





Groupe de Travail Qualité des Données Géographiques Commission Règles et Qualité du CNIG

Compte-rendu de réunion du 7 avril 2022

Participants:

Nom Prénom	Organisme	Présent	Excusé
Geoffrey ALDEBERT	Etalab - data.gouv.fr		х
Bernard ALLOUCHE	Cerema / DG / supervision des données		Х
Jean-Marie ARSAC	AZIMUT	Х	
Mathieu BECKER	ISOGEO		Х
Jérôme BOUTET	Idéo Ternum Bourgogne-Franche-Comté	Х	
Benjamin CHARTIER	Consultant - anim. GT Open Data Afigéo		Х
Laurène DEBRAY	Office international de l'eau / Sandre	Х	
Olivier DISSARD	MTE/CGDD Admin données algos codes	Х	
Arnauld GALLAIS	Cerema Ouest	Х	
Thomas GRATIER	Etalab - data.gouv.fr	Х	
Guillaume GRECH	UMS Patrinat, OFB - MNHN		Х
Marie LAMBOIS et Marion LACROIX	IGN / Cellule normalisation		х
Sebastien LAUNAY	IGN / Dépt Contrôles Qualités	X	
Mathieu Le Moal	Axes Conseil	X	
Marie MAHIER	OFB / SIG	X	
Steven PIEL	OFB / Service des référentiels marins	Х	
Nicolas PY	IGN Centre Est	Х	
Clémence RABEVOLO	IFREMER		Х
Mathieu RAJERISON	Cerema Méditerranée		Х
Gessica REYNAUD	Géomap - Imagis		Х
Stéphane ROLLE	CRIGE PACA		Х
Benoît SEGALA	Consultant - Cabinet d'études	Х	
Pierre VERGEZ	IGN / Mission CNIG		Х
David VIGLIETTI	Office international de l'eau / Sandre	Х	

Ordre du jour :

- Validation du <u>précédent compte-rendu</u>. Points d'info et d'actu.
- Qualité des données géographiques du Sandre
- Classification des usages de l'information géographique
- Retour d'expérience d'implémentation du GUF dans Géonetwork v4
- · Point divers : script pour QGIS

Prochaine réunion : 30 juin 2022 à 9h30

Les documents relatifs à cette réunion sont disponibles ici.

1. Validation du précédent CR - Points d'info et d'actu

Revue des actions du précédent compte-rendu :

Le compte-rendu du précédent GT CNIG QuaDoGéo est relu et validé.

Actions réalisées :

- Le registre des mesures QuaDoGéo a été actualisé sur le Géocatalogue mais A. Gallais regrette de ne pas avoir la main sur l'actualisation : demande doit être faite au BRGM, processus long. Le BRGM souhaite justement s'orienter vers un processus où les registres pourraient être édités par des personnes extérieures (cf. <u>présentation des évolutions du Géocatalogue</u> en Commission Règles et qualité du 31 mars).
- S. Rolle a republié les deux fiches méthodologiques précision de position et précision thématique sur le site du CRIGE.

Actions présentées en séance :

- Etudier le GUF de GeoNetwork et tester la restitution sur une plateforme GeoNetwork 4 (cf. le retour d'expérience de l'IFREMER par C. Rabevolo et S. Piel)
- Consolider le 1er niveau de la nomenclature des usages

Actions à lancer ou poursuivre :

- Lister le deuxième niveau hiérarchique en s'appuyant sur les usages référencés dans le document collaboratif (J-M. Arsac, S. Rolle, J. Boutet)
- Fournir d'autres fichiers xml conformes à ISO 19157 (M. Lambois)
- Enrichir le rendu de l'outil de restitution de la qualification des données avec de nouvelles informations graphiques, développer un traducteur vers la <u>maquette</u>, établir le lien avec la nomenclature des usages (M. Rajerison)

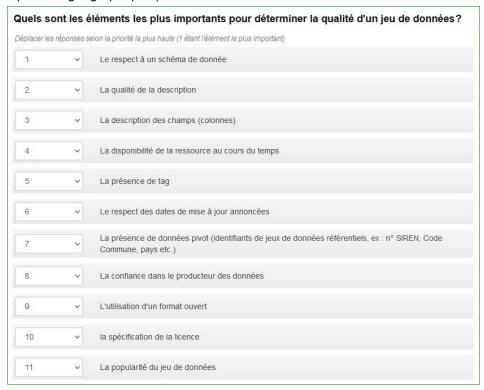
Point d'infos & actus :

