



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

IGN

INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

**Institut national de l'information géographique
et forestière**



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*



INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

VEILLE SUR LE SUJET QUALITÉ ET RETOUR SUR LE Q-KEN

Nicolas PY



1. Veille sur le sujet qualité

Outils de commentaire des utilisateurs sur les documents/spécifications : l'exemple du Sandre

[http://
id.eaufran
ce.fr/ddd/E
TH/2#Cou
rsEau](http://id.eaufrance.fr/ddd/ETH/2#CouresEau)

RÉFÉRENTIEL HYDROGRAPHIQUE

URI: <http://id.eaufrance.fr/ddd/ETH/2>
Code Sandre: ETH
Version: 2
Statut: Validé

Le référentiel hydrographique vise à décrire les entités hydrographiques présentes sur le territoire français afin de constituer un référentiel national permettant de localiser des données relatives à l'eau.

Télécharger :   

Noter cette fiche : ★★★★★

Sommaire:

BASSIN FLUVIAL	LIMITE TERRE-MER
BASSIN HYDROGRAPHIQUE	NOEUD HYDROGRAPHIQUE
BASSIN VERSANT TOPOGRAPHIQUE	OBJET HYDROGRAPHIQUE
COURS D'EAU	PLAN D'EAU
CROISEMENT	SEGMENT HYDROGRAPHIQUE THEMATIQUE
ENTITE HYDROGRAPHIQUE	SURFACE ELEMENTAIRE
ENTITE DE TRANSITION	TOPONYMES DE L'OBJET HYDROGRAPHIQUE
GENEALOGIE DES OBJETS HYDROGRAPHIQUES	TRONCON HYDROGRAPHIQUE

BASSIN FLUVIAL ✎

Zone dans laquelle toutes les eaux de ruissellement convergent à travers un réseau de rivières, fleuves et, éventuellement, de lacs, vers la mer, dans laquelle elles se déversent par une seule embouchure, ou un seul estuaire ou un seul delta.

Correspond au concept « RiverBasin » d'Inspire.

Nouvelles fonctionnalités relatives aux schémas de données sur data.gouv.fr [1/2]

Spécifier qu'une ressource respecte un schéma

Il est désormais possible de spécifier de manière déclarative, depuis l'interface d'administration de data.gouv.fr, qu'une ressource d'un jeu de données respecte un schéma de données.


La spécification de ce champ sur une ressource entraîne l'ajout d'un panneau d'information depuis la page de son jeu de données. Ce panneau permet au lecteur de découvrir la documentation relative au schéma et de vérifier que la ressource est bien conforme au schéma identifié.


Ces nouvelles fonctionnalités témoignent d'un rapprochement entre schema.data.gouv.fr et data.gouv.fr. Nous avons détaillé [dans un article de blog en anglais](#) les futures fonctionnalités envisagées : suggestion de schémas, consolidation de ressources respectant un même schéma, validation proactive de ressources, assistance aux producteurs de données etc.


Dernière version consolidée

Titre * Dernière version consolidée

Type * Fichier principal

Description  Prévisualiser

Date de publication 

Schema 

- Base adresse locale
- Budget des collectivités et établissements publics locaux
- Catalogue simplifié
- Equipements
- Lieux de covoiturage
- Lieux de stationnement
- ✓ Schéma IRVE
- Schéma SCDL Délibérations
- Schéma SCDL Subventions
- Spécification du fichier de déclaration de profil d'acheteur

URL * Taille Taille

Format * csv

Type Mime text/csv

Somme de contrôle sha1 Somme de contrôle

Annuler Enregistrer

Nouvelles fonctionnalités relatives aux schémas de données sur data.gouv.fr [2/2]

Remarques personnelles

(note: captures d'écran du 19/01/2021)

On constate que la plateforme met l'utilisateur en capacité de vérifier qu'un jeu de donnée respecte un schéma, et d'être informé des erreurs détectées.

-> On constate que le choix a été fait de favoriser la diffusion de données y compris si elles contiennent des erreurs, et de rendre l'utilisateur (averti et courageux) capable d'auditer le respect du schema

On constate que le remplissage d'un attribut 'structure de donnée' permet d'agrèger tous les jeux de données ayant le même attribut, donc la même structure.

