



**GROUPE DE TRAVAIL « OPENDATA » N°3
COMPTE RENDU**

DATE DE REUNION
Lieu de réunion

22/11/2022
Bordeaux

Affaire suivie par : Sébastien DIAS, Héroïse DESCHAMPS et Anne SAGOT-DUVAUROUX

PRESENTS :

GIP ATGeRI	Héroïse DESCHAMPS Sébastien DIAS	Anne SAGOT-DUVAUROUX Emeric PROUTEAU
ALPI	Didier ROBINO	
ATD 16	Romain LACOURARIE	
ATD 24	Cécile SOULINGEAS	
BORDEAUX METROPOLE	Gabriel DOS SANTOS	
CA BRIVE	Mathieu ANGLARD	
CA PAYS-BASQUE	Simon ESTINES	
CC DE L'ILE DE RE	Florence DELAMARCHE	
CC MACS	Marie DELHORBE	
CD 40	Amandine SOUMAILLE	
CEREMA SO	Stephane DOMINGO	Anne-Laure JAUMOILLIE
COBAS	Christelle JAZE	
CRIGE PACA	Stéphane ROLLE	
CU GRAND POITIERS	Sylvie LE MAT	
DREAL NOUVELLE-AQUITAINE	Pascal PREVOT	
EPTB VIENNE	Anne-Charlotte JEAN	
IGN	Laurent MONGIE	
PETR MARENNES-OLERON	Stephane ROGER	
PORT DE BORDEAUX	Fabrice KLEIN	
REGION NOUVELLE-AQUITAINE	Dominique LOUIS	Charlène CASSANTE
SCOT DE LA HAUTE GIRONDE	Tiphaine MAURIN	
BLAYE - ESTUAIRE		
SDIS 64	Adrien CARPENTIER	
SIAEBVELG	Rémi MALRAISON	
SIEPAL	Chloë LEGRAND	Anne-Sophie PIERRE
UNIVERSITE DE POITIERS	Saida RAVE	
VILLE DE CENON	David DELHORBE	

1. Ordre du jour

- Introduction L'Open Data en France et en Région Nouvelle-Aquitaine
- Présentation des travaux sur la qualification des données du groupe de travail du CNIG QUADOGEO par Stéphane ROLLE (CRIGE PACA)
- Proposition d'actions de valorisation des données dans le catalogue PIGMA
 - Mise en place d'un indicateur de qualité
 - Intégration d'outils de dataviz

2. Compte rendu

2.1. Introduction

Le GIP ATGeRi remercie les participants à cette troisième session du groupe de travail PIGMA sur l'Open Data.

Un tour de table est effectué pour que chacun des participants détaille la structure représentée et sa fonction.

2.2. Introduction L'Open Data en France et en Région Nouvelle-Aquitaine

2.2.1. L'Open Data en France

Le GIP ATGeRi présente les chiffres de l'[Observatoire Open Data des territoires](#), porté par [Open Data France](#).

Les chiffres présentés sont issus de la [page « Résultat »](#) de l'observatoire des territoires.

En France, 897 organisation publient au moins un jeu de donnée en open data, dont 51,1% sont des communes.

2.2.2. L'Open data en Nouvelle-Aquitaine

Le GIP ATGeRi a repris [les données brutes de l'observatoire](#) pour les adapter au périmètre de la Nouvelle-Aquitaine.

La réalisation de cette vue permet de montrer que la Région Nouvelle-Aquitaine suit la même tendance qu'au niveau national, avec une part majoritaire de communes (60,2%) mais une part plus importante de groupement de collectivités territoriales (21,2%)

En Nouvelle-Aquitaine, 118 organisations publient au moins un jeu de donnée en open data.

La Nouvelle-Aquitaine est également la région ayant le plus d'organisations qui publient au moins un jeu de donnée en open data.

Le [catalogue PIGMA](#) (Plateforme d'Echange de données en Nouvelle-Aquitaine) quant à lui recense :

- 28 organismes ayant des jeux de données open data
- 1189 jeux de données open data

2.2.3. Objectifs de PIGMA

Les précédents chiffres montrent que la donnée Open Data existe et qu'un nombre relativement important de données est diffusé. Afin d'aller plus loin tout en maintenant l'objectif de continuer les

publications de données, le GIP ATGeRi rappelle son objectif de travail : valoriser la donnée pour une meilleure réutilisation en mettant en place une qualification de cette dernière.