- Le <u>Bilan annuel du GT CNIG QuaDoGéo</u> a été présenté à la Commission Règles et qualité du 31 mars (1^{ère} réunion de la Commission sous la présidence de M. Lambois).
- Cette Commission a également été l'occasion d'une présentation de schema.data.gouv.fr
- Le GT CNIG Métadonnées, animé par M. Lambois, repart en trombe : trois réunions très riches se sont tenues depuis début janvier. <u>Cf. Présentation du GT Métadonnées</u> en Commission Règles et Qualité du 31 mars. Les liens étroits entre GT CNIG Métadonnées et GT CNIG QuaDoGéo se traduisent par la participation croisée de M. Lambois et A. Gallais. Le GT Métadonnées voit un intérêt à ce que le modèle conceptuel du lot de données soit transmis via les métadonnées.
- B. David et L. Lemaire ont présenté le projet de catalogage catalogue.data.gouv au <u>GT CNIG Métadonnées du 24 février</u>. Il en est ressorti que "(...) c'est un chantier dont le GT Métadonnées (pour les thèmes) et le GT QuaDoGéo (pour les usages de l'information géographique) peuvent se saisir."
- Le <u>GT Star-Elec</u> a intègré les recommandations du <u>Modèle pour les chapitres qualité et métadonnées des Géostandards</u> du GT QuaDoGéo, qui est un embryon de "standard des standards". Après le standard éclairage public, Star-Elec est le deuxième standard à intégrer ces recommandations. Un retour d'utilisation avec des propositions d'évolutions a été fait à cette occasion.
- A. Gallais signale le lancement de nombreuses Commissions et de nouveaux groupes de travail sous l'égide du CNIG renouvelé :
 - La nouvelle Commission "Besoins et usages" recueille les besoins qui se traduisent

15/04/22 2 / 9

potentiellement en nouveaux standards.

- Les groupes de travail <u>Bâti-ID</u>, <u>ERP</u>, <u>Risques</u>, <u>Friches</u> ont été lancés.
- Le <u>GT CNIG Risques</u> est animé par le département de normalisation de l'IGN et entreprend l'action de refonte des géostandards COVADIS PPR etc.
- Le <u>GT Bâti-ID</u> est propulsé par deux entrepreneurs d'intérêt général (EIG).
- Le <u>GT CNIG Friches</u> (cf. <u>Github</u>) est placé sous l'égide du CNIG en coordination avec Etalab schema.gouv.fr
- Le <u>GT Atlas des Paysages</u> est en passe d'être lancé, son mandat a été validé par la Commission des Standards et il sera probablement animé par le département de normalisation de l'IGN.
- S. Rolle informe que la prochaine formation "Qualité des données géographiques" se déroula du 2 au 13 mai dans une version renouvelée et toujours gratuite!
- M. Rajerison se réjouit de la création d'un groupe sur la standardisation au sein d'Etalab et il souhaite une meilleure prise en compte de la dimension géographique via des spécifications de qualité qui guident le contrôle des données et permettent de meilleures restitutions de l'évaluation de la qualité. Il informe également d'un <u>questionnaire</u> de data.gouv.fr sur les critères d'évaluation de la qualité d'un jeu de données (non spécifiquement géographiques) :



La qualité sera un axe de développement important pour Etalab, notamment en renforçant les liens entre data.gouv.fr et schema.data.gouv.fr. Data.gouv.fr s'intéresse à la qualité des méta-données, à la popularité des jeux de données, la fiabilité des producteurs, leur réactivité dans l'actualisation des jeux de données. A. Garrone, de la DINUM recherche des bêta-testeurs sur le portail, la qualité, etc... notamment la nouvelle version du portail qui intégrera des scores pour les jeux de données.

- En 2021, la France a confirmé sa position de leader européen en matière d'open data, comme en atteste le <u>classement de l'Open Data Maturity Report 2021</u> de la Commission européenne, qui a placé la France à la première place devant l'Irlande et l'Espagne.