-> Agréger ne signifie pas dé-doublonner si nécessaire, ni corriger...

-> Cette approche me semble risquée, elle conduit aisément à ce que le ré-utilisateur néglige de s'interroger sur la qualité. A l'inverse, l'approche GPU est stricte sur le respect des spécifications, au prix d'un coût d'animation (DDI-DREAL-IGN-DHUP)

The screenshot shows a modal window on the data.gouv.fr website. The modal is titled "Schéma associé" and contains the following text: "Cette ressource est associée au schéma **etalab/schema-irve** ce qui signifie qu'elle doit en respecter la spécification." Below this, there is a "Documentation" section with a button "LIRE LA DOCUMENTATION" and a "Validation" section with a button "VALIDER LA RESSOURCE". The background shows a list of resources, including "irve-driveco-210112.csv".

Une consolidation de données est effectuée à partir de ce schéma. Vous pouvez retrouver le jeu de données résultant de cette consolidation sur data.gouv.fr

The screenshot shows a data card for "Fichier consolidé des Bornes de Recharge pour Véhicules Électriques (IRVE)". The card includes the etalab logo, the title, a description: "Aménageurs, collectivités, producteurs de données : retrouvez ici la documentation complète pour référencer vos bornes. Contexte Depuis le décret n° 2017-26 du 12 janvier 2017,..." and metadata: "30 11", "22 stars", "France", and "Point d'Intérêt".

Consolidation 21713 erreurs

La dernière version de ce jeu de données consolidé comporte 16898 lignes et un total de 21713 erreurs. Le dernier rapport de validation peut être [consulté en ligne](#).

BAL 1.2

AITF SIG Topo sur Twitter :

"Publication de la révision 1

.2 du format d'échange de données voies-adresses locales dit

"format #BAL«

https://twitter.com/aitf_sig_topo/status/1335948386340573185

-> Commentaires QuaDoGeo transmis à l'AITF par Clément

Pistes d'évolution des spécifications pour le format BAL

Retour du GT CNIG QuaDoGeo

Lien : http://cnig.gouv.fr/?page_id=18183

Le GT CNIG QuaDoGeo porte l'organisation du contrôle qualité de données géographiques autour de 5 grands critères.

Un groupe de participants a élaboré un parcours qualité afin d'étudier les points de contrôle critiques qui permettent d'assurer une analyse fine des données Adresses.

L'organisation du contrôle de ces données et du format BAL associé a permis de soulever plusieurs pistes d'évolution que le GT CNIG souhaite partager avec l'AITF. Pour chaque critère analysé, nous proposons de vous remonter les informations suivantes :

- 1. Critère d'exhaustivité :**
L'enregistrement des « Lieux-dits » doit être précisé. Est-ce que la présence d'un numéro est obligatoire pour ce type d'entité ?
- 2. Critère de cohérence logique :**
La notion de domaines de valeurs est primordiale dans l'analyse du jeu de données. Nous proposons une prise en compte plus large des pratiques locales sur la position (ex : position = « secours »)

Le contrôle des valeurs obligatoires est aussi un point important et nous pensons qu'il faudrait rendre obligatoire la présence des coordonnées dans un des deux systèmes.

Il serait aussi important d'imposer l'égalité des coordonnées si elles sont présentes dans les deux systèmes.
- 3. Précision thématique :**
L'analyse du contenu d'un attribut tel que le nom de voie nous oblige à définir des règles d'écritures plus strictes. Nous pensons qu'il serait important de préciser la manière d'écrire les noms de rues, en interdisant les abréviations ou encore en cadrant l'utilisation des apostrophes.
- 4. Précision de position :**
Afin d'assurer le meilleur usage des données par des tiers, il serait important d'intégrer au modèle de données des attributs relatifs à la qualité du positionnement, comme nous pouvons en retrouver dans d'autres jeux de données nationaux.
- 5. Qualité temporelle :**
Un certain nombre de contrôles complémentaires peuvent aussi être ajoutés concernant les attributs de type date. Ainsi, nous pourrions demander à assurer une certaine cohérence des dates de mise à jour avec les dates de création de rues ou entre les dates de mise à jour des adresses d'une même rue.

