

GT QUADOGÉO DU CNIG

Présentation des travaux sous l'angle du FAIR

Nicolas PY
Stéphane ROLLE
Animateurs du GT



La qualité est un sujet actuel



Utilisateur« Piloter par la donnée »Bien choisirAvoir confiance

Posture de producteur insuffisante, quelque part entre l'acquisition et la diffusion

Observatoire OpenDataFrance, 2020, « La qualité intrinsèque des données est encore à améliorer [...] A titre d'exemple, la fraicheur des données n'est pas respectée dans près de 92% des cas !»

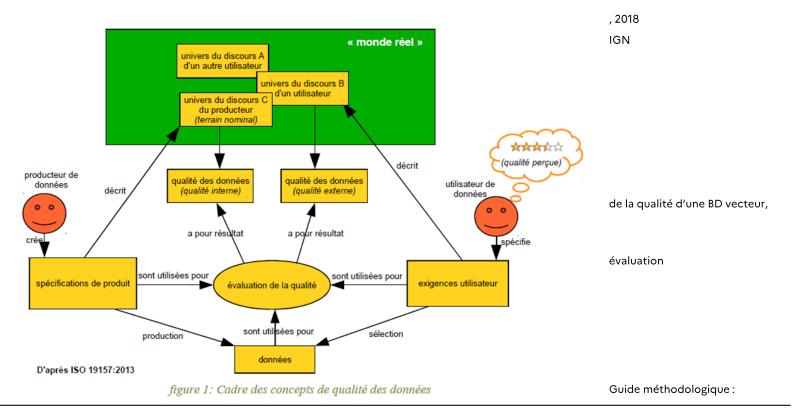
Cnépote, Schignard, 2020, « 10 ans d'ouverture des données publiques »

Bothorel, 2020, rapport pour une politique publique de la donnée, des algorithmes et des codes sources Montchalin, 2021, Transformer l'action et la fonction publiques par la donnée - Feuille de route 2021-2023 du ministère de la

Transformation et de la fonction publiques



Le sujet en un schéma





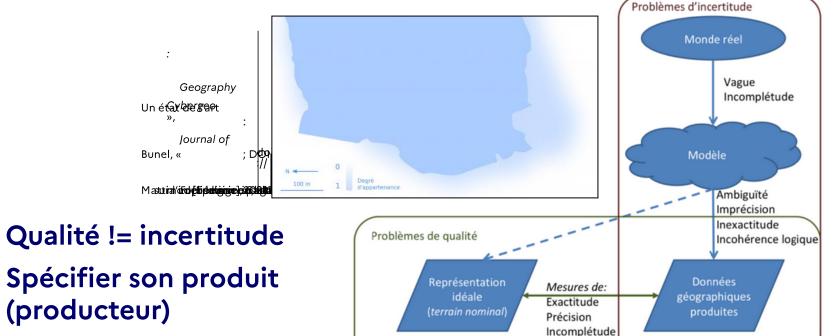
Le GT

- Le Conseil national de l'information géolocalisée organise la coordination et accompagne l'évolution de l'information géolocalisée en France (présidence Bertrand Monthubert)
 - https://cnig.gouv.fr/spip.php?page=sommaire
- Commission règles & qualité (présidente Marie LAMBOIS)
 - https://cnig.gouv.fr/commission-regles-qualite-a653.html
- Groupes de travail (GT)
 - https://cnig.gouv.fr/liste-des-groupes-de-travail-de-la-commission-a25951.html
- GT QuaDoGéo pour QUAlité des DOnnées GÉOgraphiques (animation Nicolas PY (IGN) et **Stéphane ROLLE (CRIGE PACA))**
 - https://cnig.gouv.fr/gt-quadogeo-a18183.html
- 3+1 réunions par an, ouvertes, livrables et documents publiés*





La qualité, avec ou sans FAIR, préalables



- Spécifier son produit (producteur)
- Spécifier son besoin (utilisateur)

Figure 2.4. Cadre conceptuel de l'incertitude

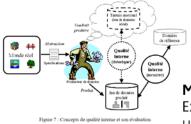
geogra

donne

httbbess//w



En tant que producteur, s'engager pour la qualité et au FAIR



Modélisation

Ex: adresse, attribut unique ou combinaison d'attributs ?

