

Pôle Entreprises – Industries AFIGEO

Réunion du Groupe de travail OGC - Open Data
Jeudi 25 septembre 2014 – 10h00 à 16h00 – Saint Mandé
Compte rendu

Liste des participants

BESSAGUET	Maud	PPIGE	présente
BREUIL	Landry	CRAIG Auvergne	Présent téléconférence
BUCHAUT	Romain	CRIGE PACA	présent
BURAC	Maurice	GéoMartinique	présent
CEBELIEU	Gilles	IGN équipe Géoportail	présent
CHARTIER	Benjamin	GéoPicardie	présent
CHAUVEAU	Chloé	PIGMA	présente téléconf
CHIGNARD	Simon	Etalab	Présent téléconférence
de BLOMAC	Françoise	DecryptaGeo	présente
DEJOUR	Nathalie	Consultante	présente téléconf
DESBOEUF	Jérôme	SGMAP / Etalab	présent
DUMONTET	Pauline	Geo Bourgogne	présente
GAUBERT	Pauline	CRAIG Auvergne	présente téléconf
GUILLERM	Anthony	CG92- Projet open data	Présent
IRATCHET	Bruno	REALIA	Présent
JONGLEZ	David	Esri France	présent
KLIPFEL	Jean-Pascal	CIGAL Alsace	Présent téléconférence
LADURELLE	Elise	AFIGEO	présente
LAGARDE	Pierre	BRGM	présent
LEOBET	Marc	MEDDE MIG	Présent
LORY	Pascal	IGN	présent
MAFFREN	Sylvain	CRIGE PACA	Présent
MEILLON	Julien	IFREMER	présent
MOURA	Julien	ISOGEO	présent
PEREGO	Andrea	JRC	Présent téléconférence
PORTIER	Thomas	DEPTH - CRIGEOS	Présent téléconférence
QUEST	Christian	OSM France / Etalab	présent
ROMAIN	Pascal	CG Gironde (33) - CODF	présent téléconférence
RYCKELINCK	Guillaume	CIGAL Alsace	Présent téléconférence
TEIXEIRA	Jérôme	SI17	présent
TINAT	Christel	CG92	présente
VERGES	Pierre	IGN	présent

Sommaire

Liste des participants	1
1. La passerelle Inspire Open Data	2
1.1. Nouvelle version de data.gouv.fr et fonctionnement de la passerelle - Jérôme DESBOEUFS – SGMAP / Etalab..	2
1.2. Echanges sur le référencement du nom de l'organisme producteur.....	3
1.3. Echanges sur l'ouverture des données géographiques.....	3
1.4. Echanges sur l'organisation générale, multi acteurs / multi territoires :	5
1.5. Retour d'expériences sur l'intégration de la passerelle – Jérôme TEIXEIRA, SI17	5
2. L'open data dans les métadonnées Inspire	7
2.1. Les travaux du CNIG – Guide de saisie de métadonnées Inspire - Marc LEOBET-MIG/MEDDE	7
2.2. Projet d'alignement Inspire et DCAT-AP – Andrea PEREGO – JRC remplacé par Jérôme Desboeufs – Etalab	7
2.3. Analyse de Pascal ROMAIN CG33 – Collectif open data France.	8
3. Vers des données géographiques plus ouvertes.....	9
3.1. Les données, les licences d'utilisation – Simon CHIGNARD – Etalab.....	9
3.2. La BANO : point d'avancement, processus de construction - Christian QUEST – OSM France	11
4. Photos	14

1. La passerelle Inspire Open Data

Avec la sortie de la nouvelle version du portail data.gouv.fr, la passerelle Inspire / Open Data (qui vise la valorisation des données géographiques sur data.gouv) sera fonctionnelle mi-septembre. L'occasion de présenter ces nouveautés, leur fonctionnement et d'échanger sur leur appropriation.

1.1. Nouvelle version de data.gouv.fr et fonctionnement de la passerelle - Jérôme DESBOEUFS – SGMAP / Etalab

Le projet de passerelle inspire / Open data (<http://afigeo.datacircl.es/>) est adossé et dépend du portail Etalab en terme d'API notamment (<http://data.gouv.fr>) dont la V.3 vient de sortir. Elle est basée sur une nouvelle infrastructure technique dont le cœur n'est plus basé sur CKAN mais sur UDATA (<https://github.com/etalab/udata>, solution open source développée en interne).