Nous proposons aussi d'intégrer un attribut pour porter la date de la dernière vérification « confirmation terrain » de l'existence de l'adresse.

L'ensemble des documents annexes à ces propositions d'évolution sont disponibles sur la page dédiée au GT QuaDoGeo sur le site internet du CNIG.

Pour rappel, le projet de registre français des métadonnées de qualité des données géographiques fait l'objet d'un appel à commentaires jusqu'au 31 décembre. <http://cnig.gouv.fr/?page=24438>

Pistes d'évolution des spécifications pour le format BAL
Retour du GT CNIG QuaDoGeo

14/12/2020

L'open data des communes sur le territoire grandlyonnais

Le PDF : <https://www.millenaire3.com/content/download/35654/458236>

<https://www.millenaire3.com/ressources/L-open-data-des-communes-sur-le-territoire-grandlyonnais>

-> raconte (**quasi verbatim**) la démarche de standardisation/collecte/diffusion OD de données produites par les collectivités (bureaux de vote, parcs et jardin, toilettes publiques, panneaux d'affichage, établissements recevant du public, stationnement pour Personnes à Mobilité Réduite [PMR]) du Gd Lyon.

-> Sa lecture (le style est agréable) ne vous/nous apportera rien que nous n'ayons déjà au moins entrevu dans QuaDoGeo, mais l'ensemble dresse un panorama qui complète l'observatoire de l'OD.

à la définition d'une structure homogène pour chaque jeu de données. L'objectif de l'expérimentation n'était en effet pas seulement que les communes publient des jeux de données similaires, mais aussi que la structuration de ces données soit identique. Pour cela, un travail itératif s'est effectué entre la Métropole, le groupe de travail et les communes pour collecter les informations et les homogénéiser avant la diffusion sur data.grandlyon.com.

sensibiliser à ces enjeux. Elle contribue également à structurer l'information sur les jeux de données choisis, et, pour certaines communes, d'en faire la donnée de référence pour les services. Pour la Métropole de Lyon, l'expérimentation contribue à enrichir et faire évoluer sa stratégie en matière de données à plusieurs titres (conventions partenariales, plateforme technique, cible de la politique, etc.).

L'ouverture des données est un processus long qui requiert un investissement conséquent de la part des agents engagés. Or, ce travail est souvent peu visible et peu valorisé au sein des administrations ou par les hiérarchies, l'ouverture des données

L'ouverture des données implique un travail conséquent de production des données qui pose la question de l'industrialisation du processus

L'analyse de la première phase a permis de mettre en évidence que l'information était éparpillée au sein de multiples systèmes d'information métiers qui ne communiquent pas toujours entre eux. À cet égard, de manière contre-intuitive, la grande taille de la collectivité peut être un frein à l'ouverture des données : si les grandes collectivités ont davantage de moyens financiers, techniques et humains, elles ont également des systèmes d'information plus lourds à faire évoluer, ce qui est moins présent chez les communes plus petites qui se trouvent alors davantage en position d'agilité. Du côté de la Métropole, la première phase d'expérimentation a montré que l'offre actuelle d'accompagnement est appréciée des communes, mais implique un temps non négligeable d'animation du collectif et de nettoyage des données. Si le déploiement de cette offre d'accompagnement est envisagé à une plus grande échelle, il paraît nécessaire de penser l'industrialisation du processus pour faciliter les mises à jour et éviter la multiplication des bases de données similaires au sein des collectivités.

RENNES: Le Service métropolitain de la donnée - Deux années d'amorçage

Le communiqué (

<https://rudi.datarennes.fr/deux-annees-damororage-du-service-public-de-la-donnee-de-rennes-metropole>

) est repris et commenté par <https://www.banquedesterritoires.fr/rennes-metropole-tire-un-bilan-mitige-pour-son-service-public-de-la-donnee>

