

Méthode pratique pour qualifier des données

version de travail, 5 juin 2020

Préambule

Le présent document constitue une méthode pratique pour qualifier les données géographiques. Il vise à proposer des éléments de mise en œuvre des [fiches méthodologiques](#) produites par le Cerema, et est réalisé dans le cadre du [groupe de travail du CNIG](#).

Les différentes étapes de la méthode sont illustrées par deux cas pratiques : la base adresse locale et les équipements collectifs.

Sommaire

1. Définir le périmètre et les objectifs.....	2
1.1 Que qualifie-t-on ?.....	2
1.2 Pourquoi ?.....	3
2. Cerner les usages.....	4
2.1 Thèmes.....	4
2.2 Usages.....	4
3. Hiérarchiser les critères qualité.....	6
4. Représenter les résultats.....	7
ANNEXES.....	8

1. Définir le périmètre et les objectifs

1.1 Que qualifie-t-on ?

La première étape dans un processus de qualification est de bien définir le lot ou la base de données que l'on veut qualifier, pour que les utilisateurs futurs puissent comprendre facilement à quoi s'appliquent les résultats de la qualification.

Ainsi il sera nécessaire de préciser :

- de quel point de vue on se place (utilisateur, contrôleur, diffuseur, etc.)
- si l'on se place dans le cadre d'un standard (local ou national) ;
- si le lot de données en question fait l'objet de spécifications pour sa production ;
- s'il fait l'objet localement de compléments dans sa conception ;
- et de façon générale toutes les règles – techniques ou métier – encadrant la production du lot de données.

Il sera également utile de lister les lots de données connexes, qui ne font pas l'objet de la qualification.

Illustration : base adresse locale

Tout d'abord une **définition** est fournie :

« Une Base Adresse Locale est un fichier géré par une collectivité locale (habituellement une commune ou un EPCI) et contenant toutes les adresses géolocalisées de celle-ci.

Elle est publiée sous sa responsabilité, ce qui lui confère un caractère officiel. » (source : [FAQ data.gouv.fr](https://www.data.gouv.fr/fr/faq/))

Ce qui permet de **préciser le type de données** que l'on veut qualifier, en distinguant :

- une base adresse : tout utilisateur peut en avoir une ;
- et une base adresse locale : il s'agit de la base adresse du référent local ; il y en a plusieurs, et au niveau France entière on peut disposer d'une collection de BAL, de responsabilité locale.

De même, la définition permet d'**exclure certaines données connexes**, comme les noms de rues, ou les registres des voies (exemple : FANTOIR).

Cela permet de préciser les mesures qualité à faire. Par exemple, on peut mesurer deux exhaustivités :

- l'exhaustivité des noms de rues ;
- l'exhaustivité des points adresse => c'est cette exhaustivité que l'on documente ici.

Enfin, on peut préciser si l'on s'appuie sur un **standard**, ou des **spécifications**. Dans notre cas :

« Le contrôle qualité se fait habituellement sur la base adresse. Par commodité, nous nous sommes basés sur le format d'échange BAL, mais en faisant des préconisations qui dépassent ce format. »

1.2 Pourquoi ?

La qualification des données peut avoir différents objectifs. Par exemple :

- connaître la qualité de données pour les corriger ;
- réaliser la recette de données récemment acquises ;
- documenter des données avant leur diffusion ;
- comparer plusieurs données ayant le même usage (par exemple des modes d'occupation du sol) ;
- documenter le niveau de qualité de données pour en faire une base de référence.

Déterminer le ou les objectifs avant d'effectuer les mesures qualité est important, à la fois pour étudier les bons usages, et pour préparer la restitution.

Illustration : la BD ADRESSE de la Métropole de Lyon

Texte extrait de Sign@ture n°68 : <https://www.cerema.fr/fr/centre-ressources/newsletters/signature/ndeg-68-avril-2019-zoom-qualification-donnees/qualite-donnees-geographiques-metropole-lyon>

En analysant la qualité de sa Base de données Adresse, la Métropole s'est aperçue que le taux d'exhaustivité n'était pas acceptable. Or **dans la logique d'en faire la source de données de référence sur le territoire**, il a donc fallu mettre en place un contrôle qualité.

La première étape a été de qualifier les manques pour aider les producteurs à mieux produire. Pour cela, le recours à des bases de données externes – elles-mêmes qualifiées – a permis d'analyser les différences. Par exemple, la base des propriétaires de locaux de la DGFiP, avec une bonne exhaustivité, permettait de recenser les données manquantes dans la BD Adresse.

D'autre part, dans la Base Adresse de la Métropole, chaque adresse est rattachée à une parcelle. Cette dimension est importante, car la Métropole fournit un service automatique de certificat d'adressage pour les notaires : cet outil applicatif nécessite des données de qualité forte en entrée. Il était donc aussi nécessaire de contrôler ce rattachement.

Sur des données de référence (comme l'adresse), il est difficile de trouver un référentiel pour contrôler la qualité. Suite à une tentative peu concluante de construire un autre référentiel avec différentes sources, le choix s'est alors orienté vers l'échantillonnage et le contrôle terrain.