Edition

Contrôle au relâchement de clic

Contrôles périodiques, systématiques, automatiques

Ex, continuité des réseaux

Documentation utilisateur

Ex, descriptifs de continu, de livraison, de suivi des évolutions



_Produit

_Qualité attendue

De saisie

Détaillées, définitions claires, exemples

Processus

Organisation de la production, responsabilités, livrables,...

Persistance

Contrôle au moment de l'écriture physique de l'édition, au moment de la réconciliation

Contrôles ad hoc

Ex, ré-acquisition terrain par échantillonnage pour mesure des écarts

Ecoute

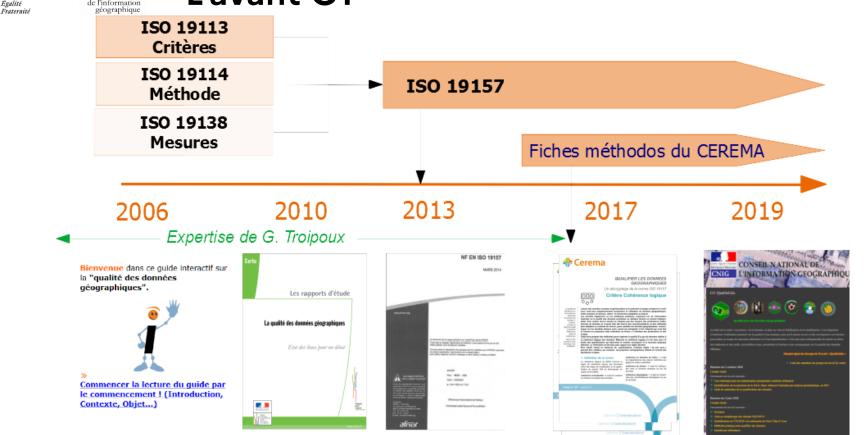
Ex, dispositif de recueil de signalements, SAV

.By design -> |





L'avant GT



Journée Interopérabilité et Innovation - 2024

7

, rapport

QuadoG

Bilan 202

Gallais

Sourregles of

oeuvre

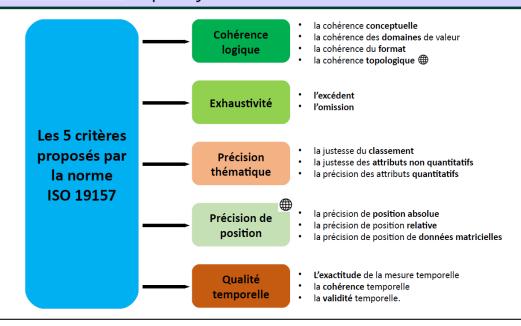


Le GT, aider à l'ISO 19157

ISO 19157



"Data producers provide data quality information to state how well a dataset represents the real world as defined by a data product specification. Users can then evaluate the data quality information to determine a dataset's fitness for use"





Le FAIR exige plus (1|4)

Etaternité L'existence de la donnée

- ·Son existence (pas de coûts d'amorcage à sa création)
- ·Sa profondeur historique, ses éditions antérieures
- ·La stabilité de sa structure dans le temps, ou l'existence
- d'un produit « série historique »

oISO 19158

- Description du contrat de confiance entre le producteur et le client
- D'autres critères, dont:
- ·Volume de livraison (~résolution des données)
- ·Calendrier / échéance de livraison
- ·Coûts de production et/ou de mise à jour

La fiche n° 47 du CNIG conclut sur la qualité des données « La qualité des données localisées n'est pas une référence absolue et résulte d'un arbitrage entre la réponse à un besoin, un coût et des délais de production... Beaucoup d'utilisateurs n'imaginent pas quels autres usages pourraient être faits des données qu'ils génèrent. L'information géographique numérique est un domaine où chaque utilisateur doit aussi apprendre à connaître les métiers connexes au sien, où les données qu'il crée pourraient être réutilisées moyennant un effort modeste »