Cette refonte permet de mettre plus en avant :

- La réutilisation des données plus que les données elles-mêmes,
- La notion d'activité autour de la donnée : historique de sa réutilisation, « time line » par profil d'utilisateur,
- L'aspect réseau social autour de la donnée,

La passerelle moissonne les portails des IDG, et les corrections apportées lui permettent fonctionner avec l'ensemble des solutions (Géosource/Geonetwork (Prodige, GeOrchestra), EasySDI, Isogeo, ...). Le mapping ISO-19139 vers JSON fonctionne, ainsi que la prise en compte du mot-clé « données ouvertes » dans les métadonnées et des flux WFS ouverts.

Illustration avec GéoBretagne : <http://afigeo.datacircl.es/services/5387a37d785588482f5acb45/datasets>

- Sur 2 165 datasets : 286 sont « marquées données ouvertes / open data »

- Le seul moyen pour les rendre exploitables et visibles sur data.gouv.fr est qu'elles soient « téléchargeable en WFS », et exclus ainsi les jeux de données qui nécessitent une authentification.
- 329 jeux de données sont concernés « téléchargeables en WFS »
- 157 jeux données si on sélectionne les deux critères « open data et WFS ».

Des tests pourront être effectués avec des IDG qui disposent des jeux de données ouverts.

Les perspectives de la passerelle :

- Publication sur data.gouv.fr et régler l'aspect « authentification » sur la plateforme
- Relation données / métadonnées : 3 manières de rattacher une métadonnée INSPIRE au service WFS
 - Référencer le service dans le jeu de donnée (40% des cas d'usages)
 - Créer une métadonnée de service qui contienne le Mapping entre métadonnées/service (méthode recommandée par le CNIG)
 - Référence à partir du GetCapabilities du service (extension Inspire)
- Le téléchargement via WFS en construisant l'url, à partir de formats de données disponibles.

Les chantiers à venir :

- Représentation des données géo dans data.gouv.fr (description, outils) une V1 devrait être proposé. Si des choses sont déjà réalisées dans les IDG (au delà de Géosource), leur contribution est bienvenue.
- Produire des statistiques en lien avec le nom de l'organisme producteur (via les métadonnées du géocatalogue). Des tests d'extraction des noms des organismes montrent des difficultés : (acronymes, nom non complet, services de l'Etat sans le territoire concerné...)

1.2. Echanges sur le référencement du nom de l'organisme producteur

- ✓ Q : dans les métadonnées d'Etalab, le producteur figure dans le champ « contact » : des choses vont-elles être corrigées à ce niveau ? ou un rapport de qualité sera-t-il fourni aux IDG afin qu'elles améliorent leurs métadonnées ?
- ✓ R : Jérôme a l'espoir d'aller au-delà d'une simple table de correspondance (trouver des nomenclatures ou des associations de nomenclatures intéressantes). A discuter : voir comment chacun se situe, puis partir des pratiques existantes pour retenir un référentiel.
- ✓ Q : Quelle est la démarche de data.gouv.fr vis-à-vis des référentiels d'organismes ?
- ✓ R : la stratégie de publication sur data.gouv comprend la création manuelle d'un organisme producteur. C'est donc l'organisme qui détermine la nomenclature. Dans le cas d'une publication automatique, la question du référentiel se posera, ce qui peut faire l'objet d'un travail croisé avec la passerelle.

1.3. Echanges sur l'ouverture des données géographiques

Pour Marc Leobet : la gestion des licences est un sérieux problème :

- Le CNIG : a recommandé l'utilisation d'un mot clé (et non des licences), facilement repéré par les systèmes informatique, mais a ses limites et dépend de la bonne volonté des services.

L'accompagnement ne donne pas de résultats rapides, et la politique nationale d'Etalab pose les bonnes questions : si on respectait la circulaire du 1^{er} ministre, la plupart des données du Géocatalogue sont des données ouvertes, mais seules 1/3 sont téléchargeables, comment améliorer ce nombre ? Pour la Commission Européenne, en 2016, à minima 96% des données du géoportail seront consultables et téléchargeables. On doit changer de rythme.

