La diffusion des données géographiques s'accélère, et **de plus en plus de données sont disponibles**. Plusieurs dynamiques y concourent.

1.1 – Éléments réglementaires concernant l'ouverture et la diffusion des données

La directive INSPIRE

Depuis 2007, la directive européenne Inspire³ impose aux autorités publiques, d'une part de publier sur Internet leurs données environnementales géographiques, d'autre part de les partager entre elles. La diffusion des données est alors devenue une composante des politiques publiques grâce à la mise en place d'infrastructures de diffusion de données, de protocoles d'échange, de standards, etc.

L'open data transport

En 2015, la loi pour la croissance, l'activité et l'égalité des chances économiques (dite loi Macron) rendait obligatoire l'ouverture des données sur les transports publics et les services à la mobilité.

1 À retrouver ici : <https://www.cerema.fr/fr/actualites/serie-fiches-cerema-qualifier-donnees-geographiques>

2 <http://www.crige-paca.org/manifestations/journees-techniques/jt-n23-journee-qualite/jt-n23-journee-technique-sur-la-qualite.html>?

3 La directive européenne 2007/2/CE du 14 mars 2007, dite directive Inspire, vise à établir une infrastructure d'information géographique dans la Communauté européenne pour favoriser la protection de l'environnement. Elle est transposée dans le droit français depuis l'ordonnance du 21 octobre 2010.

La loi pour une République numérique

Le mouvement de l'OpenData, dont la récente loi Lemaire⁴ traduit la volonté politique française, milite plus largement pour une diffusion grand public des données. De nombreuses démarches locales de collectivités – accompagnées par l'association Open Data France – ainsi que la mise en place en 2011 de la mission Etalab au niveau national, ont conduit à la multiplication des portails de données ouvertes, et à la diffusion large de données de tout type.

Ces dynamiques constituent l'opportunité de s'interroger sur **la qualité des données mises à la disposition** du public et d'autres acteurs. En effet, au-delà de la seule « ouverture » de la donnée, et dans un but de fiabilisation de la réutilisation, il est important d'informer l'utilisateur potentiel sur la qualité d'une donnée, pour qu'il puisse savoir si elle correspond à ses besoins. Or, aujourd'hui, la grande majorité des données géographiques présentes sur Internet n'est pas qualifiée, ce qui engendre un risque important de mauvaise utilisation et d'incompréhension.

1.2 – La qualité dans le cadre de la réutilisation, un enjeu reconnu mais mal cerné

Au niveau international, la norme ISO 19157 donne le cadre de la qualification des données géographiques. Elle s'applique aux producteurs de données fournissant des informations de qualité pour décrire et évaluer la façon dont un jeu de données répond à sa spécification de produit et aux utilisateurs cherchant à déterminer si des données géographiques spécifiques sont ou non de qualité suffisante pour leur application particulière.

Mais dans une logique d'ouverture des données, il est aussi nécessaire de connaître la qualité externe d'un jeu de données, c'est-à-dire l'adéquation à un besoin exprimé. En d'autres termes, il serait utile mesurer l'aptitude d'un jeu de données à satisfaire des usages qui, dans le cadre de larges réutilisations, peuvent être divers. Cette notion est encore peu utilisée et très peu documentée, la difficulté venant sans doute de la complexité à traduire scientifiquement chaque famille d'usage en exigences de qualité et sortir des approches uniquement subjectives. Une première appréciation de la qualité externe est apportée au travers des méthodes de représentation simplifiée ou synthétique présente dans les fiches publiées par le Cerema.

Au niveau de l'État, la feuille de route de la supervision générale des données⁵ stipule en page 9 que : « la mise en qualité des données est un chantier prioritaire que les producteurs de données doivent intégrer pour permettre à leurs services, mais également à des réutilisateurs externes, un usage et une valorisation les plus adaptés ». De même, le rapport au gouvernement sur les données géographiques souveraines⁶ explique qu'une donnée souveraine doit faire autorité par : « sa qualité (...) [et] son caractère documenté : elle doit donc s'accompagner des informations descriptives – ou métadonnées – indispensables à l'utilisateur pour évaluer sa qualité. »

Au niveau des collectivités et des acteurs locaux, plusieurs actions voient aussi le jour. Le sujet de la qualité est traité parmi les plateformes régionales (cf. plus haut les actions du CRIGE PACA notamment). De même, le projet Validata⁷, co-porté par OpenDataFrance, montre l'intérêt pour le sujet (sur l'ensemble des données, pas forcément géographiques).

