

 <p>RÉPUBLIQUE FRANÇAISE CNIG <i>Liberté Égalité Fraternité</i> Conseil national de l'information géolocalisée</p> <p>Commission règles & qualité</p> <p>Groupe de Travail Qualité des Données Géographiques (QuaDoGéo)</p>	<p>Groupe de Travail animé par :</p>  <p>Crige Centre de Ressources en Information Géographique Provence-Alpes-Côte d'Azur</p>  <p>IGN INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE</p>
<p>COMPTE-RENDU DE REUNION (distanciel)</p>	
<p>Objet : Réunion #24 du GT QuaDoGéo du 05 mars 2026.</p>	

Ordre du jour :

1. Accueil, Validation du précédent compte-rendu, Date du prochain GT
2. Pratique et processus de contrôle qualité de données géographiques (Erwan BIBENS, HERE Technologies)
3. Assurer la qualité de vos données géographiques avec quels outils ? (Stéphane ROLLE, CRIGE PACA)
4. Le renseignement des standards de données dans les balises de métadonnées (Alexis LEAUTIER, CGDD/SRI/ECOLAB)
5. L'utilisation de la datascience pour l'audit d'une base de données (Nicolas PY, IGN)
6. Point d'actu et questions diverses ...

Liste des participants

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

Nom Prénom	Organisme	Présent	Excusé
ARSAC Jean-Marie	AZIMUT	X	
BAATOUT Ihssen	GIP ATGeRi	X	
BALDE Mamadou-Bailo	Ineris	X	
BIBENS Erwan	HERE Technologies	X	
BOUILLON Aurélie	IGN	X	
BROCHET Laurent	INSEE	X	
CAROLIN Christian	AXES	X	
DEBRAY Laurène	Office international de l'eau / Sandre	X	
DIJOUX Salomé	SANDRE	X	
GALLAIS Arnaud	Cerema Ouest	X	
Le MOAL Mathieu	Axes Conseil	X	
LEAUTIER Alexis	Ecolab	X	
PY Nicolas	IGN Centre Est	X	
RIPOCHE Gabriel	Ecolab	X	
ROLLE Stéphane	CRIGE PACA	X	
ROUX-BARUCQ Victoire	GIP ATGeRi	X	
SEGALA Benoit	Consultant - Cabinet d'études	X	
LAUNAY Sébastien	IGN		X

Réunion #24 du GT QuaDoGéo du 05 mars 2026 (distanciel).		
Mise en relecture du compte-rendu	10/04/2026	Stéphane Rolle / Nicolas Py
Relecture du compte-rendu	13/04/2026	Les intervenants.
Validation du compte-rendu	22/04/2026	Stéphane Rolle / Nicolas Py
Prochain rendez-vous en juin 2026 en distanciel.		

1. Accueil et rappels, Validation du précédent compte-rendu

Aucune remarque sur le compte rendu de la réunion #23, 2025-3, du 7 octobre 2025, qui est donc [validé et disponible](#) sur la page du GT ([permalien](#)).

Le prochain GT est envisagé avant l'été, il pourrait contenir une intervention INSEE sur « la qualité du point de vue de statisticiens et les parallèles avec l'information géographique », sur le COG et sur la BAN.

2. Pratique et processus de contrôle qualité de données géographiques (~14h00 - > 14h40)

[Présentation](#) ([permalien](#)) de Erwan BIBENS (HERE Technologies) sur les pratique et processus de contrôle qualité de données géographiques dans l'entreprise.

L'intervention porte sur ce qu'il est possible de partager sans porter atteinte aux secrets industriel et des affaires.

HERE Technologies ne fait plus uniquement de la cartographie embarquée/automobile (sécurité, conduite autonome, conduite assistée, voiture électrique (recharge, borne rapide, ...)), mais se diversifie pour d'autres industries (logistique, indoor/outdoor, ...).

Les process ont évolué, d'une collecte sur le terrain vers désormais des automatisations, industrialisations consommant des données en Open Data, de partenaires et de traitements IA.

Il reste cependant une collecte via une flotte de véhicules Here (en propre), équipée de capteurs externes (lidar et cam360).

Cette collecte est de plus augmentée des données issues des capteurs des voitures récentes (télémétrie & capteurs ; « vehicule sensor data ») obtenue via certains accords avec certains constructeurs, ou des données issues des traces numériques (smartphones, device GPS). Sont ainsi précisées les données de vitesse et sens de circulation par exemple.