Pour Pierre Lagarde : les problèmes liés à l'identification ou la non ouverture des données :

- Problème d'ordre politique (pas de volonté) mais non majoritaire
- Problème d'identification des données Open data est récent, donc pas encore largement adopté.
- Problème d'ordre plus technique : les organismes n'ont pas mis à disposition leur données en WFS (car soit directement téléchargeables, soit une part d'erreur)
- Techniquement cette plateforme va recenser les données sur data.gouv, mais qui va prendre la responsabilité ? est ce que tout ce qui est publié dans géocatalogue serait open data ?

Pour Jérôme Desboeufs : il existe des outils qui doivent faciliter le repérage de ces données (celles disponibles en téléchargement direct non-WFS) mais cela nécessite une puissance de calcul et une consommation de bande passante supérieure côté passerelle.

Inspire ne s'intéresse pas au données open data, le geocatalogue ne doit pas être le lieu pour celles-là. Pour GéoPicardie, le service WFS pour le rapportage n'était pas actif, les données sont détectées par la passerelle mais pas par le géocatalogue.

Jérôme Desboeufs présente quelques statistiques sur les données de la passerelle (donc à prendre en l'état des développements actuels), ne distinguant pas les données avec thème Inspire :

- Données marquées OD : De 1% en juillet on est passé à 5%.
- Celles associées au service WFS : 8%
- Celles qui sont téléchargeables (WFS ouvert fonctionnel) : 4%
- Un potentiel de 251 données (1%) donc inférieure au potentiel de ce qui existe dans les IDG.

Des discussions au sein du GT CNIG « liste des données » ont porté sur la déclaration de données Inspire qui devraient être ouvertes (sauf celles avec licence payante...). Toutes les données de Collectivités, Services de l'Etat ne rentrent pas dans les cas de « restriction ». C'est un problème de formation, d'accompagnement et de temps passé à reprendre les fiches pour les taguer « données ouvertes ».

Pour David Jonglez, il ne faut pas faire de confusion entre données ouvertes et données gratuites : on peut ouvrir des données payantes et on peut avoir des données gratuites mais pas ouvertes. L'ouverture des données c'est le droit de modifier, enrichir ou de réutiliser ces données. La notion de produit dérivé est essentielle dans la notion d'ouverture. On voit, dans le domaine des transports notamment, venir une maturité croissante de mixité des modèles économiques autour des données. Cette démarche est une vraie opportunité pour les données Inspire :

- l'occasion de voir une première grande réutilisation de la donnée Inspire via le portail Etalab,
- d'augmenter la qualité des métadonnées,
- d'améliorer la façon dont on les ouvre,

- de donner de la visibilité à la valeur des données Inspire à travers leurs ré-utilisations. La valeur d'un jeu de donnée dépend directement de sa réutilisabilité.

Romain Buchaut : les IDG ont prôné Inspire qui est un moyen pour mutualiser, encourager les usages des données. La plupart des portails d'IDG proposent des données récupérées sur du long terme et mises à dispo dans un cercle d'utilisateurs défini. L'aspect accompagnement et politique de libération est important ; les démarche open data des villes fonctionnent grâce aux délibérations sur l'ouverture des données.

Benjamin Chartier souligne qu'en Picardie, seuls 80 organismes adhèrent à GéoPicardie , donc le potentiel réel des données disponibles est sous identifié.

1.4. Echanges sur l'organisation générale, multi acteurs / multi territoires :

- ✓ Q : Benjamin Chartier rappelle que de nombreuses données produites au niveau local, ne le sont pas par la Collectivité mais dans le cadre de prestation, ainsi plusieurs contacts sont identifiés dans les métadonnées (producteur, gestionnaire, contact etc..) quel résultat sur data.gouv dans tel cas ?
- ✓ R : data.gouv identifie principalement le producteur de la donnée, ne recense pas de « contacts » en tant que tel. Il y a aussi une notion de distributeur/fournisseur.
- ✓ Pour Christian Quest, il faut voir data.gouv comme une vitrine de l'existant localement, ce n'est pas le but d'avoir tous les détails. Mais dans les IDG, la visibilité du producteur de données est essentielle pour valoriser les membres.
- ✓ L'objectif est d'impliquer les producteurs, c'est aussi la raison pour laquelle ils doivent eux mêmes faire la démarche de publication. Marc Leobet, pense que les gens ne feront pas la démarche sans accompagnement.
- ✓ Pierre Lagarde s'interroge sur l'organisation, la façon dont s'inscrit cette démarche par rapport aux IDG, au Geocatalogue.
- ✓ Pascal Romain : les sociétés comme Datapublica / OpenDataSoft jouent un rôle, comme les IDG, de diffuseur, d'intermédiaire, une articulation est donc à trouver, selon une logique simple et en s'appuyant sur l'existant. Chacun doit trouver sa place dans l'écosystème.
- ✓ Quel intérêt pour les producteurs info géo d'adhérer à une IDG ? pour les producteurs open data de passer par une plateforme d'agrégation ? Quel apport ? Quelle création de valeur ?
- ✓ Une mise en correspondance, un regroupement des données, leur enrichissement, via une agrégation pour proposer d'autres valeurs. Ces sociétés offrent ensuite des services à des Collectivités qui n'ont pas les moyens pour diffuser leur données...
- ✓ Un moyen de regrouper de façon pyramidale, et éviter de moissonner divers catalogues.