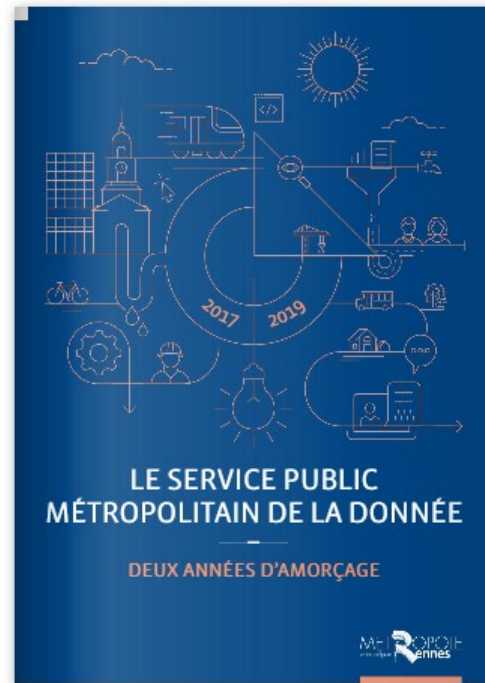
www.banquedesterritoires.fr/rennes-metropole-tire-un-bilan-mitige-pour-son-service-public-de-la-donnee

« Rennes métropole vient de publier le bilan de la préfiguration d'un "service public métropolitain de la donnée" (SPMD) pour partager et valoriser les données d'intérêt territorial. Si le nombre de données effectivement libérées est très limité, l'expérience montre que le partage des données est une question au moins aussi technique que culturelle. Un bilan sans concession qui intéressera tous les territoires intelligents engagés dans une stratégie de gouvernance de la donnée. »

« Le bilan de ces chantiers thématiques se révèle plus que mitigé : **le projet s'est rapidement heurté aux réalités des silos informatiques, aux coûts de mise à disposition et à la sensibilité de nombreuses données.** Pour l'eau, le rapport relève "des données dans des systèmes complexes" qui doivent être "restructurées" avec des "traitements plus ou moins manuels" et des coûts importants à la clef. Dans le secteur de l'énergie, les contraintes techniques se doublent de celles du RGPD qui confèrent aux acteurs privés des "marges limitées". »

« Au total, si **le catalogage des données territoriales a légèrement progressé, peu de jeux de données sont ouverts de manière pérenne.** Le nombre de jeux ouverts n'est du reste pas mentionné, le rapport se contentant de mentionner 2,03 Go de données partagées. Une misère. »

« "l'expérimentation doit être jugée non seulement au prisme des cas d'usages développés, du volume et de la diversité des données échangées, mais aussi à l'aune des résultats acquis en termes de sensibilisation d'acteurs, de création de relations de



Boîte à outils pour la gestion des données territoriales

<https://www.banquedesterritoires.fr/gestion-des-donnees-territoriales>

+ le webinar "Gestion de la donnée: quels outils et quelle stratégie pour les territoires". (28/01/2021, <https://app.livestorm.co/cdc-caisse-des-depots/gouvernance-de-la-donnee>)

Guide des bonnes pratiques contractuelles et recommandations

Pour la mise en place d'une gouvernance de la donnée territoriale

1 GARANTIR LA SOUVERAINETE DE LA COLLECTIVITE SUR SES DONNEES 8

- 1.1 Définir le statut des Données publiques 9
- 1.2 Garantir la Propriété des données 10
- 1.3 Définir les données d'intérêt général 12
- 1.4 Contrôler l'hébergement et les conditions de stockage des données publiques 14

2 PROTEGER LES DONNEES 16

- 2.1 Protéger les données à caractère personnel 17
- 2.2 Garantir la sécurité des systèmes d'information 20
- 2.3 S'engager en faveur de la sobriété dans la collecte et la conservation des données 21

3 GARANTIR LA TRANSPARENCE 23

- 3.1 Mettre en œuvre l'open data des données publiques 24
- 3.2 Développer l'open data des algorithmes 29

4 FAVORISER DES NOUVEAUX USAGES 32

- 4.1 Encadrer les droits de propriété intellectuelle sur les innovations technologiques 33
- 4.2 Garantir la réversibilité des outils technologiques 36

Gestion des données : Quels outils et quelle stratégie pour les territoires ?

An aerial night-time photograph of a city street grid. The streets are highlighted with a bright green glow, creating a complex network of lines over the dark urban landscape. The buildings and structures are visible as dark shapes between the glowing lines.