15/04/22 3 / 9

- Marie Lambois mentionne que les travaux 19157-3 (registre de mesure qualité) commencent et qu'un "appel à mesures" a été lancé afin de déterminer quelles mesures devront être considérées. Le GT QuaDoGéo devra ainsi indiquer quelles mesures seraient pertinentes à pousser dans la norme. A. Gallais suggère : "les 28 mesures du <u>registre des mesures de la qualité des infos géographiques</u>". (moins la "<u>Classe de précision au sens de l'arrêté du 16 septembre 2003</u>" spécifiquement française ?).
- T. Gratier informe de la <u>publication sur data.gouv.fr</u> de la base de données nationale des bâtiments (BDNB). Ce projet du CSTB en association avec l'ADEME, l'IGN, le CEREMA et d'autres acteurs de la donnée publique, préfigure la mise en œuvre d'un référentiel commun des bâtiments, portée par le <u>projet Bat-ID Etalab</u>, il est <u>documenté sur ce Gitlab</u>.

2. Qualité des données géographiques du Sandre

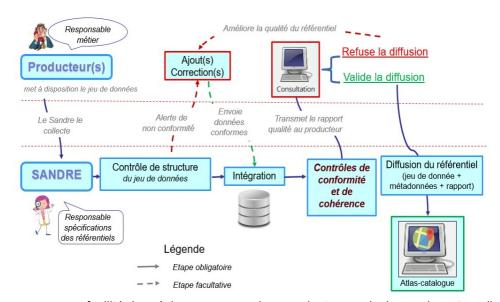
Par L. Debray et D. Viglietti. La présentation est disponible ici.

Le Service d'administration nationale des données et référentiels sur l'eau (<u>SANDRE</u>) s'est attaché à établir un langage commun pour les acteurs ou experts du système d'informations sur l'eau (SIE). Différents outils documentaires ont été élaborés tels que le dictionnaire de données de données de référence, ainsi que des scénarios d'échange de données y compris pour les services web. L'ensemble se réfère à des normes ISO.

Le « code Sandre » est un code unique attribué à chaque donnée relative à un concept du dictionnaire de données Sandre. Il lui permet d'être référencée sans équivoque dans les systèmes d'informations.

La présentation porte sur le processus de contrôle de conformité des jeux de données de référence. Jusqu'en 2016, le Sandre mettait à disposition les données de référence du SIE sans information relative à leur qualité. En 2017 le Sandre décide d'étendre sa certification ISO 9001 aux processus de gestion des jeux de données géographiques. Dans ce contexte a été décidé la mise en place d'une <u>interface de visualisation des résultats des contrôles réalisés</u>, avec le triple objectif : d'évaluer la qualité des jeux de données diffusés ; de permettre aux utilisateurs de prendre connaissance de ces informations pour adapter leur utilisation ; et de viser l'amélioration continue de la donnée.

Contexte: processus général



Le processus a facilité les échanges avec les producteurs ainsi que leur travail de

15/04/22 4 / 9

correction. Il est ensuite illustré à travers deux exemples, celui des aires d'alimentation de captage et celui des stations hydrométriques.

Ces exemples décrivent la vingtaine de contrôles effectués sur les différents référentiels, principalement des contrôles de conformité et de cohérence.

Ces contrôles sont restitués via une "échelle de criticité" à quatre niveaux : gravité nulle (c'est ok), faible (juste une info), moyenne (avertissement), élevée (erreur), avec leurs conséquences sur la diffusion et le besoin de correction. Les rapports de contrôle se présentent avec une restitution de synthèse très lisible (utilisation de codes couleurs) qui renvoie le cas échéant vers des rapports de contrôles détaillés.

Dans ces travaux sur la qualification de la donnée du SIE, le Sandre s'est appuyé sur la norme ISO 19157 et les travaux du GT CNIG QuaDoGéo pour proposer une évaluation pour chaque critère de la norme, et s'inspirer des travaux relatifs à la restitution graphique simplifiée de la qualité des données.

Les perspectives reposent sur la gestion des faux positifs ; davantage de contrôles en étendant les critères d'évaluation ; l'amélioration de la restitution des résultats ; la fourniture de rapport simplifié ; et la possibilité donnée aux utilisateurs d'effectuer des retours d'évaluations de la donnée. Les perspectives du Sandre correspondent donc parfaitement à la mise en pratique des travaux du GT ...