1.5.Retour d'expériences sur l'intégration de la passerelle – Jérôme TEIXEIRA,

SI17 <http://geoplateforme17.fr/>

Localement, il faut être pragmatique et le SI 17 s'attache à :

- sensibiliser les acteurs (une centaine de partenaires),

- suivre les recommandations du CNIG dans les métadonnées,
- mettre à disposition un catalogue, et assurer un accompagnement, de la pédagogie,

La passerelle constitue une opportunité pour sensibiliser les décideurs aux démarches liées au partage des données, elle vient illustrer et vulgariser.

Les enjeux sont d'ordre organisationnels, juridiques (délibération des Elus)... Le fait de remonter des données publiques locales au niveau national peut gêner des Elus : au SI17 un second item « données publiques » a été rajouté à « données ouvertes ».

La donnée et la métadonnée sont visibles sur le portail : geoplateforme17 : par ex :

- si la fiche métadonnée pour le cadastre signale que la donnée est publique, le WMS sera sécurisé (accessible à certains partenaires, licence soumise à la volonté du producteur)
- des tests de la passerelle ont été faits avec la DDTM de Charente Maritime qui a qualifié sa donnée comme « ouverte », un WMS public y a été associé.

3 cas de figure pour les données : données : fermées /ouvertes / publiques auxquelles sont associés des Web service publics / sécurisés (l'acronyme de l'organisme figure dans l'url).

Techniquement nous n'avons pas de soucis (travail avec Depth France) tout l'enjeu est dans l'organisationnel (avec catalogues nationaux, ou thématiques locaux qui chevauchent les territoires).

Des « catalogues enfants » émanant de la Géoplateforme sont créés sur les sites partenaires (emprise de la Collectivité) et délègue la gestion des services OGC (WMS puis WFS).

Le SI17 est confronté à 2 feuilles de routes nécessitent un point de convergence :

- WFS / Etalab
- WMS / SI17

Pour Thomas Portier, il faut clarifier la façon dont se fait le lien entre métadonnées et web service : un travail à faire pour préciser les standards à mettre en œuvre.

- Marc Leobet renvoi au Guide CNIG « Métadonnées de service » et appels à commentaire « guide implémentation d'un service... » ou comment implémenter un getcapabilities.

2. L'open data dans les métadonnées Inspire

Réaliser un point d'information sur la place de l'open data dans le modèle actuel des fiches de métadonnées au format Inspire, et les perspectives d'évolution liées aux standards existants (ISO), aux travaux autour de l'alignement des métadonnées Inspire au format DCAT-AP...

2.1. Les travaux du CNIG – Guide de saisie de métadonnées Inspire - Marc LEOBET-MIG/MEDDE

Le calendrier législatif de la MIG n'est pas le même que celui d'Etalab, très réactif.

Annonce relative à la publication des données des services de l'Etat non soumises à restriction.

Un constat : à quoi sert une licence pour les données. Un cadre légal qui s'impose à nous, 98% des données du geocatalogue ont le caractère téléchargeable.

L'idée serait d'aller plus loin, en ne se basant plus sur le volontariat des acteurs. Diffuser les données des services de l'Etat non soumises à « restriction » par défaut. L'administrateur de données devra justifier/ déclarer en quoi la donnée n'est pas publiée si elle ne l'est pas.