In <Fiche n°90 du CNIG - 2006>

<u>o</u>Evaluation indirecte

- •Métadonnées: usage suggéré, exemples d'usage, généalogie (<u>Inspire</u>!), formulation des textes libres dont le commentaire et le résumé
- ·Producteur: crédibilité, connaissance du territoire, expérience
- ·Autres utilisateurs: recommandations, feedbacks positifs, réputation du producteur



Le FAIR exige plus (2|4)

Critères intrinsèques	Définitions
Unicité	Chaque entité du monde réel est représenté par un et un seul objet métier : absence de doublon et d'amalgame
Complétude	Les données sont complètes (instances et leurs caractéristiques) par rapport à la réalité et aux processus de collectes
Exactitude	Les données sont exactes et égales à la réalité qu'elles sont censées représenter. Ce critère englobe la notion de précision et de validité.
Conformité	Respect des contraintes liés aux données
Intégrité	Les relations entre les objets métiers sont cohérentes et présentes
Cohérence	Les caractéristiques des données sont cohérentes entres elles et avec les autres données considérées.

Figure 4 - Qualité des données : les critères intrinsèques

Critères Services	Définitions
Accessibilité	Facilité d'accès aux données (synchrone / asynchrone, unitaire / masse, push / pull, open API).
Cohérence	Gestion dans le temps de la distribution des données (problématique de synchronisation et de resynchronisation).
Actualité	Rapport entre les données et le temps, ou degré de mise à jour des données. La fraîcheur des données doit tenir compte de la réalité organisationnelle et réglementaire.
Pertinence	Mesure de l'utilité d'un ensemble de données, ou de l'adéquation de ces données à des usages

Figure 5 - Qualité des données : les critères de services

Critères Sécurité	Définitions
Disponibilité	Aptitude du référentiel à remplir une fonction dans des niveaux de services définis Propriété d'une information d'être, à la demande, utilisable par une personne ou un système.
Intégrité	Les données doivent être celles que l'on s'attend à ce qu'elles soient, et ne doivent pas être altérées de façon fortuite ou volontaire. Propriété assurant qu'une information n'a pas été modifié ou détruit de façon non autorisée.
Confidentialité	Seules les personnes, les applications ou les processus autorisées ont accès aux informations qui leur sont destinées. Caractère réservé d'une information dont l'accès est limité aux seules personnes admises à la connaître pour les besoins de leurs missions, ou aux processus autorisés
Traçabilité (preuve ou imputabilité)	C'est la garantie de disposer des éléments qui apportent la preuve des traitements ou autres évènements relatifs aux informations considérées.
Lisibilité	le codage d'une information doit à tout moment être décodable. L'obsolescence des codages peut ainsi conduire à des recodages, dans l'objectif du maintient de cette lisibilité.

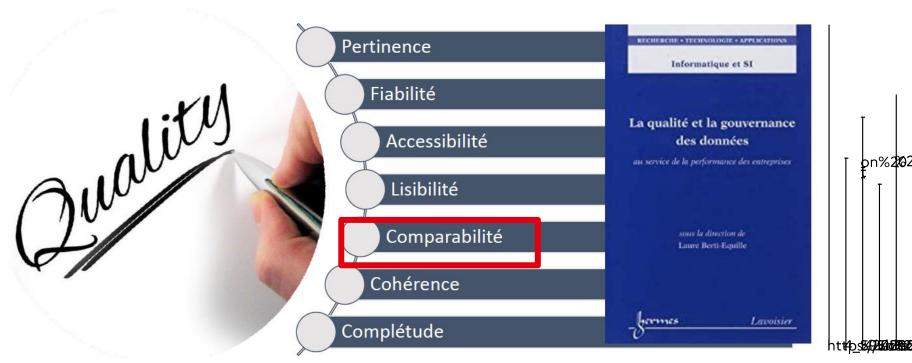
Figure 6 - Qualité des données : les critères de sécurité

Le Cadre Commun d'Architecture des référentiels de données de l'État

10



Le FAIR exige plus (3|4)



Dans l'intervention de Pascal Rivière (chef de l'inspection générale de l'INSEE) au Bercy Hub le 28/03/2023 et intitulée « Data, de quoi parle-t-on ? Parole d'expert »



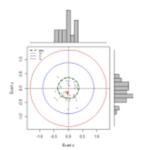
Le FAIR exige plus (4|4)