Questions / débats

- O. Dissard témoigne de l'importance, vécue à l'occasion de l'accueil d'un stagiaire, de la qualification d'un jeu de données y compris le fait d'en connaître les éventuelles anomalies au moment de son exploitation, car un informaticien ne met *a priori* pas en doute la qualité des données et ne s'attend absolument pas à y trouver des erreurs.
- J. Boutet constate l'appui efficace donné aux producteurs du Sandre pour corriger leurs données. Il remarque l'intérêt du Sandre pour les propositions du GT QuaDoGéo quant à la restitution graphique de qualification des données, et demande si le Sandre serait prêt à y contribuer, dans le cadre de son implémentation dans des outils de diffusion des métadonnées. D. Viglietti rappelle la nécessité pour l'OFB de passer par des marchés pour les développements. A. Gallais confirme que les GT du CNIG, basés sur le volontariat, atteignent rapidement leurs limites en développement d'outils logiciels, et que la question de la possibilité de faire appel à un prestataire a été et doit encore être remontée au CNIG. J. Boutet convient qu'il s'agit de mutualiser les efforts des différents groupes.
- N. Py s'interroge sur l'intérêt de publier des contrôles révélant des anomalies de "gravité importante" sachant qu'il y aura correction et re-publication des lots de données ?
 L. Debray indique que sans évaluation globale ces erreurs bloqueraient la diffusion.
 D. Viglietti complète en indiquant qu'il s'agit d'un processus d'auto-certification, et les rapports sont également utilisés comme outils de communication.
- N. Py sera intéressé d'un échange avec le Sandre quant à l'acception globale de ce processus de contrôle qualité par les producteurs du Sandre. Il propose une rencontre bilatérale Sandre & IGN pour approfondir ces sujets en présentant d'une part les contrôles de conformité en continu et automatisés à l'IGN qui pourraient être utiles au Sandre, et une plus large découverte des outils mis en place au Sandre.
- A. Gallais demande si les modèles (trames, template...) de rapports de contrôle du Sandre et des autres participants peuvent être partagés avec le GT CNIG QuaDoGéo, afin de déterminer l'essentiel d'un rapport de contrôle.
- N. Py interroge le Sandre sur la grammaire utilisée pour traduire les spécifications littérales et les algorithmes en spécifications lisibles par le logiciel. D. Viglietti répond que

15/04/22 5 / 9

les traitements sont assez hétérogènes et s'appuient sur des outils PHP et PostgreSQL et PostGIS, et évolueront potentiellement vers du XML et du code Pythonr.

- A. Gallais demande quels sont les organismes producteurs d'informations géographiques pour le Sandre. D. Viglietti en présente la grande variété en citant l'IGN, le BRGM, le Service central d'hydro-météorologie et d'appui à la prévision des inondations (<u>SCHAPI</u>), les agences de l'eau, et de façon générale tous les organismes liés au SIE.

Décision / Actions :

- Rencontre bilatérale Sandre IGN pour approfondir les questions relatives aux processus de contrôles qualité et leur restitution aux producteurs et aux utilisateurs (N.Py)
- Partager des modèles de rapport de contrôle qualité (tous)

3. Classification des usages de l'info géographique

Par J-M. Arsac. La présentation est disponible ici.

Le groupe de travail souhaite abonder la nomenclature des usages établies en 2020 par J-M. Arsac dans le cadre de la <u>méthodologie pratique pour qualifier des données</u>.

Cette nomenclature constitue un entrant indispensable à la maquette de restitution graphique de qualification de la donnée géographique.

Un <u>document collaboratif</u> (framapad) a été mis en ligne que la communauté des utilisateurs peut abonder avec les différents usages identifiés de l'information géographique.

La difficulté de l'exercice semble d'établir le juste compromis entre une trop grande généricité des usages (qui seraient passe-partout) et des usages trop précis (pas assez généraux).

Le groupe de travail indique la nécessité d'adopter deux entrées : l'une par la thématique concernée, l'autre par l'usage.

S. Rolle suggère de structurer la nomenclature de façon hiérarchique, à l'instar de la nomenclature Corine Land Cover.

Des nomenclatures existent mais, par exemple, le catalogue INSPIRE apparaît trop détaillé. La même remarque peut être faite pour la <u>nomenclature des usages</u> exposée par l'IGN sur son site Géoservices, qui ont le mérite de renvoyer à des usages très, voire trop, concrets, ou trop thématiques, pour apparaître comme génériques.