2. Proposition de livrable

Modèle de document de spécification

Les métadonnées de qualité sont un chantier Quadogeo, une autre approche serait également d'aider les producteurs à rédiger les spécifications/documentations des données qu'ils diffusent.

Pour cela, nous proposons la création d'un livrable Quadogeo qui serait un document textuel modèle de spécification:

- Chapitré,
 - et ces chapitres marqués obligatoires/facultatifs
- Illustré d'exemples, de contre-exemple
- Inspiré
 - Descriptifs de contenu (DC) et de livraison (DL) de l'IGN
 - Gebru&All2020_Datasheets for Datasets_V7.pdf
 - Mooney&All2016_Towards a Protocol for the Collection of VGI Vector Data.pdf
 - <https://collaborative-data.theodi.org/>
 - ISO19157
 - <https://github.com/sigagglcompiegne>, la description du SIG de l'agglomération de Compiègne
- Relié au registre, aux fiches, ...
- Tenant compte de besoins IGN au regard de sa stratégie (souhait d'identifiants uniques et pérenne, de métadonnées d'actualité si possible à l'objet, de capacité à générer des différentiels entre 2 dates).
- Des volontaires ? AITF, Cerema, ... ?

3. Pour votre information

Document technique

Document à destination des utilisateurs techniques

Note: une version retravaillée et complétée (contrôles serveurs, mises à jour mineures, reformulations et explications) sera mise à disposition sur le site geoservices à cote des documents habituels BDTPOPO, accompagné d'une note de blog sur le sujet

Au fil de la production

Des contrôles :

- à chaque modification d'objet (auto complétion, valeurs contraintes, ...)
- à chaque envoi d'une mise à jour vers le serveur par le collecteur
- quotidiens sur le serveur
- trimestriels avant chaque diffusion



CONTÔLES BDUNI ET BDTPOPO

Une validation continue au fil de la production...



© IGN

4. Q-KEN

An aerial night photograph of a city street grid. The streets are illuminated with a vibrant green light, creating a glowing pattern against the dark, shadowed buildings and rooftops. The perspective is from a high angle, looking down on the city's layout.

Q-KEN

Le Quality Knowledge Exchange Network échange, entre instituts cartographiques, sur le management de la qualité et la qualité des données.

Le 19 novembre 2020, l'IGN y a présenté

- Les travaux du GT CNIG QuaDoGeo
- Sa stratégie partenariale et de consommation de données de producteurs de référence (collectivités, opérateurs, ...).

Summary

1. The context

- a. National Council for Geographic Information
- b. Web geographic data
- c. Foundations
- d. Data quality, a never ending story (to tell)
- e. Seminar of birth

2. WG QuaDoGeo (qualification of geographic data)

- a. WG QuaDoGeo : Mandate
- b. #Season 1
- c. #Season 2

3. The QuaDoGeo's products

- a. Seminar's product
- b. Survey
- c. Registry proposal
- d. Also

4. French mapping agency as a consumer of geographic open data

- a. The strategy
- b. With national providers
- c. With regional providers
- d. The tools

BKG quality concept for geodata and geoservices

Influence factors and different levels of the QM

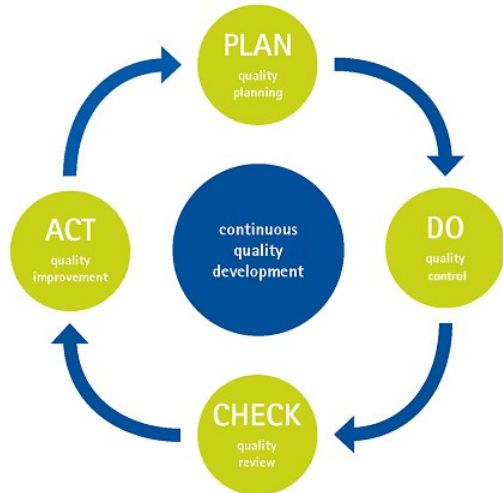
- Superordinate - strategic level



- Authority (operational) level



The 19 basic principles of the quality concept



Quality management in general

1. **QM** is a **cyclical process**.
2. The BKG's quality management is oriented towards **user needs**.
3. The BKG's quality management is based on relevant international **standards** (e.g. ISO and OGC standards) and the TR BGeoRG and is oriented to the SDI-Germany.