L'existence de documentation de la donnée

Descriptifs de contenu, de saisie, de livraison, de l'évolution

L'existence de méta-documentation de la donnée, et sa diffusion

- · Rapports de production, de contrôle qualité
- Descriptifs des processus de production (Terrain, Edition, transaction client-serveur, mise en base, archivage, ...)
- · Informer des changements, des nouveautés



IGN	Fiche A.13	SBV.BDUnI/IT/042
INSTRUCTANTINAL DE L'IMPORANTON GEOGRAPHININE ET FREESTIERE	Decharche des trous externes	Edition : 3 Date : 25/09/2018 Page : 1/1

Définition du contrôle

1 — Types/sous-types/attributs concernés Toutes les classes surfaciques Grande Échelle sauf 'Parc ou réserve', 'Forêt publique'

'Zone d'activité ou d'intérêt' et les classes de Métadonnées



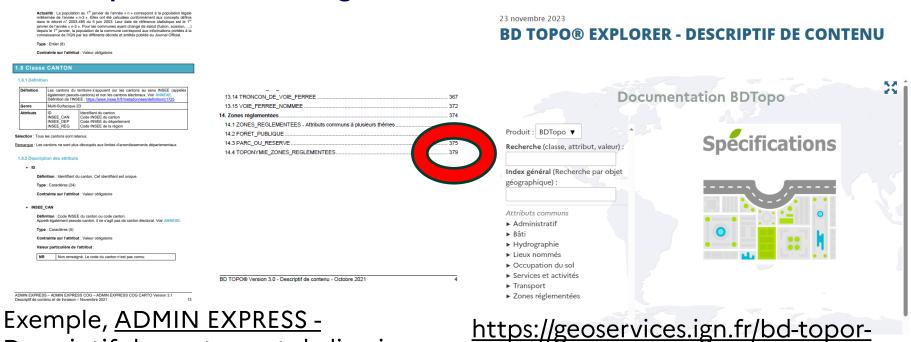




Quelques trucs FAIR de l'ère GT, @IGN (1|2)

explorer-descriptif-de-contenu

Les spécifications navigables



Journée Interopérabilité et Innovation - 2024

Descriptif de contenu et de livraison

13



Quelques trucs FAIR de l'ère GT, @IGN (2|2)

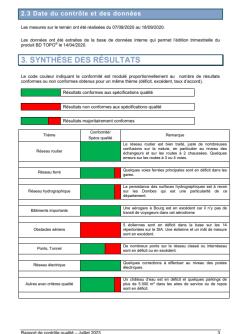
La diffusion d'une version publique des rapports de contrôle

qualité

https://geoservice s.ign.fr/document ation/donnees/ve cteur/bdtopo/rap port-controlequalite

https://geoservice s.ign.fr/document ation/donnees/or tho/bdortho/rapp ort-controlequalite





4 DECLUITATO	

4.1 Résultats comparés aux spécifications de qualité

Le tableau ci-dessous regroupe les taux de déficit, d'excédent et d'accord les plus significatifs recueillis lors du contrôle sur le ferrain.

Ces taux sont comparés aux exigences des spécifications de qualité du produit BD TOPO®; les codes couleurs utilisés font ressortir les non-conformités.

