



**GROUPE DE TRAVAIL « OPENDATA » N°4
VALORISATION DE LA DONNÉE
COMPTE RENDU**

DATE DE REUNION
Lieu de réunion

11/04/2024
Bordeaux

Affaire suivie par : Guillaume BLANCHARD, Héroïse DESCHAMPS et Ihssen BAATOUT

PRESENTS :

GIP ATGERI

Héroïse DESCHAMPS
Guillaume BLANCHARD

Ihssen BAATOUT
Paul VILVANDRÉ

AREC NA Camille ROMARY

ATD 24 Émeric PROUTEAU

BRGM Bruce AYACHE

CA BRIVE Mathieu ANGLARD

CC MACS Michael COSTY

CD 24 Arnaud LÉBOUCHER

CD 64 François ORDOQUI

CU GRAND POITIERS Sylvie LE MAT

DDT 19 Marie-Laure FRANCH

DDTM 33 Éric BALLAN

PETR GRAND LIBOURNAIS Antoine CASTAGNET

SCOT DE LA HAUTE GIRONDE Philippe CIANFARANI

BLAYE - ESTUAIRE

SIEPAL Martin JOUY

VILLE DE CENON David DELHORBE

Pierre GOMIS

1. Ordre du jour

- Etat des lieux des travaux sur la qualité des données et perspectives
- Conclusions et mises en œuvre des travaux 2023 sur les outils de datavisualisation

2. Compte rendu

2.1. Introduction

Le GIP ATGeRi remercie les participants à cette quatrième session du groupe de travail PIGMA sur l'Open Data. En Région Nouvelle-Aquitaine et plus généralement en France, la publication de jeux de données en Open Data est bien développée et s'accroît chaque année. L'enjeu aujourd'hui pour la plateforme PIGMA concerne davantage la valorisation des jeux de données sur la plateforme. Cette valorisation de la donnée se travaille sur deux axes :

- La qualité des données et des jeux de données pour favoriser leur réutilisation
- La datavisualisation pour favoriser la découvrabilité des jeux de données

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

2.2. État des lieux des travaux sur la qualité des données et perspectives

2.2.1. Rappel de la feuille de route 2023

Lors du précédent GT Open Data 3 du 22/11/2022, le GIP ATGeRi a décidé de l'organisation d'un stage en 2023, dédié à la définition d'indicateurs et de critères de qualité, à partir d'un nombre restreint de jeux de données prioritaires, dans une démarche de valorisation de la qualité.

Ce stage, effectué par Guillaume BLANCHARD, s'est déroulé d'avril à août 2023.

2.2.2. Les travaux du stage QuaDoGéo

2.2.2.1. Objectifs et constats

Les objectifs du stage étaient la représentation sous forme de Label Qualité, l'établissement de critères de qualité et l'automatisation du contrôle qualité pour un grand nombre de jeux de données simultanément, le tout ayant pour but de valoriser la qualité des jeux de données de la plateforme.

Lors du stage, 3 constats ont pu être faits :

- Les métadonnées des jeux de données sont **décrites** selon les mêmes champs dans PIGMA
- Dans son expérience, l'utilisateur est **d'abord** confronté à la qualité des métadonnées
- Une donnée ancienne ou inaccessible ne peut pas être aisément **réutilisée**

2.2.2.2. Label Qualité à 2 niveaux

Comme l'utilisateur est d'abord confronté à la qualité des métadonnées, la métadonnée doit être correctement renseignée pour fournir à l'utilisateur toutes les informations lui permettant d'estimer la pertinence du jeu de données au regard de ses besoins et, in fine, de constater la qualité de donnée.