Il devient ainsi indispensable de mettre en place des **méthodes et des outils, accessibles à tous**, permettant d'estimer et de communiquer sur la qualité des données diffusées. En particulier, les grands objectifs du groupe de travail proposé ici seront de définir des méthodes qui :

- s'appliquent à tous les acteurs, notamment les collectivités ;
- permettent de rendre opérationnel la qualification des données géographiques ;
- mettent l'utilisateur au centre de la réflexion, en explorant notamment la qualité externe.

4 Loi pour une République numérique du 7 octobre 2016.

5 A retrouver ici : <https://www.ecologie-solidaire.gouv.fr/missions-du-superviseur-general-des-donnees>

6 A retrouver ici : <http://www.ladocumentationfrancaise.fr/rapports-publics/184000486/index.shtml>

7 En savoir plus : <http://www.opendatafrance.net/projets/validata/>

2 – MISSION

Le groupe de travail « Qualification des données géographiques » a pour but d'aider à la réutilisation des données ouvertes grâce à une meilleure information sur leur qualité.

3 – ORGANISATION ET FONCTIONNEMENT

Le règlement intérieur du CNIG s'applique à l'organisation et au fonctionnement du groupe de travail « Qualification des données géographiques ».

Le groupe de travail rapporte régulièrement, au moins une fois par an, l'état d'avancement de ses travaux à la Commission Règles de mise en œuvre du CNIG.

Le groupe de travail mène sa mission dans le cadre d'une large concertation, associant en son sein des utilisateurs, des producteurs, des représentants de l'État et des collectivités territoriales, ainsi que des fournisseurs de services liés à l'information géographique et des infrastructures de données géographiques.

Il coordonne les contributions de ces acteurs.

Le groupe de travail est ouvert et à géométrie variable selon les sujets abordés. Des appels à candidatures et/ou à contributions sont organisés au regard des problématiques traitées.

La composition initiale minimale du groupe de travail pourra s'appuyer sur le réseau de partenaires déjà constitué :

- Institutions nationales : le CGDD (MIG), le Cerema, l'IGN
- Plateformes régionales : Soluris, le CRIGE PACA
- Collectivités : la CC du Pays Réuni d'Orange, le Grand Lyon
- La recherche : le GDR MAGIS
- Des acteurs privés : Azimut, Oslandia

Cette composition est destinée à s'ouvrir à tous les acteurs intéressés.

4 – MÉTHODE DE TRAVAIL

Le groupe de travail agira selon trois axes :

- **l'accompagnement des diffuseurs de données** : aide au choix des critères les plus pertinents à mesurer selon les données, aide au remplissage des métadonnées et au rapportage de la qualité ;
- **l'accompagnement des utilisateurs de données** : avec une restitution des critères qualité de façon à pouvoir répondre à la question : « cette donnée convient-elle à tel usage ? » ;
- la **prise en compte du retour utilisateur** (crowdsourcing, évaluation...).

Pour atteindre ses objectifs, le groupe de travail s'appuie sur :

- quatre réunions par an avec accès à un système de visioconférence ;
- des échanges à distance via une plateforme collaborative comprenant un espace de dépôt de documents accessibles en lecture/écriture aux membres du groupe.

5 – PLAN DE TRAVAIL PRÉVISIONNEL

Pour chacun des axes cités au §4, les étapes seront les suivantes :

- Analyse de l'existant : qu'existe-t-il actuellement comme accompagnement des diffuseurs et utilisateurs de données ?
- Écoute des besoins : à partir des conclusions du séminaire de février, d'un échange avec

les participants du GT, et d'une enquête « utilisateurs » (qui peut s'appuyer sur l'enquête lancée par le Cerema au premier semestre 2018⁸), il s'agira de formaliser les besoins auxquels on veut répondre, et de les transformer en actions. Les actions potentielles sont par exemple : réaliser un tableau de choix des critères qualité en fonction du type de données, faire des préconisations sur la mesure de la qualité grâce au retour utilisateur, proposer une restitution simple de la qualité (« 5 étoiles ») en harmonisant la façon de la produire.

- Mise en œuvre des actions, avec définition d'un calendrier et de responsables par actions.

Actions concrètes réalisées dans la première année :

- Production d'un synopsis pratique pour la qualification : face à un type de données, quelle méthode et quel ordre des critères à mesurer ?
- production d'une première fiche qui validera la méthode évoquée ci-dessus (le groupe de travail déterminera le processus à mettre en place : type de fiche, choix éventuel d'un lot de données test, etc.).

6 – MOYENS MIS EN ŒUVRE

L'animation du groupe de travail est assuré par la MIG.

Le secrétariat technique est placé sous la responsabilité du Cerema.

7 – DURÉE DU MANDAT

Le présent mandat est défini pour une durée d'un an renouvelable.

Fait à Paris, le ___/___/ 2018, en 2 exemplaires

Pour la Commission
RMO du CNIG

Le président
Marc Léobet

Le