2.3. Présentation des travaux sur la qualification des données du groupe de travail du CNIG QUADOGEO par Stéphane ROLLE (CRIGE PACA)

2.3.1. Intervention du CRIGE PACA

Stéphane ROLLE, chef de projet SIG au Centre de Ressources en Information Géographique (CRIGE) PACA, présente sa structure et l'équipe avant de faire un rappel sur les dates clés autour de la qualité.

Depuis la parution de la première revue sur la qualité en 1997 rien ne s'est imposé dans la qualité de la donnée.

En 2018 a eu lieu un séminaire sur la qualité des données au CEREMA de Lyon. Ce séminaire a lancé la création d'un GT national [QuaDoGéo](#) (Qualité Données Géographiques) porté par le CNIG l'année suivante.

Depuis 2019, le GT QuaDoGéo, co-organisé entre le CRIGE PACA et le CEREMA, a pour rôle de porter au niveau national la dynamique mise en place localement le 8 février 2018 au séminaire qualité à Lyon.

En 2022, le GT a organisé 4 rencontres avec un axe de veille avec des experts, proposition d'accompagnement technique et assurer un support dans les régions.

Le CRIGE PACA intervient en tant que relai technique avec le niveau national et local pour adapter les méthodes.

Le GT QuaDoGéo est basé sur 3 axes principaux :

- Accompagnement des diffuseurs de données,
- Accompagnement des utilisateurs des données (réutilisation)
- Prise en compte du retour utilisateur

Stéphane ROLLE présente brièvement les livrables actuels du GT dont les [fiches méthodologiques](#) produites en 2018 par le CEREMA et qui décryptent la norme ISO 19157 (2013) ainsi que des [webinaires de formation](#) « qualité des données géographiques » (2021/ 2022) qui sont gratuits, ouverts à tous et expliquent ces fiches.

Il relève que la plupart des personnes font de la qualité sans le savoir et que ces formations, en collaboration avec le CEREMA, ont vocation à trouver un point commun entre les démarches de tout le monde.

En 2022, le CRIGE PACA a également développé un script FME de traitement générique basé sur certains indicateurs de qualité de la norme ISO 19157 avec édition de rapports automatiques d'analyse d'erreurs. Ces travaux ont été réalisés dans le cadre d'un stage de 4 mois.

Ce script sera consolidé et mis en production en 2023.

Le CRIGE, en lien avec le CEREMA et OPENIG, propose également des formations orientées qualité sur postgres/ postgis (2022/2023).

Stéphane ROLLE fait part du constat suivant : avec l'arrivée de l'open data se pose la question de la qualité et de la gouvernance des données.

Il soulève différents enjeux :

- Interopérabilité
- Confiance dans les données (et producteurs)

- Exploitation opérationnelle des jeux de données
- Critères d'exigence communs
- Compréhension des résultats

Il note également que la qualité doit être traitée dès le départ sinon cela peut impacter les prises de décision, avec une amplification d'erreurs.

Pour éviter cela, il y a 2 aspects de la qualité à identifier :

- Interne : s'appuyer sur la conformité, la standardisation, les spécifications de la donnée, les géostandards (ex validata, validateur GPU, IBAN)
- Externe : trouver l'adéquation entre la donnée et le besoin de l'utilisateur, soit définir « l'utilisabilité » de la donnée.

Le producteur indique la qualité de son jeu de donnée, applique un traitement générique pour attester la qualité au regard de la norme

Stéphane ROLLE finit son intervention en faisant un focus sur les fiches CEREMA qui décryptent la norme européenne ISO 19157 adoptée le 09/11/13 et évalue la qualité des données géographiques. Cette norme agrège 3 normes ISO : 19113 (2001), 19114 (2005) et 19138 (2006).

Ce sont 10 fiches sur les critères de qualité et méthodes et qui posent la question du rendu de ce que l'on veut représenter.

2.3.2. Echanges

[PORT DE BORDEAUX] Le système de notation est-il envisagé à l'issue des travaux de qualification des données ? Comment arriver à passer d'un indicateur à un aspect graphique simple et évocateur ?