Ainsi, les travaux de qualification ont permis l'établissement d'un Label Qualité à deux niveaux :

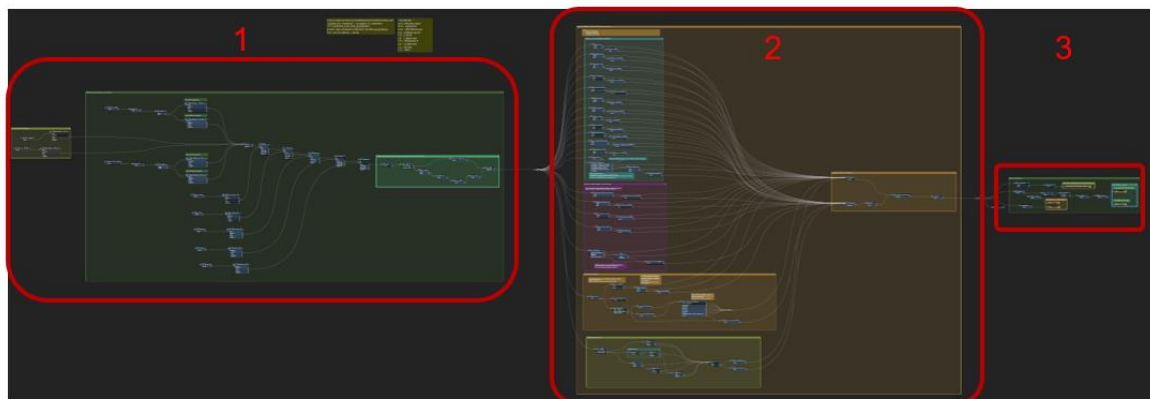
- Le **Label Minimal**, centré sur la qualité des métadonnées et qui analyse :
 - Le renseignement de **champs indispensables** (nom, description, mots-clés, licence, dates de publication et de mise à jour du dataset, fréquence de mise à jour, type de données, contact, thématiques, granularité)
 - La présence d'une **ressource principale réutilisable** avec nom, date de mise à jour, format (GeoJSON, Shapefile, flux OGC, CSV, JP2, etc.)
 - L'obligation de disposer d'une **donnée fraîche ou millésimée**, avec une borne temporelle à 2 ans (compromis entre la fraîcheur de la donnée et la contrainte pour l'éditeur d'une mise à jour trop régulière)
- Le **Label Optimal**, qui analyse une qualité avancée des métadonnées en intégrant un premier niveau d'analyse de la qualité des données et qui :
 - Reprend tous les critères du Label Minimal
 - Analyse le détail et la concordance des informations de métadonnées
 - L'absence de défauts topologiques et d'erreurs majeures dans la donnée

Pour le Label Minimal, **l'absence d'un seul champ** indispensable ou l'échec à une des analyses empêche le jeu de données de recevoir Label Minimal.

Pour le Label Optimal, l'attribution du Label se fait par **pondération des différents critères** analysés, hormis quelques critères devant être obligatoirement remplis (absence de doublon dans la donnée, absence d'erreurs topologiques, etc.)

Les travaux d'automatisation se sont penchés sur les deux niveaux de Label. La présentation des résultats lors du présent GT concerne uniquement le Label Minimal.

2.2.2.3. Script de qualification et livrables produits



Le script de qualification du Label Minimal mis en place via le logiciel FME se décompose en 3 grandes parties :

1. La récupération des informations de métadonnées depuis la base de données (propre à PIGMA)
2. Les différentes analyses du Label Minimal
3. La production d'exports
 - a. Les CSV pour l'analyse du Label Optimal
 - b. Les rapports HTML pour les travaux de montée en qualité

Les éléments, champs et valeurs analysés dans ce script sont très **dépendants du modèle de données** de la plateforme. Une adaptation du script au modèle de données est nécessaire pour les organismes qui souhaiteraient exécuter le script en interne pour des données non-référencées dans PIGMA.

En parallèle des scripts, plusieurs documents ont été produits :

- [Un document rassemblant les critères de qualité des métadonnées](#) retenus pour les deux niveaux de Label
- [Un document rassemblant les critères de qualité de la donnée](#) retenus pour le Label Optimal
- [Un document de correctifs à apporter aux jeux de données](#), en fonction des erreurs remontées dans les [exports HTML](#), guidant l'utilisateur dans sa démarche de montée en qualité, pas à pas

2.2.2.4. Résultats avant et après montée en qualité

Lors du stage, sur 530 jeux de données analysés, seuls 79 jeux de données (soit 15% du total) recevaient le Label Minimal lors des premières exécutions du script. Si les champs indispensables étaient globalement bien renseignés, **certaines champs avaient une grande importance dans l'obtention** ou non du Label Minimal comme l'obligation de millésime ou la présence d'un format réutilisable.