Le GT QuaDoGéo s'est prononcé pour la sélection d'une courte liste d'actions (un verbe à l'infinitif) correspondants à une dizaine d'usages génériques : recenser ; localiser ; analyser ; cartographier ; adresser ; naviguer / acheminer ; planifier ; gérer ; suivre / observer / évaluer ; servir de donnée référentielle ou et/ou de donnée pivot

Il convient de rattacher à ce premier niveau, un deuxième niveau plus détaillé, dans une organisation hiérarchique de la nomenclature des usages.

J-M Arsac propose d'identifier en tant que telles, les données dont l'usage est également de constituer une donnée pivot (adresse, code INSEE, identifiant universel, etc.)

A l'instar de la nomenclature Corine Land Cover dont les usages sont fondés sur l'échelle d'utilisation et la représentation, il propose un deuxième niveau de nomenclature qui serait fondé sur le thème d'exploitation et sur différentes classes d'échelles du 1/1000 au 1/1000000, éventuellement traduites en classes de précision de position (centimétrique à

15/04/22 6 / 9

kilométrique) ou bien à une classification davantage identifiable par le grand public : de la parcelle au pays.

Les thèmes relatifs à l'usage pourraient s'appuyer en première approche sur <u>ceux identifiés par l'IGN</u> : agriculture ; aménagement du territoire ; biodiversité ; chasse ; climat ; culture ; défense ; eau ; éducation ; énergie ; espace ; europe ; fiscalité locale ; forêt ; innovation et numérique ; médico-social ; mer et littoral ; prévention des risques ; santé ; sécurité ; social ; télécoms ; tourisme ; transports et mobilité. Il convient de les compléter, par exemple : planification / urbanisme ; accessibilité, etc.

En conclusion le critère d' "utilisablité" ou d'usage des données peut se fonder sur :

- 1/ l'action (pour quoi faire)
- 2/ l'échelle / l'étendue du territoire
- 3/ la thématique (le domaine, le métier)

Questions / débats

Un nouveau tour de concertation est nécessaire pour finaliser le 1^{er} niveau de nomenclature des usages et pour bien définir ce que sous-entend chaque verbe car les termes utilisés sont polysémiques. Des usages peuvent être regroupés, la difficulté demeure de trouver le juste compromis entre une nomenclature trop générique, ou trop détaillée.

Dans une nomenclature hiérarchique, le second niveau découle du premier. Or, dans la proposition qui combine action et thématique, il s'agirait plutôt d'une nomenclature de type matriciel (2D).

Suggestion est faite d'ajouter "inventorier" à "recenser", et un usage supplémentaire de type "informer / diffuser" et de soumettre la proposition de nomenclature à un groupe miroir. Chacun peut encore contribuer en utilisant le <u>document partagé</u>.

Décision / Actions :

- Consolider la nomenclature de 1er niveau (tous)
- Lister le deuxième niveau en s'appuyant sur les usages, l'étendue, la thématique (J-M. Arsac, S. Rolle, J. Boutet)

4. Retour d'expérience d'implémentation du GUF dans Géonetwork v4 Par C. Rabevolo et S. Piel. La présentation est <u>disponible ici</u>.

Le SIMM nous fait un retour d'expérience de l'implémentation du retour des utilisateurs (GUF) dans le <u>système d'informations Sextant</u>, en utilisant la version 4 de Géonetwork. Ce retour d'expérience est particulièrement intéressant du fait que des <u>tentatives précédentes</u> avaient échoué au niveau de l'installation.

L'outil de retour utilisateurs existe en natif dans Géonetwork mais n'était jusqu'ici pas activé. L'outil a été mis en ligne en janvier 2022 simultanément à la mise en production de Géonetwork v4.

Cela se traduit par l'apparition d'un menu "Retours utilisateurs" dans l'interface de présentation des données et métadonnées de Sextant, leur permettant d'apporter un commentaire une fois authentifiés.

Le SIMM signale que l'outil de base propose plus de catégories pour l'évaluation qu'un unique champ texte, mais qu'il a dans un premier temps été décidé de n'en garder qu'une par difficulté d'identifier des catégories suffisamment claires et pertinentes, et pour ne pas noyer et décourager l'utilisateur dans un trop grand nombre de catégories à renseigner.