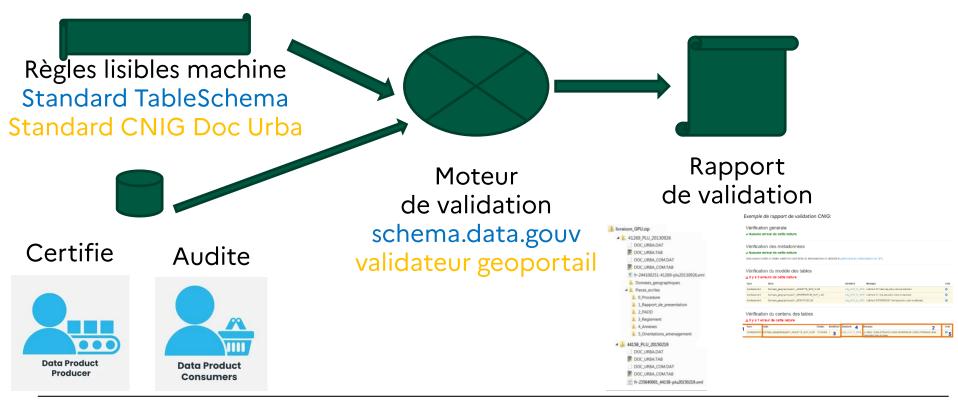
	Exigences				Dépa	rtement 01		
	ince aux spéci lité de la BDU				Résultat Contrôle Départemental			
Taux de déficit	Taux d'excédent	Taux d'accord	Nature	Taux de déficit	Taux d'excédent	Taux d'accord	Nb Terrain	
-			Transport					
2%	5%	98%	Route classée	<0,1%	0,2%	99,9%	2 005 507	lg
1%	2%	96%	Réseau routier non classé	0,1%	0,1%	99,4%	2 656 130	lg
1%	2%	98%	Réseau revêtu	0,1%	0,1%	99,3%	4 730 166	lg
2%	5%	98%	Route Numérotée (complexe)	0%	0%	100%	362	nb
3%	5%	97%	Liens vers route nommée (tronçon)	<0,1%	<0,1%	>99,9%	2 024 783	lg
		95%	Parking-Péage	20,2%	0%	79,8%	658 470	sur
1%	1%	99%	Voies Principales et LGV	0,9%	0%	99%	225 070	lg
1%	1%	95%	Aire de triage	0%	0%	100%	509 826	sur
		100%	Piste d'aérodrome en dur	0%	0%	100%	119 865	sur
		100%	Piste d'aérodrome en herbe	0%	0%	100%	76 258	sur
		95%	Aéroport - Catégorie	0%	0%	100%	3	nb
5%	5%	90%	Transport par câble	0%	0%	100%	8 697	lg
			Hydrographie					
5%	5%	95%	Tronçon hydrographique permanent	0,2%	0,1%	99,8%	927 627	lg
3%	5%	97%	Surface hydrographique permanente	<0,1%	0,1%	89%	42 576 390	sur
5%	5%	90%	Détail Hydrographique	4,5%	0,4%	95,5%	561	nb

Rapport de contrôle qualité - Juillet 2023

Taux de déficit	Taux d'excédent	Taux d'accord	Nature	Taux de déficit	Taux d'excédent	Taux d'accord	Nb Terrain		
			Service et Activités						
		100%	Préfecture	0%	0%	100%	1	nb	
		100%	Sous-préfecture	0%	0%	100%	3	nb	
		95%	Mairie	2,5%	2,5%	97,5%	202	nb	
		95%	Gare	0%	11,1%	100%	9	nb	
		95%	Aérogare	-	100%	-	0	nb	

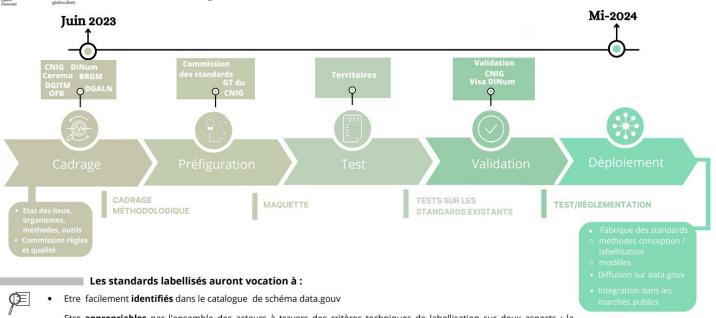


Quelques trucs FAIR de l'ère GT, les validateurs





Quelques trucs FAIR de l'ère GT, la fabrique des standards incluant des seuils qualité



Exemple, à terme, pour être réputé respecté, le standard Y exige un taux d'exhaustivité de 95%.