Des **travaux de montée en qualité** ont été effectués à partir des documents produits, ainsi que des modifications dans le fonctionnement du script pour mieux prendre en compte certains comportements. À l'issue de ces travaux, 222 jeux de données (soit 42% du total) ont reçu le Label Minimal, permettant de valider **la faisabilité d'une montée en qualité** des jeux de données.

2.2.3. Expérimentation de la qualification avec des partenaires PIGMA

Pour poursuivre les travaux du stage, le GIP ATGeRi a souhaité **expérimenter le processus de qualification avec des partenaires PIGMA** volontaires pour s'inscrire dans cette démarche de montée en qualité et de valorisation des jeux de données. Parmi les partenaires sollicités, les travaux de qualification ont pu être menés **avec la CC MACS et le BRGM**.

2.2.3.1. Méthodologie de travail

Lors de ces travaux, le GIP ATGeRi a pu mettre en place la méthodologie de montée en qualité suivante :



Après une réponse favorable de la structure, le script de qualification est exécuté et un **premier rapport est produit**. Une **réunion de concertation** entre l'équipe PIGMA et la structure est organisée pour analyser les résultats et déterminer les axes de travail de la montée en qualité (fiches à reprendre, jeux de données obsolètes à supprimer, modification à apporter au script). La structure corrige ensuite les erreurs et une nouvelle analyse s'effectue, avec envoi des rapports, cette étape pouvant se répéter plusieurs fois.

En fin de processus, les jeux de données de la structure obtiennent le Label Minimal. **L'affichage du Label Qualité** pour les jeux de données qualifiés de la plateforme **est actuellement en cours de développement**.

2.2.3.2. Retours d'expérience des partenaires

Michael COSTY de la CC MACS explique que la CC disposait de plus d'une centaine de fiches dans le catalogue PIGMA, pour la plupart privées et dont seules quatre étaient publiées en Open Data. La communauté de communes s'est lancée depuis 2019 dans une démarche Open Data, longue à mettre en place avec les services de l'intercommunalité. En 2024, cette **démarche de valorisation** est relancée, avec un long **travail d'accompagnement et d'acculturation** à la qualité de la donnée à destination des services de l'intercommunalité.

À la suite de la réunion de concertation, une suppression des fiches privées a été effectuée, suivi d'un nettoyage des 4 fiches Open Data, notamment sur les fréquences de mise à jour renseignées. Les 4 fiches reçoivent toutes le Label Minimal à l'issue du processus. La CC MACS souligne que l'outil mis à disposition par PIGMA est **simple à prendre en main**.

Bruce AYACHE du BRGM rappelle le contexte du catalogage pour le BRGM qui dispose de plusieurs catalogues moissonnés par PIGMA. Un grand nombre de fiches ont été intégrées dans le nouveau catalogue PIGMA depuis 2022, ce qui concerne quelques 400 fiches. Ceci s'inscrit dans le positionnement du BRGM en tant que centre de ressources pour la thématique littorale pour partager au mieux ces données.

Bruc AYACHE explique que le travail de gestion des métadonnées est de longue haleine et peut toujours être amélioré et précise que **l'outil mis en place par PIGMA permet de corriger les problèmes de qualité** des métadonnées.

L'expérimentation s'est effectuée sur une vingtaine de fiches prioritaires, celles en open data avec des ressources téléchargeables. La qualification a fait remonter un manque de renseignements pour certains champs de métadonnées ainsi que la présence de **ressources principales au format ZIP**, format systématiquement utilisé pour échanger des données Shapefile.

La montée en qualité et l'affinage du processus de qualification ont permis de passer de 4 à 16 fiches recevant le Label Minimal, sur un total de 18 fiches.

Le GIP ATGeRi cherche désormais d'autres structures souhaitant s'inscrire dans cette démarche de qualification et de montée en qualité de leurs jeux de données. La DDTM 33 et la ville de Cenon, insérés chacun dans une démarche de qualité depuis plusieurs années, se sont manifestées pour participer à l'expérimentation.