15/04/22 7 / 9

L'accent a donc été mis sur la présentation et le remplissage d'un seul champ de type "commentaire", quitte à faire le nombre de champ par la suite.

Les retours des utilisateurs authentifiés de Sextant sont directement publiés en ligne, les autres utilisateurs doivent préalablement renseigner leur nom et leur adresse mail, et leur avis est modéré par l'équipe de Sextant, qui est notifiée de tous les retours d'utilisateurs à la fois par mail et dans la console d'administration (admin Sextant / Geonetwork).

Le SIMM regrette la difficulté à transmettre les commentaires directement (par l'outil) aux personnes concernées, ceci se fait donc manuellement par les administrateurs, ce qui reste possible tant que le volume reste limité.

L'outil de retours utilisateur offert par Geonetwork v4 est transmis aux portails qui utilisent l'API Sextant permet d'essaimer la fonctionnalité de retour des utilisateurs dans d'autres sites web, par exemple celui du SIMM qui bénéficie ainsi du même procédé.

Les améliorations futures portent sur :

- la suppression des commentaires plutôt destinés à être adressés directement par mail sans recours au dispositif ;
- le tri des fiches de métadonnées en fonction de la notation attribuée : ce tri permis par Géonetwork mais n'a pas encore été activé ;
- l'intégration des balises liées aux retours des utilisateurs (GUF) dans le XML de la fiche de métadonnées afin que ces champs puissent être moissonnés par d'autres catalogues.

En conclusion de ce retour d'expérience C. Rabevolo rappelle le peu de recul et la faible portée très peu de commentaires sur une période de deux mois d'essai. Cependant, il convient d'améliorer la consigne donnée aux utilisateurs afin qu'ils utilisent ce dispositif à bon escient, c'est à dire pour évaluer réellement et/ou donner un avis sur un jeu de données, plutôt que pour faire part d'une anomalie et d'une demande de correction. Mais le dispositif fonctionne dans Géonetwork v4 et permet à un utilisateur de pouvoir interagir avec le producteur ou diffuseur des données.

Suite à cette première phase de rodage, une communication est enivisagée lors de la journée annuelle Sextant, et il s'agit d'un chantier à suivre... lors d'un prochain GT QuaDoGéo.

Questions / débats

S. Piel convient que cette fonctionnalité permet un retour de l'utilisateur a priori au moment où il découvre les métadonnées, ce qui peut paraître contre-intuitif car il est encore à la phase de découverte des données, et pas encore dans la phase de leur exploitation.

<u>Décision / Actions :</u>

- GUF Géonetwork v4, un dispositif à suivre : M. Lambois, C. Rabevolo

5. Points divers

Thomas Gratier présente son développement d'un script dont l'objet est de créer depuis des définitions de "schémas" publiés sur schema.data.gouv.fr un projet QGIS présentant la structure prête à l'emploi.

Les deux variables en entrée sont "url_schema" et "create_from_scratch" (en bon français dans le texte...;-). Ce script reste sujet à améliorations et T. Gratier pourrait ajouter une interface dans un plugin QGIS par exemple. A ce stade le script n'est pas encore publié, mais peut être obtenu sur demande auprès de T. Gratier pour procéder à des tests.

[Hors réunion] M. Rajerison indique que même si nous disposons de gabarits pour la plupart des standards CNIG, l'intérêt de ce script est de produire un environnement d'édition sous QGIS qui intègre automatiquement les règles d'édition définies dans le

15/04/22 8 / 9

GT CNIG QuaDoGéo

schéma de données : les champs à valeurs restreintes se matérialiseraient par exemple par des listes de valeurs, les valeurs min et max par des champs numériques, les dates par des champs calendrier, etc.

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

- Les fiches méthodologiques du Cerema : sur le <u>site du Cerema</u> 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
- 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.
- *(en cours)* Le prototype d'un <u>outil de restitution graphique de la qualification</u> des données géographiques suivant la <u>maquette</u>.
- (en cours) la Nomenclature des usages génériques de l'information géographique
- (en cours) Modèle de parties qualité et métadonnées des géostandards
- Le fil de discussion Géorezo : "Qualité des données géographiques"

15/04/22 9 / 9