Calendrier et ressources : Les Ministères de l'écologie et de l'agriculture, lors de leur dernier comité de coordination information géographique, ont acté la décision. Les systèmes informatiques internes doivent être prêts. Les systèmes informatiques internes (Prodige, GeolIDE, Carmen...) doivent être adaptés pour que soient exportées les données ouvertes en interne vers le web. A terme, 70 000 jeux de données sont susceptibles de sortir. Il faudra ensuite gérer l'accompagnement...

DCAT et Inspire ont des modèles de données proches. Andrea PEREGO travaille sur une matrice plus importante : sur les métadonnées de services.

Le consensus avec les autres Etats membres :

- trop investis dans Inspire, pas de changement prévus pour fabriquer des données en DCAT (qui peut être incomplet : par ex il y manque les coordonnées).
- Mais convaincus que l'infrastructure, les données, métadonnées, services Inspire, doivent alimenter d'autres systèmes moins orientés « info géo », et favoriser leur valorisation / réutilisation par des développeurs non Géomaticiens pour de nombreux (applis mobiles etc..). Sanctuariser les infrastructures.

2.2. Projet d'alignement Inspire et DCAT-AP – Andrea PEREGO – JRC = présenté par Jérôme Desboeufs – Etalab

Voir support de présentation

L'objectif du format DCAT : une standardisation générique de la donnée (et pas uniquement géographique).

C'est un format par essence sémantique (issu de l'écosystème W3C / RDF).

Le projet d'alignement vise donc à établir une recommandation de correspondance entre Inspire et vocabulaires sémantiques associés à DCAT. En effet DCAT ne couvrant pas Inspire, il faut s'entendre sur les compléments à retenir, de préférence en s'appuyant sur des vocabulaires existants (par exemple vCard pour les contacts).

Pour le JRC, le projet Are3NA est un moyen de financer ce type de travaux. Il faut voir cette démarche comme un complément pour l'interopérabilité d'Inspire.

2 niveaux :

- Core : 1er niveau qui intègre les éléments d'Inspire dans DCAT-AP
- Extended : 2ème partie : travailler sur ce qui manque dans DCAT

Marc Leobet précise que le groupe de maintenance se réunit dans 15 jours : étendre un standard c'est faire du non standard : pourquoi utiliser DCAT pour des métadonnées de service ? Ce n'est pas fait pour ça. DCAT-AP est différent de DCAT, on assiste à une floraison de système, l'objectif est de défendre les spécifications Inspire.

Etalab travaille avec un modèle JSON « maison », compatible DCAT. Il n'est pas prévu pour le moment de passer sur un modèle purement sémantique. La passerelle fonctionne sur une approche équivalente.

2.3. Analyse de Pascal ROMAIN - CG33 / Collectif open data France.

DCAT-AP doit préciser la façon dont est implémenté DCAT : favoriser l'expression de la sémantique dans les métadonnées. Cette convergence vise à éviter que les producteurs aient à modifier des choses, leur permettre d'être interopérables avec d'autres communautés.

Cette dimension Web de donnée, rejoint les travaux du W3C sur la traçabilité de la donnée (provenance, modifications, réutilisations, enrichissement des métadonnées...) et permet au producteur de connaître ces changements. **Ces éléments doivent guider nos efforts, un bon terrain de dialogue entre communautés qui ont identifiés l'importance des métadonnées...**

L'intérêt :

- Faciliter les collaborations entre OSM et producteurs institutionnels...
- Permettre aux acteurs Open Data d'avoir une vision plus pérenne de l'évolution de la donnée, ainsi bâtir des modèles économiques plus stables (leur évite d'avoir à re-moissonner les données, mais plus complexe que scraper un page web).

Un terrain porté par OGC, W3C doit avoir comme fil rouge : faciliter l'interopérabilité, rajouter des champs /descripteurs selon divers contexte sans perdre la matière.

Question de **l'évolution d'ISO-19115:2014** : pour Marc, les Etats membres ne se sont pas liés à des institutions de normalisation. Inspire n'est pas conforme à ISO-19115 :2003 et les réglementations ne bougeront pas. En revanche, les futurs guides techniques doivent prendre en compte la double conformité, comme les guides CNIG.

3. Vers des données géographiques plus ouvertes

L'objectif est d'échanger pour mieux identifier les données géographiques et les freins à leur diffusion, fournir des éléments aux géomaticiens, animateurs de plateformes en terme de licences les plus adaptées, de techniques pour encourager les réutilisations, en lien avec les notions Inspire.