2.2.4. Echanges

[CD 64] Parfois, les données au CD 64 sont mises à jour automatiquement par des scripts mais en réalité la donnée est la même que précédemment et n'est pas plus fraîche. Quelle est la réflexion de PIGMA à ce sujet ?

GIP ATGeRi : Les jeux de données dont la donnée est mise à jour automatiquement par des scripts sont minoritaires sur la plateforme PIGMA. Même la question de la fraîcheur réelle des données est pertinente, le premier travail effectué sur PIGMA concerne d'abord la mise à jour des données et des jeux de données, dont plusieurs sont anciennes. En tant que plateforme généraliste et multithématiques, PIGMA peut assurer que le fichier contenant la donnée a été mis à jour à une certaine date mais ne peut pas garantir plus que cette information. La garantie de disposer de la donnée la plus à jour possible est du ressort du producteur.

[BRGM] Le résultat du script sous forme de rapports HTML peut-il intégrer le nombre de vues et de téléchargement des jeux de données ? Cela permettrait de montrer et de prendre la réutilisation des jeux de données aux services des collectivités et de prioriser le travail de montée en qualité.

GIP ATGeRi : Ces informations sont stockées dans les bases de données et peuvent être extraites si besoin. Les informations de consultation et de téléchargement des jeux de données sont plus pensées pour être affichées à terme dans les fiches du catalogue. Cependant, elles peuvent être pertinentes pour la montée en qualité et doivent pouvoir intégrer les rapports à l'avenir.

2.2.5. Conclusion et perspectives

La feuille de route 2024 pour la qualité des données présente différentes actions à mener par le GIP ATGeRi :

- Nettoyage des fiches propres au GIP ATGeRi
- Accompagnement des partenaires volontaires dans leur démarche de qualification
- Analyse de la qualité pour les fiches indexées

Le processus ne fonctionne actuellement que pour les **jeux de données intégrés** dans la plateforme. **L'objectif** est de pouvoir exécuter le processus pour les jeux de données **indexés** depuis d'autres catalogues, afin de qualifier l'ensemble des jeux de données de la plateforme.

Des développements sont en cours pour afficher le Label Minimal dans le catalogue. Une fois le système de qualification du Label Minimal appliqué à tous les jeux de données, celui du Label Optimal pourra être mis en œuvre.

Pour participer à la valorisation des jeux de données de la plateforme, un système de **notifications de mise à jour et d'abonnement** à des jeux de données sera déployée sur la plateforme en 2024.

2.3. Conclusions et mise en œuvre des travaux 2023 sur les outils de datavisualisation

2.3.1. Rappels des besoins de l'outil et de la datavisualisation dans PIGMA

En introduction, le GIP ATGeRi rappelle les grands principes de la datavisualisation et notamment les quatre grandes étapes que sont la définition d'un objectif, la préparation des données, le choix du mode de représentation et la création d'un rendu .

Aussi, il rappelle les besoins de la datavisualisation dans PIGMA et de l'outil intégré à la plateforme, tels qu'exprimés lors de l'atelier Dataviz qui s'est tenu le 25/04/23.

Le premier besoin était de disposer d'un **outil permettant de déployer des datavisualisations pré-constituées** par les éditeurs ou administrateurs, cet outil devant être différent d'un outil de création de dataviz à la volée. Cet outil doit pouvoir être **intégré à la plateforme** et **utiliser la donnée déposée, sans retravail à priori** de cette donnée. Seuls les éditeurs des organisations et les administrateurs de la plateforme sont amenés à constituer ces datavisualisations, du fait de leurs connaissances techniques ou thématiques de la donnée.

L'intégration dans le catalogue doit pouvoir se faire au sein de l'onglet Données (ce qui suppose une gestion des droits de consultation) et autoriser d'emblée plusieurs types de graphiques (anneau, chiffre-clé, histogramme, courbe, treemap).

Une **feuille de route 2023** pour la datavisualisation a été produite à l'issue de cet atelier et un cahier des charges a été soumise auprès du prestataire de la plateforme PIGMA (NeoGeo).

2.3.2. Confrontations des besoins aux solutions techniques

La proposition technique de NeoGeo a constitué en une étude de faisabilité avec un preuve de concept, centrée autour de l'outil Metabase.