CRIGE PACA : il n'y a pas beaucoup de solutions (smiley, note, étoile). L'étoile est dérivée d'une note (5 niveaux max), l'aspect de la couleur c'est toujours la même chose qui revient et qui parle mais qui est plus difficile à appliquer, ça pose la question de l'homogénéisation sur l'ensemble des indicateurs GIP ATGeRi : la notation est une réflexion parallèle à la qualification des données

[SCOT DE LA HAUTE GIRONDE BLAYE – ESTUAIRE] L'objectif des fiches qualité est à appliquer pour les données à publier, est-il prévu de faire un retour qualité sur les données déjà publiées en open data ?

GIP ATGeRi : l'objectif concerne les données à publier et certaines déjà publiées dans le catalogue PIGMA dont celles retenues comme données testées dans le cadre de ce GT.

2.4. Propositions d'actions de valorisation des données dans le catalogue PIGMA

2.4.1. Mise en place d'un indicateur de qualité

En introduction, le GIP ATGeRi rappelle l'objectif du GT Open Data 3 : définir des règles de qualité pour une meilleure réutilisation et diffusion de la donnée.

Aussi, il informe les participants de la nécessité que l'indicateur de qualité calculé par PIGMA puisse faire l'objet de traitement automatique. Ce sont donc des critères communs à toutes les données, appréciables des non thématiques de la donnée et présents dans le modèle de jeu de donnée PIGMA qui ont été retenus dans le cadre de ses premiers travaux.

L'indicateur de qualité concerne ici à la fois la métadonnée (fiche d'information sur la donnée) et la donnée elle-même.

Une méthode commune a été définie et testée sur 5 premiers jeux de données : Plages Nouvelle-Aquitaine, Etablissements de santé et médico-sociaux, Zonages des Documents d'urbanisme et l'Occupation du Sol en Nouvelle-Aquitaine.

Ces 5 données sont toutes régionales, certaines sont non standardisées, d'autres respectent un standard (local ou national) et sont accessibles en licence ouverte 2.0 ou ODbL.

Indicateur qualité métadonnée

Le GIP ATGeRi propose de mobiliser cet indicateur qui s'appuie sur le score qualité mis en place par datagouv.

C'est un score en phase d'expérimentation, le poids du critère sera ajusté en fonction des retours réutilisateurs/ producteurs et de nouveaux critères pourront être ajoutés progressivement.

Au 07/11/22, datagouv propose [6 critères](#) : description des données, mise à jour, licence, métadonnées des ressources, couverture spatiale et couverture temporelle.

Pour la Nouvelle-Aquitaine, le GT PIGMA OPEN DATA décide de retenir les 5 critères suivants (hors couverture temporelle) qui sont présents dans le modèle de jeu de donnée PIGMA :

- Description des données
- Mise à jour
- Licence
- Métadonnées de ressources
- Couverture spatiale (granularité spatiale)

Les 5 métadonnées testées disposent bien des éléments d'information faisant l'objet du score qualité.

Indicateur qualité donnée

Le GIP ATGeRi propose de mobiliser cet indicateur qui s'appuie sur les fiches méthodologiques produites par le CEREMA.

La norme ISO 19157 (2013) qui mesure la qualité des données, est composée de [5 critères](#), certains ayant des sous-critères : cohérence logique, exhaustivité, précision thématique, précision de la position et qualité temporelle.

Pour la Nouvelle-Aquitaine, le GT PIGMA OPEN DATA décide de retenir 3 critères et d'écarter certains sous-critères en proposant de vérifier les éléments suivants :

- cohérence logique
 - o cohérence conceptuelle : unicité identifiant unique, cohérence code INSEE/ zone géographique, respect géométrie de la table
 - o cohérence du format : respect format fichier, encodage, projection, nom champ, format attribut code géographique et date
 - o cohérence topologique : nombre de superposition, auto-intersection et auto-chevauchement
- précision de la position : le GIP ATGeRi propose d'écarter les sous-critères et de vérifier à la place la mention d'une échelle d'utilisation
- qualité temporelle : exactitude mesure temporelle, cohérence et validité temporelle

Les éléments de vérification proposés par le GIP ATGeRi peuvent être vérifiés en grande majorité de manière automatique (via script).