3.1. Les données, les licences d'utilisation – Simon CHIGNARD – Etalab

Support de présentation : <http://fr.slideshare.net/schignard/open-data-ouvrir-ses-donnees-cest-mieux-en-contrler-lusage>

Simon CHIGNARD collabore avec Etalab, responsable de la partie éditoriale de data.gouv.fr.

Le thème « ouvrir ses données c'est bien, en contrôler l'usage c'est mieux » est volontairement provoquant, avec l'idée d'explorer les modalités, les différents outils disponibles pour accompagner une démarche d'ouverture de données.

Les éléments de contrôle de la réutilisation présentés sous forme d'un schéma de la fusée : chaque étage / outils contribuent à orienter une démarche / trajectoire de réutilisation de données.

Tous ces outils ne sont pas forcément à déployer mais chacun peut être utile. Beaucoup de débat sur les données du transport entre Ministère et Etalab avec des tentations de créer des licences spécifiques selon des jeux de données... ce qui n'est pas nécessaire.

La licence de données et la licence de contenus sont différentes mais se font écho.

Creative commons ; plutôt que d'avoir un copyright, c'est un moyen de moduler les droits, rendre la possibilité de contrôler l'usage qui est fait. « CC » puis des déclinaisons :

- CC-zéro : domaine public, le producteur de contenu cède l'intégralité de ses droits et laisse l'utilisateur en faire ce qu'il veut. « Public domain data licence » mais personne ne l'utilise.
- CC-By correspond à la licence ouverte d'Etalab
- Cc-by-Sa = odbl (celle d'OSM).

Un contrat de licence propose de nombreux avantages par rapport à un modèle contractuel classique (le téléchargement). Dans une licence : il y a un volet « droit » et « devoir ». Dans l'ODbl et la licence ouverte ; les droits sont les plus larges et quasiment les mêmes : on peut modifier, adapter, réutiliser de façon commerciale les données, et même les commercialiser.

Mais le niveau d'obligation de réutilisation est différent !

- PDDL : aucune obligation
- Licence ouverte ou CC-BY : obligation d'attribution : mentionner la source et la date de mise à jour (utilisée par l'Etat, Ministères, Agences, Collectivités,...)
- ODbL ou CC-BY-SA : mêmes obligations mais en plus le partage à l'identique : avec des réalités diverses. 2 voix : licence de biens communs ou n'encourage pas l'utilisation des données par les entreprises ; côté viral : contaminant de la donnée = projet OSM et des collectivités.
- Des acteurs privés (SNCF...) utilisent des variantes de la licence ODbL.

Les éléments de marque : des conditions d'utilisation de la marque RATP existent par ex, pour éviter la confusion pour l'utilisateur final dans les services officiels et non officiels) (optique de protection).

Cet élément comporte aussi des optiques positives. Un moyen de préciser ce que l'on peut ou pas faire.

CGU du portail, API et webservices : les CGU d'un site peuvent être complétées d'infos sur l'usage de données (le détenteur de la donnée n'est pas responsable des risques liés à son usage ou précise la notion de « délais de prévenance » en cas d'interruption d'un flux de données, l'usage qui ne doit pas porter atteinte à l'image du producteur...)

Actions d'animation : il est possible dans les règles d'un concours de réutilisation ; d'orienter par exemple la gratuité du service développé à partir de données ouvertes pendant une durée, éléments qui ne peuvent figurer dans une licence.

Les Processus de labellisation, comme dans le Programme Keolis Rennes : « approuvé par Star », offrent une meilleure visibilité des applications développées, dans les guide touristiques etc.... en retour il peu être demander la remontée de statistiques sur l'usage de l'appli,...

Le message clé : on réinvente parfois des licences maison, alors qu'il existe de nombreuses possibilités d'orienter des choses dans d'autres espaces. S'appuyer sur les licences existantes évitent la confusion et respectent l'objectif d'interopérabilité des données disponibles sur divers territoires.