À la suite de l'analyse de la proposition, le GIP ATGeRi a pu dresser **plusieurs constats**.

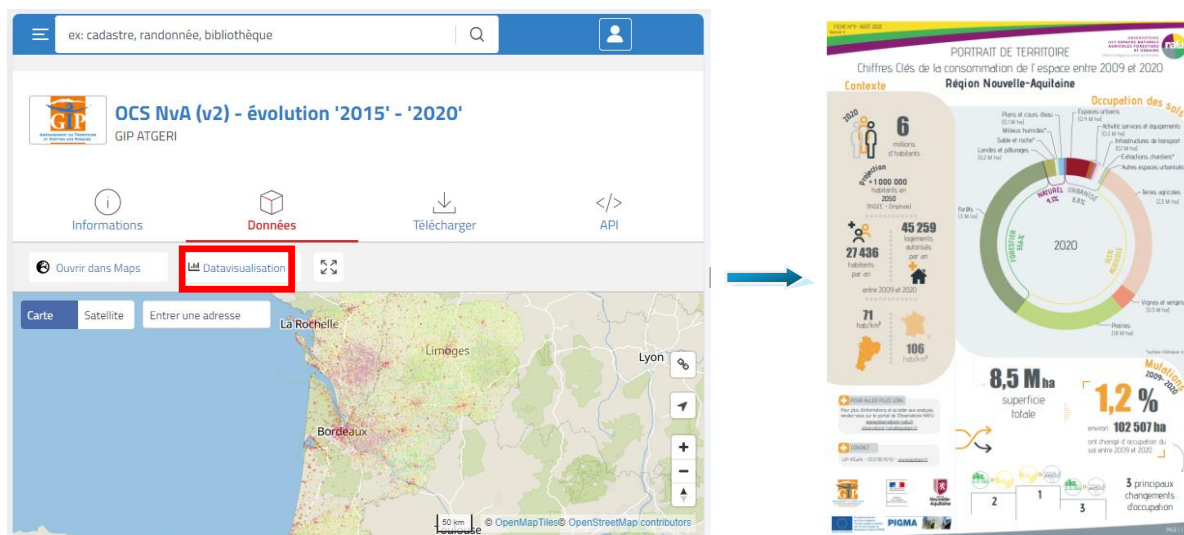
La solution détaillée dans l'étude ainsi que les ajustements et améliorations proposées par la suite ne constituent **pas une solution définitive ni opérationnelle**. Les développements visant à mettre en place un POC, des développements ultérieurs seraient nécessaires pour amener l'outil dans une forme opérationnelle, livrable sur la plateforme.

Second constat : les outils de datavisualisations nécessitent de **consommer des données traitées** en amont. Ce traitement a donc pour but de modifier les données initiales pour les rendre consommables par l'outil de datavisualisation. La conséquence directe de ce fonctionnement est une **complexification de l'administration et de l'architecture** de la plateforme, avec la création d'une base de données séparée pour la dataviz et la constitution d'un processus via ETL, non-intégré et dépendant de l'administrateur de la plateforme pour fonctionner.

De plus, les outils analysés ne permettent **pas de gérer simplement** ou à moindre coût **les droits et autorisations** de consultation sur les dataviz.

Ces différents constats ont fait prendre conscience que **les outils n'étaient pas assez mûrs** pour une solution de datavisualisation répondant aux besoins exprimés.

La datavisualisation dans PIGMA sert à **valoriser les jeux de données du catalogue** et vont de pair avec les travaux sur la qualité des données et des jeux de données. **Pour répondre à cet objectif de valorisation des jeux de données**, une solution intermédiaire a été proposée : celle de **valoriser les datavisualisations existantes sur un jeu de données** en permettant la redirection vers ces datavisualisations via un bouton accessible depuis l'onglet Données.



2.3.3. Exemples de datavisualisations partagées par des partenaires

Le premier exemple ci-dessus concerne la donnée d'occupation du sol, diffusée par le GIP ATGeRi. Donnée géographique de suivi de la consommation des espaces, des portraits de territoire ont été produits en 2022. Ces portraits au format PDF sont disponibles **via le bouton Datavisualisation**.

