RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

CNIG
Conseil national de l'information géolocalisée

Commission règles & qualité

Groupe de Travail Qualité des Données Géographiques (QuaDoGéo)

Groupe de Travail animé par :

Crige
Centre de Ressources en Information Géographique
Provence-Alpes-Côte d'Azur

IGN
INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE

COMPTE-RENDU DE REUNION (distanciel)

Objet : Réunion #17 du GT QuaDoGéo du 18 octobre 2023.

Ordre du jour :

1. Accueil
2. Validation du précédent compte-rendu
3. Date du prochain GT
4. Informations sur les formations
5. Suivi des actions
6. Partage d'expériences
7. Mandat
8. Points divers

Liste des participants

Animateurs : Stéphane Rolle (CRIGE PACA) et Nicolas Py (IGN)

Nom Prénom	Organisme	Présent	Excusé
AÏT-AMMAR Vincent	Ecolab	X	
ARSAC Jean-Marie	AZIMUT	X	
BALDE Mamadou-Bailo	Ineris	X	
BLANCHARD Guillaume	GIP ATGeRi	X	
BOUTET Jerome	Idéo Ternum Bourgogne-Franche-Comté	X	
DESCHAMPS Héloïse	GIP ATGeRi	X	
GALLAIS Arnauld	Cerema Ouest	X	
GERVAIS Bertrand	someware.fr		X
GRIGNON Paul	Ecolab	X	
LAUNAY Sebastien	IGN / Dépt Contrôles Qualités	X	
Le MOAL Mathieu	Axes Conseil	X	
PY Nicolas	IGN Centre Est	X	
ROLLE Stéphane	CRIGE PACA	X	
THIARD Jérôme	someware.fr	X	

Réunion #17 du GT QuaDoGéo du 18 octobre 2023 (distanciel).		
Mise en relecture du compte-rendu	06/10/2023	Stéphane Rolle / Nicolas Py
Relecture du compte-rendu	17/11/2023	Les participants.
Validation du compte-rendu	25/01/2023	Stéphane Rolle / Nicolas Py
Prochain rendez-vous le 25 janvier 2024 matin		

1. **Accueil**

2. **Validation du précédent compte-rendu**

Aucune remarque à signaler sur le précédent compte-rendu disponible sur la page QuaDoGéo ici (permalien).

3. **Date du prochain GT**

La prochaine réunion est envisagée en janvier 2024, entre le 15 et le 26 janvier ; choix via un sondage envoyé aux membres du GT.

Répondre à https://framadate.org/5OLw2rxvF7teCE4J
Un rappel sera effectué le 20/11/2023.
Le sondage sera clos le 24/11/2023.

4. **Informations sur les formations**

Cet automne concentre une actualité riche sur le plan des formations à la qualité des données géographiques.

Via **Idéo Ternum Bourgogne-Franche-Comté** tout d'abord, qui porte en novembre le rejeu des webinaires gratuits CEREMA/CRIGE PACA « Qualité avec QGIS » et « Qualité avec PostGIS » :

- Le 7 novembre, « Qualité des données géographiques avec QGIS », https://ideo.ternum-bfc.fr/agenda/7-novembre-2023-classe-virtuelle-qualite-des-donnees-avec-qgis (permalien)
- Les 14, 11 et 28 novembre (3 sessions à distance d'environ 1h30), « Qualité des données avec PostGreSQL /PostGIS», https://ideo.ternum-bfc.fr/agenda/14-novembre-2023-classe-virtuelle-qualite-des-donnees-avec-postgresql-postgis-session-1 (permalien)

Ces webinaires ne sont pas réservés uniquement aux Bourguignons et Francs-Comtois bien qu'au besoin priorité sera donnée aux habitués du réseau Idéo Ternum BFC. Le public cible est prioritairement les géomaticiens publics.
Note **S.Rolle** : Il y a une forte demande pour ces types de formation. Les participants à la première édition ont été contactés par Stéphane afin de savoir s'ils avaient transformé leur participation en une action/process dans leur structure. Deux retours positifs exprimés, qui seront valorisés lors d'une prochaine réunion QuaDoGéo.