Echanges et discussions :

- ✓ Q : Françoise de Blomac questionne la notion « d'enrichissement » ? A quel niveau ? Et qu'est ce qui déclenche la notion de « partage à l'identique » ?
- ✓ R : l'idée est que tous ceux qui veulent utiliser / améliorer doivent aussi contribuer. Pour OSM, c'est compliqué car c'est une base multithématique, mais pour BANO le partage est nécessaire seulement s'il y a une amélioration de la base adresse en tant que telle.
- ✓ C'est complexe car il n'existe pas de jurisprudence sur ce point. Ex du fichier des arbres remarquables de la ville de Rennes : on peut l'enrichir en remarquant qu'il en manque, et en ODbL : l'obligation ne tombe que lorsque l'usage devient public (pas en privé).
- ✓ On voit qu'il y a d'un côté la loi vs et la morale.

- ✓ Il y a aussi des licences et des pratiques communément acceptées : quelle source citée pour une donnée du MEDDE dispo sur data.gouv.fr ? c'est le producteur et non le diffuseur.

- ✓ Pour Simon Chignard, ce point doit figurer dans d'autres types de documents, il faut des recommandations plutôt que du formalisme juridique.

- ✓ Q : OSM devient peu à peu un système de référentiel, utilisé comme base / fond et on ne va plus questionner la production elle-même, ODbL, est-elle donc adaptée à ce type d'usages ? C'est tellement complexe d'entretenir un référentiel que l'idée est d'être autant contributeur qu'utilisateur.

- ✓ Simon Chignard n'a jamais vu de licences spécifiques sur les données de référence/registre clé. Quand, politiquement la donnée est considérée comme un « bien commun » : est-ce qu'ODbL n'est pas plus adaptée ?

- ✓ Pour David Jonglez, il y a un vrai vide juridique pour l'information géographique : l'ODbL protège la donnée en tant que Base de données (coordonnées géo et attributs) mais pas les

produits dérivés (ex une représentation RASTER issue d'un jeu de données ODBL, la représentation topologique d'un réseau issu des données OSM, ...) et ne sont pas impactés par l'ODbl.

- ✓ Pour Christian Quest : la question se posera quand IGN ouvrira ses jeux de données, quelle licence sera choisie ? si la donnée est un bien commun il faudra réfléchir à cette notion de « partage à l'identique ». Marc interroge la légalité de la Licence ODbl pour une autorité publique : une contrainte sur la réutilisation des données publiques est-elle contraire à la loi ? car « la loi n'est pas la morale ». Cela fait appel à la notion : valorisation de l'effort, que doit-on mettre en place dans nos territoires ?
- ✓ Exemple de Julien Moura : pendant le procès lié à la tempête Xynthia ; un Elu rejetait la faute sur les équipes techniques, la qualité des données... les collectivités consultent les juristes qui ont un rôle important.
- ✓ Le legal working group d'OSM se pose à l'inverse des questions sur l'intégration de données extérieures, leur origine, le droit d'usage ;

La question de l'animation : 3 fonctions principales pour un portail open data :

- 1/ mise à disposition des données
- 2/ documentation de la démarche
- 3/ une ligne éditoriale en lien

Pas d'animation des villes sur data.gouv, il faut le faire sur les territoires avec l'écosystème local. La création de nouveaux portails interrogent : ils sont nombreux avec au final trop peu de visites. Peut-on envisager des boites à outils exploitables en local ?

Etalab propose des espaces aux producteurs dans lequel ils peuvent se présenter, valoriser des mises en avant...

L'open data a tout son sens dans l'aspect local et territorial de l'animation.

3.2. La BANO : point d'avancement, processus de construction - Christian QUEST – OSM France

Voir support de présentation.

Christian Quest présente Bano1 et la version 2 plus institutionnelle récemment intégrée à Etalab.

L'Adresse est une donnée à fort enjeux, pour des usages importants (secours, communication, itinéraires,) le géocodage de fichier permet le passage de fichiers non géographiques à des fichiers géographiques. Constat : des rapports, des créations d'adresse multi acteurs, des bases de données en silos...

Les données disponibles dans Bano : ont une double origine : open data ou issues du cadastre.

Des exemples montrent les données BD Adresse de l'IGN / cadastre ont en général une grande proximité, avec quelques différences au niveau de la position.

La constitution de Bano :

- 1^{ère} extraction complète en 10 j de la saisie manuelle à la saisie automatiquement.
- Couverture des données adresse avec code couleur : selon l'origine en open data, ou les données manquantes, incomplètes ou mal renseignées.
- 18 millions d'adresse en licence ODbL car une volonté de partage de leur amélioration.
- 3 formats ; csv, shapefile (pour SIG), RDF (turtle) pour les adeptes du web sémantique.