Le SMEAG, le syndicat du bassin versant de la Garonne, possède un catalogue propre reposant sur la même solution que la plateforme PIGMA. Cette plateforme abrite de nombreuses datavisualisations concernant la qualité des eaux, la biodiversité, l'étiage, etc. Ces datavisualisations découlent de

données tabulaires dont les profils hydrologiques des points nodaux de Lamagistère, visualisables ci-dessous.



Enfin, Camille ROMARY de l'AREC NA rappelle le contexte dans lequel les datavisualisations de l'Agence Régionale d'Évaluation Environnement et Climat ont été produites, notamment la spécificité de la diffusion de ces données via deux plateformes : [Terristory](#) et [OREGES](#). Elle expose les différentes contraintes techniques liées à la production de ces datavisualisations et rappelle les questions à se poser avant de produire une datavisualisation : quel est l'objectif, la cible et ce que l'on veut faire visualiser ? Les données de suivi du stockage carbone font ainsi l'objet de datavisualisations spécifiques.



Ces trois datavisualisations ont pu être valorisées au sein du catalogue PIGMA avec la fonctionnalité présentée lors de la présente réunion. Le GIP ATGeRi cherche désormais d'autres structures possédant des datavisualisations de jeux de données existants ou à référencer, afin de **permettre la découvrabilité des données du territoire**.

2.3.4. Échanges

La Ville de Cenon, représentée par David DELHORBE, possède un portail géographique qui dispose d'un tableau de bord permettant de faire parler les données, à destination des élus et des décideurs. Cet outil, développé en interne, repose sur la même base de données et est alimenté par différents indicateurs calculés depuis la base, à la demande, en fonction de l'expression des besoins. Ainsi, la ville possède différentes datavisualisations, co-construites avec les services métier, qui pourraient être valorisables dans le catalogue PIGMA.

L'Agglomération de Brive, représentée par Matthieu ANGLARD, monte actuellement un observatoire basée sur GeoKey (Business Geographic), avec un système de vues implémentées dans l'outil ou connectées aux outils spécifiques de différents services. La consultation de ces datavisualisations est pour le moment interne à l'agglomération mais une réflexion est en cours pour ouvrir la consultation, à terme.

Le SCOT de Haute-Gironde, représenté par Philippe CIANFARANI, utilise une [solution à façon de datavisualisation à la demande](#), basée sur Lizmap et possède des indicateurs spécifiques au SCOT. Le CD 64, représenté par François ORDOQUI, possède des [datavisualisations sur son portail Open Data](#), reposant sur la solution OpenDataSoft, très utilisées pour des croisements entre des données sémantiquement liées. Les utilisateurs manipulent peu l'outil par eux-mêmes mais consultent beaucoup les datavisualisations mises à disposition et déjà pré-travaillées, ce qui rejoint l'objectif d'améliorer la découvrabilité des données par la datavisualisation de ces données.

2.3.5. Actions PIGMA en 2024

La fonctionnalité de partage de datavisualisations existantes est **un arbitrage, une solution intermédiaire** permettant de répondre partiellement aux besoins exprimés compte tenu de la maturité actuels des outils, et à l'objectif principal de valorisation des jeux de données du catalogue.

La fonctionnalité présentée est actuellement **en cours de développement** et sera déployée sur la plateforme au premier semestre 2024. Les éditeurs auront ainsi la possibilité de renseigner directement dans la fiche d'un jeu de données le lien vers la datavisualisation existante.

Un travail de remontée des datavisualisations existantes sur le territoire doit également être mené, afin de valoriser les jeux de données déjà référencés ou à référencer.

Enfin, si les outils ne sont pas assez mûrs aujourd'hui, le GIP ATGeRi reste en **veille sur ces différents outils de datavisualisations** pour faire évoluer la fonctionnalité à l'avenir.

3. Relevé de décisions

Le GIP ATGeRi propose :

- **L'accompagnement de partenaires dans la montée en qualité** de leurs jeux de données
- **La mise en place progressive d'un Label Qualité** des jeux de données sur la plateforme
- **Le recensement des datavisualisations existantes** chez les partenaires
- **Le déploiement d'une fonctionnalité de partage** des datavisualisations existantes