Un aparté licence est effectué. HERE Technologies, pour des questions juridiques et contractuelles, ne peut pas utiliser les données sous licence ODBL (se privant ainsi des données de la ville de Paris). Il y a cependant une bascule qui est en train de s'opérer sur les licences des données ouvertes en France, de l'ODBL vers davantage de licence ouverte LO2 (cf. <https://www.data.gouv.fr/pages/legal/licences> , [permalien](#)).

Des contrôles sont effectués à chaque étape des process ; dont la mise au format pivot Here (identique monde entier) des données disparates récoltées. Certaines de ces vérifications sont automatiques, d'autres manuelles, y compris au besoin par appels téléphoniques.

Cela est complété de tests terrain tous les ans pour vérifier que le niveau de qualité spécifié soit respecté, et le respect des consignes de saisie (ex, terreplein central) et la précision planimétrique.

L'entretien d'une BD est compliqué, c'est un constat récurrent de l'entreprise et un sujet régulier de discussions avec les fournisseurs dont les collectivités.

Cet entretien s'effectue également par une écoute des remontées des utilisateurs (route qui n'existe pas/plus, restrictions de tonnage ou hauteur erronées, ...).

HERE n'a pas encore adapté ses process pour informer le producteur initial (ex. IGN, collectivité, BAN, ...) qu'une incohérence a été détectée.

Here propose une application grand public gratuite : HERE WeGo (https://fr.wikipedia.org/wiki/Here_WeGo).

3. Assurer la qualité de vos données géographiques avec quels outils ? (~14h40 -> 15h10)

[Présentation \(permalien\)](#) de Stéphane ROLLE (CRIGE PACA) sur les conclusions du questionnaire CRIG/QuaDoGéo sur les outils et pratiques de contrôle qualité.

L'intervention porte sur le dépouillement et l'analyse du questionnaire CRIG/QuaDoGéo sur les outils et pratiques de contrôle qualité. L'action a été décidée à la réunion du 13 janvier 2025 (#21, 2025-1), une constitution des question coconstituée avec des membres du GT, un lancement début avril 2025 et un suivi intermédiaire des répondants partagé à la réunion du 13 mai 2025 (#22, 2025-2) avant une clôture en juillet 2025.

Le constat de départ était l'existence de nombreuses données (ouvertes) mais peu normalisées ou vérifiées. L'emploi de l'IA pour cette tâche n'était pas prévue dans le questionnaire, à l'époque l'engouement n'y était pas (du tout).

Le questionnaire a eu 115 répondants, plutôt bien répartis si on ventile par type de structure.

Le support est suffisamment autoporteur pour ne pas le paraphraser.

Dans les quelques résultats saillants, on notera que 52% des répondants qui effectuent du contrôle qualité utilisent FME mais qu'inversement 43% d'entre eux se débrouillent sans ETL, majoritairement via des outils open-source insérés dans des scripts R ou Python. Postgre/gis est plébiscité, et les SGBD portables/fichiers géopackage (OGC) et géodatabase (ESRI) sont plus courant qu'Oracle Spatial ou SQL Server ! Pas d'outils inconnus cités dans le questionnaire.

La cohérence (domaines de valeur, relations, ...) est la dimension qualité la plus contrôlée, de manière assez évidente car c'est la plus automatisable. L'exhaustivité est annoncée plutôt bien contrôlée, mais les réponses sont décevantes car aucun commentaire ni explication des process n'a été partagé.

Enfin, 2/3 des répondants ne connaissent pas l'ISO19157.

Merci aux membres du GT ayant participé !

Un lien vers la page QuaDoGéo et ce CR sera fait là où le sondage avait été annoncé.

4. Le renseignement des standards de données dans les balises de métadonnées (~15h20 -> 16h00)

[Présentation \(permalien\)](#) de Alexis LEAUTIER (CGDD/SRI/ECOLAB) sur le renseignement, dans les balises de métadonnées, du standard auquel adhère la donnée.

Au 04/03/2026, 94% des données publiées data.gouv n'est liée, ne se réfère à aucun schéma/standard.

La présentation, autoporteuse, suggère qu'il soit ajouté une métadonnée permettant de pointer au sein d'un registre des standards le standard que respecte cette donnée, le cas échéant.

Après discussion, les participants conviennent de l'intérêt de la proposition, mais le GT métadonnées (<https://cnig.gouv.fr/gt-metadonnees-a958.html>, [permalien](#)) est davantage légitime à se prononcer et se mobiliser, sans préjuger de l'appui au besoin de membres du GT QuaDoGéo.