S.Rolle/CRIGE PACA intervient dans le cursus de 2 formations :

- Au sein d'https://www.idgeo.fr, via un module d'1,5 jour en demi classe ; pour une partie théorie et une partie pratique
- Au sein du master 2 "Géomatique et conduite de projets territoriaux" d'Avignon, pour une séquence générique sur la qualité

L'**IGN-ENSG** a dispensé la formation « Management de la qualité appliquée aux SIG » les 12 et 13 octobre derniers (2 jours) ; rarement dispensée faute d'inscrits en nombre suffisants. Son contenu https://ensg.eu/fr/management-de-la-qualite-applique-aux-sig comprends https://ensg.eu/files/ensg/2023-02/formation_220.pdf soit :

- Management de la qualité à l'échelle de l'organisation
- Annotation/dénotation, échantillonnage,

- Contrôle qualité image
- Qualité intrinsèque, extrinsèque, interne, externe, by design
- Unité contrôle qualité de l'IGN, méthodologie générale, charge de travail, détails sur les méthodes de contrôle de produit

Discussion collective sur d'autres formations qui pourraient exister.
J.Boutet indique le récent https://www.univ-amu.fr/fr/public/actualites/participez-au-nouveau-webinaire-du-cedre-qualite-des-donnees-donnees-de-qualite (données non géographiques), qui permet à **S.Rolle** de rappeler notre souhait de recrutement de contributeurs 'recherche' dans le GT ; **A.Gallais** suggère les enseignants chercheur des écoles d'ingénieurs et université.

5. Suivi des actions

Une action concerne la reprise des documents QuaDoGéo sur le nouveau site du CNIG. Stéphane et Nicolas disposent depuis peu des droits de rédaction des pages QuaDoGéo ; il manque une consigne CNIG sur le nommage à adopter pour les répertoires et documents.

Une action concernait l'interrogation des participants aux formations/webinaires CEREMA/CRIGE PACA « Qualité avec QGIS » et « Qualité avec PostGIS » afin de savoir s'ils avaient transformé leur participation en une action/process dans leur structure. Le **Grand Lyon** interviendra lors de la prochaine séance QuaDoGéo pour présenter le contrôle qualité de leurs acquisitions de nuages de points PCRS.

Une action d'animation, au long cours, concerne la prise de contacts pour susciter des présentations et retours d'expérience dans le GT. L'INSEE, le SHOM, la DGAC ont été approchés, avec un accord de principe INSEE suite à une réunion **S.Rolle/N.Py/INSEE** de présentation mutuelle le 10/10/2023.

Une action concernait la rédaction d'un billet pour Geotribu (https://contribuer.geotribu.fr/) afin de présenter le GT, son historique, ses livrables, et proposer de nous rejoindre. **J.Boutet** s'est déclaré volontaire pour initier cette tâche.

Décision/actions :
J.Boutet: Leader sur l'action de rédaction d'un billet pour Geotribu
N.Py/S.Rolle : Courriel aux membres du GT pour adjoindre des équipiers à Jérôme

Une action portée par **M.Lacroix** concernait le modèle de standard CNIG, avec l'objectif d'y insérer des items qualité. Cette action a vraisemblablement vocation à migrer au sein de « la fabrique des standards » évoquée au conseil plénier du CNIG de mai 2023.