Etre **appropriables** par l'ensemble des acteurs à travers des critères techniques de labellisation sur deux aspects : la méthode de construction et le formatage technique (respect d'un cahier des charges commun)

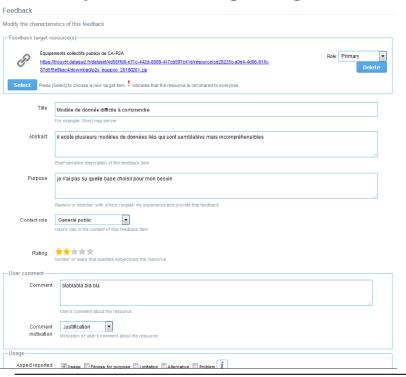


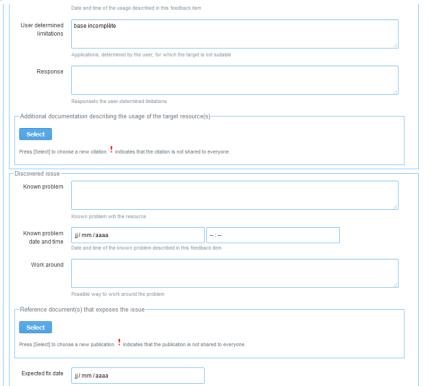
Rendre possible leur intégration à la réglementation, ou dans le cas où la réglementation crée un standard, répondre à cette obligation. A plus long terme, sous réserve de faisabilité juridique, l'intégration dans les CCAG des marchés publics pour assurer l'adhésion générale aux standards labellisés.



Quelques trucs FAIR de l'ère GT, le geospatial user feedback

https://www.ogc.org/standards/guf







Quelques livrables du GT

- Formations/webinaires à la qualité de la donnée
- Sélection restreinte des 21 critères iso les plus utiles -> « Syst. de publication de registres » du Géocatalogue
- Parties qualité et métadonnées des modèles de standards
- Prototype de métadonnées qualité et de leur représentation





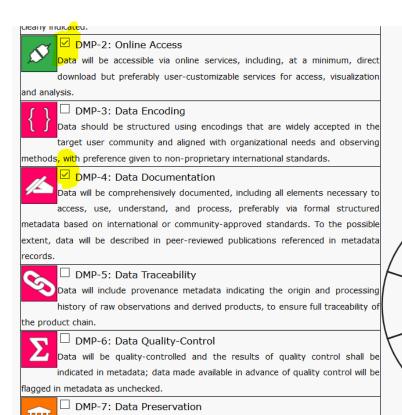
PS: le FAIR a des amis

The GEO Label

The intergovernmental Group on Earth Observations (GEO), through its members and participating organisations has been a strong proponent of open data. [...]

For this reason GEO developed the Global Earth Observation System of Systems (GEOSS) Data Management Principles (DMP) to promote and encourage best practices that will result in users being able to derive the most benefit from data obtained via GEOSS. [...]

A data provider can self-declare conformity of their own datasets with the GEO label. The data provider should review their operations and the dataset against the specifications set out for each data management principle. They can then declare conformity with between 1 – 10 data management principles underpinning the GEO label.



19

GE3label

DMP 2016-1



Nous rejoindre? Prochaine réunion le 25/01/24 @ https://webconf.numerique.gouv.fr/GTCNIGQuaDoGeo2024

oLe GT Quadogéo est documenté via un espace de stockage partagé Google Drive, dont le propriétaire est le compte <u>gtquadogeocnig@gmail.com</u>. Cet espace héberge les documents, les comptes rendus et les travaux finalisés.

Certains travaux sont cependant publiés là où ils sont le plus utiles, tel le registre de mesures Quadogéo https://data.geocatalogue.fr/ncl/mesuresQuaDoGeo.

Pour ses travaux, le GT Quadogéo emploie tout outil nécessaire, notamment le répertoire https://github.com/cnigfr/quadogeo de l'espace Github du CNIG https://github.com/orgs/cnigfr.

Le GT Quadogéo communique à l'extérieur (mise en ligne des comptes rendu, communiqués) via les mécanismes de mailing "lettre d'info CNIG"

Le GT Quadogéo publicise son action sur différents médias :

- •Le forum GeoRezo, https://georezo.net/forum/ à l'initiative et en leur nom des membres du GT, ou via le compte https://georezo.net/forum/profile.php?id=93961 QuadoGeo
- •Le réseau social Twitter via le compte https://twitter.com/quadogeo

Pour participer (produire), se tenir informé, faire/partager des retours d'expériences







Liberté Égalité Fraternité

MERCI DE VOTRE ATTENTION