Ont été cités dans la présentation les données HVD (<https://guides.data.gouv.fr/autres-ressources-utiles/donnees-de-forte-valeur-modalites-de-rapportage>, [permalien](#)) ainsi que la Fabrique des Standards du CNIG (<http://cnig.gouv.fr/gt-fabrique-des-standards-a26189.html>, [permalien](#))).

Connue de seulement 3 participants, et avec un intérêt explicité en séance, il est convenu que ladite Fabrique soit présentée lors de la prochaine séance du GT ; son fonctionnement, ses livrables, la cohérence entre les différentes usines à standards.

5. L'utilisation de la datascience pour l'audit d'une base de données (~16h00 -> 16h30)

[Présentation](#) ([permalien](#)) de Nicolas PY (IGN) sur l'emploi de la datascience (comptage, exploration graphique, analyses bivariées et multivariées, modélisation) pour détecter de potentielles incohérences dans une base de données géographique.

NPY partage ainsi le travail de 2 années d'alternance de Gabriel BREGAND sous sa supervision et via la formation 'BUT - Science des données' à 'I.U.T LUMIERE Lyon 2'.

La science des données est « La science des données (en anglais : data science) est un domaine interdisciplinaire qui utilise les mathématiques, les statistiques, le calcul scientifique, les méthodes scientifiques, les process, les algorithmes et les systèmes informatiques automatisés pour extraire et extrapoler des connaissances à partir de grandes quantités de données brutes, structurées ou non[1],[2],[3]. Elle est souvent associée aux données massives, à l'analyse des données, aux techniques d'apprentissage automatique comme l'apprentissage automatique (machine learning) et l'apprentissage profond (deep learning). » ([Wikipedia](#)).

Il s'est agi ici d'une analyse de la base de données de production (celle dans laquelle, au quotidien, des modifications sont apportées par les agents de l'IGN pour refléter l'évolution du territoire) de l'IGN, appelée BDUi.

L'analyse a été de complexité croissante, et s'est focalisée sur des contrôles de cohérence réalisables au bureau car ne nécessitant pas l'acquisition d'une donnée de référence.

Il s'est d'abord agi de vérifier la conformité de la base aux règles métier enregistrées dans le méta-modèle. De cette vérification, il a par exemple pu être proposé de toiletter les listes de valeurs autorisées pour un attribut (qui pouvait contenir des valeurs nulles ou des valeurs en doublon), ou bien de rendre effectif la vérification du respect d'expression régulières, prévues mais ne faisant à l'époque pas l'objet de scripts d'audit.

Gabriel a ensuite effectué des analyses bivariées, c'est-à-dire recherché des couples d'attributs corrélés pour y détecter des incohérences, par comptages et analyses graphiques. Un exemple d'écart aux spécifications de saisie concerne les tronçons de route, où le niveau d'importance du tronçon est par exemple dépendant de la capacité d'un véhicule léger à le parcourir.

Plus complexe, des analyses multivariées. Cas particulier, plusieurs attributs corrélés dont l'exemple le plus parlant est constitué des informations d'altitude des bâtiments (altitude min et max du sol, min et max du toit).

En dehors de ces analyses multivariées ad-hoc, le travail a enfin porté sur la fabrication d'une multitude de modèles statistiques. L'idée est de prédire un attribut à l'aide des autres attributs de la table, et de suspecter une incohérence lorsque la valeur prédite et la valeur en base diffèrent.

La technique du Random Forest a été employée, avec comme spécificité au regard de ses emplois habituels de générer un ensemble d'attributs descripteurs de la géométrie (en effet, un graphe autoroutier est plutôt composé de tronçons longs et rectilignes, avec peu d'angle entre eux, là où à l'inverse une route de montagne sera composée de tronçons plus courts et avec de nombreux virages).

Plusieurs analyses sont décevantes (la BDUni est une base de donnée d'excellente qualité, fortement contrôlée ; il s'agit donc de trouver une aiguille dans une meule) mais quelques-unes prometteuses, par exemple la nature des tronçons de routes où la méthode pourrait proposer d'examiner 40.000 objets suspects avec vraisemblablement 13.500 corrections à effectuer. Cela pourrait permettre d'identifier environ 13 500 corrections sur l'attribut nature. La majorité des incohérences ainsi corrigées concerneraient des objets de nature routes à chaussée (1 ou 2 chaussées), qui pourraient être de nature : type autoroutier, bretelle, ou rond-point.

6. Point divers et conclusion

M.LE MOAL partage une synthèse d'expériences d'analyses qualité (notamment sur des données descriptives de réseaux humides) dans le document [« Qualité des données : évaluer ou juger ? » \(permalien\)](#).

GT CNIG QuaDoGéo



Prochaine réunion (présentiel) en juin 2026.