La fabrique des standards : de l'artisanat d'excellence à l'industrialisation

Constat

- Besoin d'outils (standards) pour suivre et piloter la planification écologique au niveau territorial. Les instances de standardisation ont besoin de réunir les compétences métiers de chaque thématique.
- **Les travaux du chantier FNV ont identifié plusieurs besoins :**
  - **Centraliser les standards sur une plateforme** afin que les utilisateurs les moins expérimentés puissent s'y retrouver.
  - **Clarifier les rôles** et **organiser l'articulation** entre les standards et les schémas de données de la transition écologique en créant des **règles communes sur les modèles de standards.**
  - **Une coordination multi-échelle** à tous les niveaux des territoires. Le pôle de coordination avec les territoires du CNIG assure la coordination pour l'information géolocalisée
- Le CNIG est largement identifié comme étant un acteur de référence pour la coordination et la production de standards de données géolocalisées avec une expertise technique de ses usages, ainsi que de nombreux travaux en cours en commission des standards et en commission règles et qualité.
- La règlementation encadre la création et l'évolution de référentiels pour un certain nombre de sujets (eau, biodiversité, milieux marins). Certaines structures de référentiels sont reconnues par les acteurs de leur secteur (SANDRE, SARM, CARET, GIS Sol, SINP).
- La plateforme **"schema.data.gouv"** diffuse des schémas de données, test d'intégration de schémas du CNIG en cours.

Propositions

Le CNIG initie avec les autres acteurs de la standardisation la mise en place d'un dispositif national avec :

- Une méthode de conception et un modèle commun de standards.
- Un dispositif de labellisation et la possibilité d'intégrer des standards dans la règlementation.
- Une plateforme unique de diffusion des standards labellisés : **schema.data.gouv.**

Feuille de route

Juin 2023 — Mi-2024

CNIG DiNum Cerema BRGM DGITM DGALN OFB — Commission des standards GT du CNIG — Territoires — Validation CNIG Visa DiNum

Cadrage | Préfiguration | Test | Validation | Déploiement

- Etat des lieux, organismes, méthodes, outils
- Commission règles et qualité

CADRAGE MÉTHODOLOGIQUE | MAQUETTE | TESTS SUR LES STANDARDS EXISTANTS | TEST/RÈGLEMENTATION

- Fabrique des standards
  - méthodes conception / labellisation
  - modèles
- Diffusion sur data.gouv
- Intégration dans les marchés publics

Les standards labellisés auront vocation à :

- Etre facilement **identifiés** dans le catalogue de schéma data.gouv
- 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.

Décision/actions :
L'action est éteinte dans le fichier de suivi QuaDoGéo, avec cependant évidemment une incitation à contribuer au sein du GT à venir « Fabrique des standards ».

6. Partage d'expériences, le projet Ecosphère

Paul GRIGNON (Chef de projet data/IA and Green Data for Health SRI/**ECOLAB**) rappelle l'objet des projets Ecosphère(s) et green data for health (gd4h)
CF. présentation (permalien)

L'ambition du projet Ecosphères est de référencer de manière pérenne et sur chaque territoire les données (ouvertes et restreintes) utiles à l'appui des politiques publiques des MTECT - MTE répondant aux chantiers France Nation Verte. Le projet est porté par le Commissariat général au développement durable (CGDD) et son Ecolab, en s'appuyant sur les grands producteurs de données environnementales, les territoires et le portail data.gouv.fr.

- ☐ Il existe des difficultés à découvrir les données, auquel l'équipe pense remédier en créant des parcours utilisateurs centrés sur des cas d'usages débouchant sur des jeux de données mis en cohérence pour être rassemblés en pack, en y adjoignant un éclairage sur la qualité des données et des conseils d'utilisation.

Le projet green data for health https://gd4h.ecologie.gouv.fr/ est une plateforme de référencement (pas d'hébergement) de données environnementales pour les acteurs de la santé environnement contenant actuellement environ 150 ressources.