Ces fichiers sont dispo sur github ce qui permet d'avoir une vision d'évolution,

Schéma du fonctionnement : un cycle complet journalier de mise à jour.

Prochaine étape :

- **un guichet unique** pour identifier les erreurs, un fichier déconnecté d'OSM pour faciliter la participation d'acteurs institutionnels et rediffusé sous d'autres licences afin d'alimenter d'autres partenaires, tels que l'IGN...
- Création d'une API de type Rest pour interagir sans l'interface web, si un réel besoin en WFS-T pourquoi pas mais dans un second temps. Un outil de géocodage sera également dispo.

Animation : après un 1^{er} Bano Camp fin juin pour poser les premières briques du guichet unique, réfléchir sur l'API, il est prévu : un Banocamp, voir un Bano Tour pour aller au plus près du terrain...

BANO 2 : s'inscrit dans la stratégie numérique de l'Etat. La DISIC, chargée de mettre en place le SI de l'Etat, interministériel et faciliter la mutualisation des données s'est fixée comme objectif : la création d'une infrastructure de données Adresse (soutien officiel d'Etalab, et de la plateforme)

- Objectif à 6 mois
- 1,5 ETP dédiés au projet (animer le projet, rechercher de nouvelles sources, le développeur travaillant sur la partie géocodage et mise en place de la plateforme)
- et des contributeurs OSM poursuivent le travail,
- un bel avancement sur les « lieux dits », à creuser du côté du cadastre vecteur pour voir si on peut extraire de nouvelles données.

Feuille de route, notamment sécuriser des partenariats nationaux, locaux,

- Démarrage d'un outil de geocodage
- Codesign de la plateforme : co conception souhaitée

Echanges : Guichet unique :

- ✓ Le guichet unique : est-il le lieu où les communes inséreront leur mise à jour d'adresse ? Leur délibération ? l'objectif est d'avoir un outil facile à l'usage des petites communes, et services associés (remontée d'erreur...). Automatiser le processus, et rendre cohérent la démarche...
- ✓ Les guichets uniques se multiplient et malgré leur simplicité les petites communes ont du mal à s'y retrouver.
- ✓ Au-delà du développement d'une API, quel calendrier pour les relations avec les acteurs locaux, l'accompagnement... ?
- ✓ BANO2, il faut le voir comme une plateforme d'API qui doit permettre de déployer au plus près du terrain, aux IDG d'assurer un rôle d'animation.

Jeux d'acteurs :

- ✓ Marc Leobet signale qu'IGN avait proposé des choses dans ce sens.
- ✓ Il y a bien un objectif de convergence avec l'IGN, qui avait débuté la BAN en 2013 mais qui prenait du temps. Associer l'IGN, là où il est fort (un bel outil industriel) OSM et Etalab, La

Poste, ... et la connexion avec les initiatives de mise à jour souvent à l'échelle locale. La Commission donnée du CNIG abordera ce sujet le 2 oct.

- ✓ Question du circuit de l'information, remontée des mises à jour... Tout peut être construit sur 5 grandes briques de base, et c'est nécessaire...
- ✓ Tout cela est en pression et un système vertueux devrait en sortir. Dans la durée, maintenir des référentiels, des API est possible si on s'appuie sur la logique des IDG régionales, départementales. Il faut définir une articulation, créer de véritables flux de données (et non des envois de fichiers....)

Autres :

- ✓ En terme de communication, la présentation ne rend pas visible la stratégie nationale, ce que ne va pas rassurer les communes qui produisent cette données.
- ✓ La donnée en licence ODbI pose problème : l'adresse étant produite par les Collectivités c'est une donnée publique. Cette licence ne leur permet pas de fournir la donnée aux acteurs du privé sans condition, qui garantissent une visibilité de leur territoire (cartes embarquée...). Cela semble aussi difficile pour l'IGN d'en bénéficier en l'état.
- ✓ Un changement d'orientation avec la notion de registre clé (plutôt que référentiel), encore dans le domaine de la réflexion, mais une réelle prise de conscience, et une volonté que ces bases bénéficient de contribution pour être alimentées.

4. Photos