Organisational structure of QM, responsibilities

4. The **quality manager** monitors the quality management system.
5. **Product managers** are responsible for maintaining and improving the quality of their product for the **entire life cycle**.
6. A **central inspection unit independent** of production is responsible for the quality control of the geodata products.

Quality Control

7. For product managers, the **editing and analysis rights for demand reports** are secured.
8. The **quality targets** are defined in the smallest possible units for each product group or product-specific, they result in the target status, are consistent in all reference documents, are updated annually and compared with the TR BGeoRG.
9. The quality targets are **transparent and measurable** using suitable test scenarios.
10. The test methods and tools are selected on a **product-oriented** basis.
11. **Quality control** is carried out separately from the production process and documents the quality of the geodata products on the basis of test results.

Documentation

12. At BKG, **all work steps of the product creation** are documented.
13. The **results of the quality control** and **final check** are documented for each product and the documents are stored permanently.
14. The quality information is documented in **machine- and human-readable form**.
15. Documents of BKG products are stored in a **central location**, free access is ensured.
16. The central reference document of QM is the cross-product **quality manual (handbook)**.
17. The **reference documents** are **published online** and made feely available to users.
18. The **in-house documents** are immediately updated and versioned in case of changes.
19. The **public documents** are updated at least annually.



Nota QuaDoGeo:

Les principes 9-10 font écho à la démarche QuaDoGeo de proposer une sélection d'indicateur qualité selon le produit

Le principe 11 rappelle l'importance d'un contrôle qualité indépendant de a production (à l'IGN, unité contrôle qualité dédiée)

Le principe 14 impose la difficile cohérence et synchronisation entre paramètres d'outils et spécifications

Les principes 17 et 19 rendent disponibles à l'utilisateurs les spécifications de qualités

New common method for declaring data quality in Denmark

The declaration template



Why?

Need for authorities to talk same language when assessing the quality of data from widely different domains. The goal is to make better use of data across the public sector.

How?

By using four quality elements the new data quality declaration should enable a better and more secure way to have a dialogue about data quality across the public sector.

What?

A common public declaration template and underlying vocabulary.

Who?

Public authorities and private companies are on voluntary basis welcome to use the declaration method.

DQ Dimension	Measurement and description <i>Describe the quality of the dataset in relation to each dimension.</i>	Assessment in relation to use <i>Fill in the appropriate number of stars.</i>
completeness		☆☆☆☆☆
accuracy		☆☆☆☆☆
currentness		☆☆☆☆☆
reusability		☆☆☆☆☆

☆☆☆☆☆	Rated unsuitable for the given application
★☆☆☆☆	Is judged to be poorly used
★★☆☆☆	Can be used subject to a greater number of errors / unintended results
★★★☆☆	Can be used subject to errors / unintended results
★★★★☆	Can be used subject to individual errors / unintended results
★★★★★	Can be used immediately and unconditionally

		number of stars
Completeness	<p>What proportion of the individuals' dataset describing according to its specification is currently represented in the data set? - Is this an estimate?</p> <p>The completeness check is carried out by sampling in different numbers of samples and the number of grids. The inspection is performed as a visual inspection of the use of spring orthophoto from the year in which the data is updated (support data for centerline of rivers and lake from two additional years of spring orthophoto). SDFE, based on DS / EN ISO 19157: 2014, designates objects for the samples.</p> <p>99 percent of buildings visible in spring orthophoto from 2017 are represented in the topographic basic data set building theme. The percentage is calculated from random checks carried out in August 2018 according to DS / EN ISO 19157: 2014. The GeoDanmark specification allows for the completeness of 99 percent of buildings visible in Orthophoto.</p>	★★★★☆

Same as critical for
a user

Questions:

1. How to define criticality for intended use?
2. Should all core dataset be treated as critical to all users?
3. Can dataset have different data quality description for different feature classes? - If yes how do you produce a quality description/markings for the entire dataset?
4. How do you communicate quality in different levels to the end user?