- ☐ Il existe un besoin de mieux documenter les conditions d'accès, le contenu, la qualité des données.
- ☐ Cela peut passer par une charte OpenData proposée aux utilisateurs, afin qu'ils deviennent acteurs de la montée en qualité des données et métadonnées, au travers notamment du recueil des réutilisations faites et du feedback (sur le catalogue, sur les métadonnées, sur la qualité des données, sur la documentation, sur l'usage fait, ...)
- ☐ Au premier semestre, gd4h a fait l'objet de défis de tous types (dataviz, jointure, dashboard, ...) https://challenge.gd4h.ecologie.gouv.fr/defis/ avec des développeurs bénévoles pour constituer des POC transférables et pérennes (cadre de naissance de l'outil qualigeoenvi).
- ☐ Naissance d'un GT CNIG « GT données géolocalisées en santé» (https://cnig.gouv.fr/gt-donnees-geolocalisees-en-sante-a26103.html) au sein de la Commission besoins & usages ; prochaine session le 09/11/2023.
- ☐ **A.Gallais** interroge sur les données emblématiques, qui sont les données environnementales utiles pour la santé-environnement. **N.Py** signale les travaux Pigma (intervention de ce matin notamment) pour le bon remplissage des métadonnées pour une expérience inspirante en sus évidemment des mises en œuvre sur le catalogue data.gouv. **N.Py** interroge sur la contradiction entre principe OpenData de moindre friction imposée à l'utilisateur et l'ambition de la charte. **P.Grignon** confirme que la charte constitue le principal levier. Un autre levier pourrait être de favoriser la prise en compte des enjeux de qualité des données dans la façon de saisir les réutilisations (e.g. associer un data paper). Des data papers ont ainsi été produits pour certains des défis développés dans le cadre du Challenge Green Data for Health. Voici le lien vers l'ensemble des outils : https://gitlab.com/data-challenge-gd4h, vers un exemple de jeu de données publié sur data gouv et de data paper associé (défi HistoIRIS porté par Santé publique France) : https://www.data.gouv.fr/fr/datasets/historique-des-codes-iris/ et https://hal.science/hal-04176155. **N.Py** interroge sur la capacité organisationnelle des producteurs à recevoir des critiques/commentaires (une personne nommée, un guichet pour recevoir des signalements, ... Exemple IGN https://espacecollaboratif.ign.fr/). **P.Grignon** confirme être en train d'échanger à ce sujet avec des producteurs de données. L'objectif est évidemment de prendre en compte les capacités organisationnelles de traitement des feedbacks par les producteurs de données.

7. Partage d'expériences, l'outil Qualigeoenvi

Mamadou-Bailo BALDE, géomaticien à l'**INERIS**, présente https://qualigeoenvi.fr/ (code source https://gitlab.com/data-challenge-gd4h/qualigeoenvi/-/tree/main/, licence MIT) né dans le cadre des défis gd4h.
CF présentation (permalien)

Les fonctionnalités de qualigeoenvi sont de permettre de détecter, sans être expert, des doublons, coordonnées inversées, auto-intersections, géocoder, vérifier la cohérence interne coordonnées/adresse géocodée dans un jeu de données au format CSV.
L'INERIS apprécierait un portage de pérennisation. **A.Gallais** suggère une piste en proposant un jumelage/fusion avec le validateur https://validata.fr/.

8. Partage d'expériences, ECOLAB, métadonnées qualité dans le SI d'Ecolab/MTE

Vincent AÏT-AMMAR, **ECOLAB**, présente les actions qualité de la donnée à Ecolab et revient sur l'atelier Ecosphère(s) du 13/10/2023.
CF. presentation (permalien)

L'ambition du projet Ecosphère(s) est de créer des bouquets de données pertinents et utiles au pilotage des politiques de la transition écologique, avec une forte dimension territoriale. Le référencement des données s'appuie sur des cas d'usage, **S.Rolle** a cité lors de l'atelier la typologie QuaDoGéo.
Après enquête, deux des freins à la réutilisation sont la découvrabilité et la qualité des données.
La feuille de route ministérielle prévoit depuis 2021 une série d'actions sur la qualité des données. Cette année, Ecolab s'intéresse notamment à l'appréciation de la qualité des données à partir de leurs métadonnées. Cette action a fait l'objet d'un atelier ayant eu lieu le 13/10/2023, recentré sur les besoins des gestionnaires de patrimoines de données.