Le GIP ATGeRi rappelle les 2 cas de données à qualifier :

- Données avec standard (national, local) : qui ont des spécifications propres à vérifier, dont certaines peuvent recouper avec les vérifications proposées par le GIP ATGeRi.
- Données sans standard : vérifications du GIP ATGeRi, qui s'appuient sur la recherche d'informations aberrantes/ anomalies, à appliquer en fonction des conditions remplies.

Indicateurs Qualité donnée

Les résultats de mesure de la qualité des métadonnées et des données testées servent de base pour le calcul de la note finale.

Le type de notation reste à définir (choix, rendu).

La notation finale

La notation peut être de 3 types : complexe (mesures brutes), simplifiée (toile d'araignée/ schéma radar) ou schématique (une seule information).

La note sous forme schématique est ici présentée et ne doit pas dépasser 5 niveaux. Elle serait représentée sous forme de note, d'étoile ou smiley.

De son côté, le GT PIGMA OPEN DATA acte un système label dont l'intitulé reste à définir, qui serait appliqué aux jeux de données à partir d'une certaine note. Ce label permettrait de valoriser certains jeux de données dont la qualification est notable et faire améliorer la qualité d'autres jeux de données.

Enfin, l'équipe PIGMA se tient à disposition pour toute question sur les travaux de qualification réalisés et peut approfondir la présentation si besoin.

2.4.2. Mise en place d'outils de datavisualisation

Le GIP ATGeRi présente les grands principes de la datavisualisation et quelques exemples associés. En lien avec la précédente partie, le GIP ATGeRi rappelle le rôle central de la qualité des données pour aboutir à des datavisualisations claires et cohérentes.

Le GIP ATGeRi propose pour 2023 d'intégrer un premier niveau de service de datavisualisation dans la plateforme PIGMA.

La proposition de socle minimal à intégrer concernerait l'intégration d'un outil de datavisualisation dans le catalogue PIGMA (ex : nouvel onglet « Graphique » ou « Datavisualisation »).

Le GT PIGMA OPEN DATA décide que les datavisualisations soient réalisées par les administrateurs de la plateforme et les éditeurs des organisations. Il ne s'agirait donc pas d'un outil de création de datavisualisation à la volée. Pour justifier cette décision, le GIP ATGeRi met en avant que ces outils de création de graphique à la volée sont souvent difficilement utilisables et que la connaissance des données est souvent un préalable important pour extraire des informations pertinentes.

Le GIP ATGeRi propose de co-construire ce premier niveau de service au sein d'un sous-groupe de travail datavisualisation. Ce sous-groupe aura pour objectif de définir les spécifications de développement pour l'intégration de cet outil datavisualisation dans la plateforme PIGMA.

2.4.3. Echanges

Mise en place d'un indicateur de qualité

[ALPI] Comment on s'intercale avec validata ou data.gouv ?

GIP ATGeRi : on est sur des tests de standards, PIGMA vise un standard de qualification générique qui peut venir en complément de ce qui est déjà fait au niveau national.

[VILLE DE CENON] Cela signifierait de mettre une barrière à la publication dans le catalogue PIGMA ?

GIP ATGeRi : non, mais certaines données seraient plus valorisées que d'autres. Il faut faire monter en qualité tout le monde. Tout le monde n'a pas le même niveau financier et technique. C'est pour cela que le système label serait le plus adapté.

Il faut avoir des critères objectifs dans les relances de mise à jour.

PIGMA ne fait qu'héberger les données et est tributaire de la mise à jour, qualité de publication par les producteurs mais se doit d'accompagner ses partenaires pour améliorer cette qualité.

[SCOT DE LA HAUTE GIRONDE BLAYE – ESTUAIRE] *Attention au fait de ne pas prendre en compte le critère localisation qui est un problème très souvent rencontré à l'échelle de la collectivité. Il faudrait bien spécifier ce qui est pris en compte, préciser les règles vérifiées par PIGMA si un label est appliqué pour ne pas induire en erreur ?*

Pour le cas des données erronées retravaillées, doit-on faire un retour directement au producteur ?

GIP ATGeRi : c'est pour cela qu'il est important de travailler sur la pondération des critères. La localisation est un critère difficile à qualifier.

Le retour au producteur supposerait un système de retour type chat pour indiquer que la donnée a été retravaillée par un réutilisateur et des développements à l'outil actuel.