2. Cerner les usages

Le choix des critères qualité à prioriser, ainsi que des modalités de restitution des mesures, dépendent des usages principaux des données qualifiées.

Une fois les lots de données à qualifier clairement définis, il est donc nécessaire de préciser les usages pour lesquels cette qualification doit être réalisée.

Ci-dessous : reprise du document de JM Arzac.

L'objet de cette partie est de définir, ou du moins d'initier la définition, des usages génériques de l'information géographique afin d'améliorer sa qualification et d'aider les différents publics utilisateurs à déterminer son « utilisabilité ».

Ces usages pourront être repris, précisés et complétés dans le cadre d'une démarche de qualification. Le tableau synopsis présenté dans la partie suivante reprend le principe de qualification suivant l'usage.

Il semble intéressant de qualifier en utilisant des couples thème / usage qui seraient relativement aisés à saisir et exploiter. Une proposition de thèmes et d'usages figure ci-dessous.

2.1 Thèmes

- Administratif et réglementation
- Défense et secours
- Gestion technique
- Planification et prévention
- Loisirs / tourisme
- Santé
- Environnement

2.2 Usages

- Recensement/dénombrement
- Localisation
- Cartographie (grand public, réglementaire, souveraine)
- Acheminement – navigation
- Analyse (statistique, spatiale)
- M2M, IoT (à préciser ?)

Illustration : les usages définis pour la Base adresse locale

Afin de préciser les modalités de qualification d'une BAL, nous avons défini plusieurs cas d'usage, regroupés en thèmes.

- ADMINISTRATIF
 - administratif : gestion / certificats d'adressage
 - fiscalité (localisation d'un local, déclaration d'une propriété)
- DÉFENSE / SECOURS
 - DECI (défense externe contre les incendies)

- Secours
- VIE QUOTIDIENNE
 - Navigation
 - Publipostage
 - Livraison postale
- TECHNIQUE ET GESTIONNAIRE
 - Localisation (ex : signalement citoyen de problèmes sur le territoire)
 - Géocodage
 - Aménagement du territoire (Gestionnaires FTTH)

3. Hiérarchiser les critères qualité

Le GT CNIG a conçu et testé un « synopsis pratique pour la qualification » : il s'agit, face à un type de données, de déterminer une méthode et un ordre des critères à mesurer.

Le principe du tableau « synopsis »

Le « synopsis » prend la forme d'un tableau avec :

- en colonne les mesures qualité retenues dans les fiches méthodologiques du Cerema ;
- en ligne différents champs de description permettant d'aider à la réalisation de la mesure concernée (voir le descriptif des lignes ci-dessous).

Importance intrinsèque du critère	Note de 1 à 5 permettant de montrer si la connaissance du résultat de la mesure qualité est indispensable à l'utilisation de la base (note de 5) ou si elle est anecdotique (note de 1).
Cas d'usage pour lesquels le critère est important	Avant de remplir cette ligne, il est nécessaire de définir des cas d'usages du lot de données (dans un document à part). On indique ensuite, pour chaque mesure, si elle est particulièrement importante pour un ou plusieurs usages.
Points de vigilance	Conseils pour la mesure de la qualité.
Pistes d'évolution des spécifications	Ligne facultative : elle permet d'indiquer les limites des spécifications.
Dépendance à des données externes	Ligne facultative : liste des données de référence ou nécessaires pour la bonne réalisation de la mesure.
Besoin d'outils complémentaires	Ligne facultative : permet de liste des idées pour des outils d'aide à la qualification.

Ce synopsis a été testé sur deux lots de données type (BAL et EquiCo), et les tableaux remplis, ainsi que des notes d'accompagnement, sont disponibles en annexe.

4. Représenter les résultats

La représentation des résultats de la qualification est fondamentale, pour aider les utilisateurs à comprendre quels usages ils peuvent faire des données.

Il s'agira de représenter de façon claire :

- les principaux résultats de la qualification ;
- les usages utiles et ceux déconseillés.

Il est également recommandé d'autoriser la contribution des utilisateurs.

RÉSUMÉ : PLU de Numérac

NOTE DE L'EVALUATEUR	NOTE DES UTILISATEURS (16)	VOTRE AVIS	
★★★★★	★★★☆☆	DÉPOSER UN AVIS >	
SOUS-NOTES			
EXHAUSTIVITE	★★★★★	PRECISION THEM	★★★★★
PRECISION GEO	★★★☆☆	COHERENCE LOGIQUE	★★★★☆

Usages recommandés

- ✓ Analyse statistique
- ✓ Calculs de surface

Usages proscrits

- ✗ Instruction des autorisations d'urbanisme
- ✗ Représentation cartographique

Illustration 1: Maquette pour une représentation possible de résultats qualité

ANNEXES

Une carte mentale permet de retrouver pour chaque critère :

- les définitions ;
- l'ensemble des mesures selon qu'elles sont présentes dans le synopsis, dans les fiches méthodologique, ou dans la norme.

Et pour chacun des cas étudiés (BAL et EquipCo) :

- un arbre de décisions pour aider l'utilisateur à qualifier les données ;
- le synopsis rempli ;
- une note d'accompagnement.