- ☐ **J.Boutet** signale les dispositifs d'observation de plateforme : https://geobs.cnrs.fr/, https://doi.org/10.3166/rig.2019.00074 ; ainsi que la charte de la BAN https://adresse.data.gouv.fr/bases-locales/charte
- ☐ **A.Gallais** demande si les métriques qualité sélectionnées par QuaDoGéo (le registre) seront considérées ; Vincent annonce avoir échangé à ce sujet avec M.Lambois et le GT Métadonnées, notamment son mapping DCAT https://github.com/cnigfr/metadonnee.
- ☐ Un prochain point d'étape pourra être présenté, lorsqu'un POC sera disponible.

9. Partage d'expériences : société Someware, QDeliveryCheck

Jérôme Thiard, de la société Someware (https://www.someware.fr/) intervient pour présenter la mise au point d'un outil open source de contrôle de données topographiques pour de nombreuses villes bretonnes : QDeliveryCheck (https://someware.gitlab.io/deliverycheck/qdeliverycheck/).
CF. presentation (permalien)

L'outil a été développé pour des collectivités clientes bretonnes et pays de Loire (3 inputs DXF, un input shapefile), autour de la conformité de la structure des données au regard du format/standard PCRS GéoBretagne (https://cms.geobretagne.fr/referentiel-topographique). Il y est adjoint 2 contrôles optionnels permettant la détection d'aberrations planimétriques et altimétriques si l'utilisateur fournit en sus [un fichier de points de contrôle (de SRID défini) ; un MNT] et les écarts maximaux attendus. L'outil (plugin QGIS, avec des modules Python) génère des erreurs et des warnings et permet en leur absence d'exporter au format gml.
Le code source est ouvert sous licence GPLv3 pour la partie plugin (compatibilité avec les exigences d'un plugin QGIS), et avec une licence plus permissive pour les modules ; à l'url https://someware.gitlab.io/deliverycheck/qdeliverycheck/.
Des utilisateurs ayant développé une chaine FME se sont intéressés à migrer sur l'outil.

- ☐ **N.Py** questionne la syntaxe/grammaire de définition de la structure de données (inspirée voire conforme tableschema donc schema.data.gouv par exemple). Réponse négative, chaque contrôle est codé comme une fonction. **N.Py** cite pour complément https://github.com/IGNF/validator.

10. Partage d'expériences : déploiement sur la catalogue PIGMA de l'indicateur qualité

Guillaume BLANCHARD (GIP ATGeRi, **PIGMA**) revient sur le déploiement de son stage, visant à définir des critères qualité généraux pour le catalogue PIGMA afin de mettre en valeur les données vertueuses dans leurs aspects qualité.
CF. presentation (permalien)

La plateforme comprend environ 1500 jeux de données ; même constat (similaire mais antérieur !) qu'Ecolab : la réutilisation est à augmenter, en travaillant la dataviz et la mise en valeur de la qualité
L'idée est de mettre en lumière les acteurs vertueux, sans blâmer ceux dont des progrès seraient souhaitables.
Guillaume rappelle les critères retenus, vérifiables par un automate, qui conduisent à une qualification des données en 3 classes [aucun label ; label minimal (champs indispensables, ressource principale réutilisable, donnée fraiche ou millésimée) ; label optimal (label minimal + informations détaillées concordantes, pas de défaut topologique)].
Quelques champs obligatoires discriminent à eux seuls la majorité des données [pas de label VS label minimal], un travail de reprise des données/métadonnées a été réalisé.

- ☐ Le travail est fortement inféodé à l'architecture PIGMA, bien qu'avec un effort adaptable à toute plateforme.
- ☐ La proposition de labels va être discutée via les discussions régionales, afin de valider les critères, seuils et pondérations. D'ores et déjà par exemple dans L'atelier Valorisation de la donnée des Rencontres régionales PIGMA 2023 ; Télécharger la présentation de l'équipe PIGMA (permalien); Télécharger le compte-rendu détaillé de l'atelier (permalien)

11. Mandat

Suite au premier cycle de relecture critique de la version martyre du mandat **S.Rolle** a rédigé une nouvelle version plus aboutie, discutée en séance et dont les amendements qui en sont issus font l'objet d'un appel à commentaire.