[DREAL NOUVELLE-AQUITAINE] *Il faudrait peut-être mettre un commentaire de mise en garde de l'utilisateur sur la qualité/ défaut de la donnée ?*

GIP ATGeRi : l'analyse commencera en priorité sur les données niveau 1 (régional) et open data (partenaires), niveau départemental si besoin.

[PORT DE BORDEAUX] *réflexion engagée par PIGMA assez cohérente, pose la question de la valorisation la plus forte possible entre ces différentes données, qui à son sens, doit passer par un croisement d'informations.*

[VILLE DE CENON] *Il y a des projets de standardisation des données de niveau 2 et 3 pour homogénéiser ?*

GIP ATGeRi : non, tout ne pourra pas être traité mais cela peut entraîner certaines dynamiques.

[REGION NOUVELLE-AQUITAINE] *PIGMA apporte une expertise technique des données, qu'en est-il de l'expertise des usagers ? Il serait intéressant de dédier un endroit pour remonter les avis des usagers ? La prise en compte des avis des usagers peut permettre de connaître l'usage de la donnée. Pose également la question de la réutilisation.*

GIP ATGeRi : le nombre de réutilisations n'est pas forcément un indicateur parlant, il y a plus des retours négatifs que positifs. Une réflexion est en cours sur les données à proposer.

Techniquement ça peut se mettre en place et ça se réfléchit dans le cadre du programme PIGMA. Aujourd'hui, PIGMA est en capacité de valoriser la donnée par la qualité pour favoriser son usage.

[PORT DE BORDEAUX] *Les données sont liées à des licences (un usage est défini) et concédées à une entité. Dans le domaine maritime on cherche d'où viennent les données utilisées entre bureaux d'études souvent brutes. Lors d'une commande, pourquoi pas demander de fournir la licence des données utilisées en plus des études, pour identifier la source et avoir une étude plus qualitative. Il faudrait pouvoir tracer les données utilisées.*

[SIEPAL] *Dans les données priorité 1, est-ce que vous avez la possibilité d'accéder aux statistiques de téléchargement ?*

GIP ATGeRi : pour le cas de l'occupation du sol, après discussion avec la Région et la DREAL notamment, un formulaire avant le téléchargement avait été mis en place pour identifier la réutilisation

Mise en place d'outils de dataviz

[CU GRAND POITIERS] *On a fait un datasandwich sur la dataviz, le support est aussi disponible sur le site data.grandpoitiers.fr à vous de jouer. Les dataviz préconstituées sont préférées.*

<https://data.grandpoitiers.fr/pages/datasandwich29042021/>

<https://data.grandpoitiers.fr/pages/evenements/>

[ALPI] *Quel outil dataviz ?*

GIP ATGeRi : c'est l'objet du sous-groupe datavisualisation

[PORT DE BORDEAUX] *même questionnement : quels sont les outils de visualisation ?*

Dans le cadre d'un projet de mesures de salinisation de l'estuaire, le Port de Bordeaux a opté pour un outil simple d'utilisation avec une liste de couches et éditeur simple (type qgis) pour les

décideurs et pour le monde scientifique un autre outil plus complexe (notebookjupyter) à utiliser avec du code python pour générer des graphiques et tableaux qui nécessitent des compétences. On valorise une donnée et pas des gros croisements de données.

[REGION NOUVELLE-AQUITAINE] intéressée de savoir si la dataviz couvre tout le territoire Nouvelle-Aquitaine ? Quel est l'objectif de la plateforme ? Quel usage est fait de la dataviz ?
GIP ATGeRi : le but est de mettre en valeur la donnée et de proposer un outil ayant un socle de fonctionnalités minimal.

3. Relevé de décisions

Le GIP ATGeRi propose :

- l'organisation de la **prochaine réunion du GT courant juin 2023** avec la validation d'un indicateur qualité (notation et affichage) et des données analysées (en priorité régionales et open data)
- le recrutement d'un stagiaire pour poursuivre les travaux déjà engagés
- la mise en place d'un **sous-groupe datavisualisation au 1^{er} semestre 2023** avec les partenaires intéressés d'y participer (pour le moment : SCOT HAUTE GIRONDE, REGION NOUVELLE-AQUITAINE, ATD 24, UNIVERSITE DE POITIERS). Ce sous-groupe permettra entre autre de définir plus précisément les objectifs.