Nous attendons vos éventuels retours d'ici le 16 novembre prochain.

Décision/actions :
Membres QuaDoGéo : relecture du mandat.

12. Points divers

L'**IGN** diffuse maintenant ses rapports de contrôle qualité sur les produits BDOrtho et BDTopo. L'information sera annoncée publiquement via la newsletter d'octobre.
Les rapports sont regroupés sur le site geoservices.ign.fr :

- BDOrtho
  https://geoservices.ign.fr/documentation/donnees/ortho/bdortho/rapport-controle-qualite
- BDTopo
  https://geoservices.ign.fr/documentation/donnees/vecteur/bdtopo/rapport-controle-qualite

S.Launay nous en fera une présentation lors de la prochaine séance, incluant le temps nécessaire à ces tâches. D'ici là, n'hésitez pas à en faire une lecture attentive afin de partager à l'issue de l'intervention de Sébastien vos ressentis, incompréhensions, suggestions, ...

Décision/actions :
Membres QuaDoGéo : prise de connaissance des rapports de contrôle qualité IGN pour partage en séance des étonnements.

Livrables du GT

A ce stade, la production effective ou en cours du GT CNIG QuaDoGéo comprend :

- Les fiches méthodologiques du Cerema : sur le site du Cerema (permalien) ou en accès individuels : Introduction - généralités - contexte du contrôle qualité - éléments statistiques - méthodes d'échantillonnage - mode de représentation - cohérence logique - exhaustivité - précision thématique - précision de position - qualité temporelle ;

- Les webinaires de formation à la qualité par le CRIGE PACA et CEREMA: :
Saison 2021 module 1 : 25 novembre, module 2 : 6 décembre, module 3 : 8 décembre, module 4 : 13 décembre, module 5 : 15 décembre
Saison 2022 lundi 2 mai 2022 généralités et éléments de contexte pour le contrôle qualité ; mercredi 4 mai 2022 statistiques basiques, échantillonnage et critère d'exhaustivité ; vendredi 6 mai 2022 critère cohérence logique et précision de position ; mardi 10 mai 2022 critère qualité temporelle et précision thématique ; vendredi 13 mai 2022 statistiques avancées et méthode de représentation
Thématique « Qualité avec PostGreSQL/PostGIS », jeudi 17 novembre 2022 Session 1 – Modélisation et contrôle a priori ; jeudi 24 novembre 2022 Session 2 – Règles assouplies et contrôle a posteriori ; jeudi 1er décembre : Session 3 – (Géo)analyses avancées
Thématique « Qualité des données avec QGIS », 7 novembre 2023
Thématique « Qualité avec PostGreSQL/PostGIS », saison 2023, 14+21+28 novembre 2023

- Le registre des mesures de qualité des données géographiques et le registre en ligne sur le Géocatalogue ;

- La carte mentale au sujet de la qualification des données suivant la norme ISO 19157 ;

- Deux pistes exploratoires de qualification de données (synopsis, logigramme) sur des thèmes particuliers ;

- La méthodologie de qualification de données *(à consolider…),* comprenant :
- une ébauche de nomenclature des usages génériques de l'information géographique ;
- une maquette de restitution de la qualité des données ;

- Une présentation du GUF, outil de retour utilisateur quant à la qualité des données.

- La Nomenclature des usages génériques de l'information géographique : notice de lecture et tableau de correspondance usage/action.

- *(en cours)* Le prototype d'un outil de restitution graphique de la qualification des données géographiques.

- *(en cours)* Modèle de parties qualité et métadonnées des géostandards

- Le fil de discussion Géorezo : "Qualité des données géographiques"

- Le *Github QuaDoGéo*

- L'espace collaboratif pour l'animation du GT, Google Drive, dont le propriétaire est le compte gtquadogeocnig@gmail.com. Cet espace héberge les documents, les comptes rendus et les travaux finalisés.

Prochaine réunion (visioconférence) le 25 janvier 2024