

Compte-rendu

Réunion du Groupe de travail Open Data

Mardi 30 mai 2017 de 10h00 à 13h00 chez Etalab

Réunion animée par Benjamin Chartier

Compte rendu co - rédigé

Liste des présent(e)s

BOUTET	Jérôme	IDO BFC	présent
CHARTIER	Benjamin	NéoGéo Technologies	présent
DESBOEUFS	Jérôme	SGMAP / Etalab	Présent
DESCHAMPS	Héloïse	GIP ATGERI PIGMA	présente (teleconf)
HAUSTANT	Axel	Etalab	
IRATCHET	Bruno	REALIA	présent
KAZMIERSKI	Mathieu	CRIGE-PACA	présent (teleconf)
LADURELLE	Elise	AFIGEO	présente
LEBARD	Christophe	GIP ATGERI PIGMA	présent (teleconf)
LECOMTE	Carole	GéoNormandie DREAL	présente (tel conf)
MEILLON	Julien	IFREMER	présent
PORTIER	Thomas	DEPTH	présent (teleconf)
QUEST	Christian	Etalab	présent
ROBINET	Elodie	SOLURIS	présente (tel conf)
THOUMAS	David	Open Data Soft	présent
TINAT	Christel	CG92	présente
VERGEZ	Pierre	IGN	présent

Excusé(e)s absent(e)s

BESSAGUET	Maud	PPIGE
BURAC	Maurice	GeoMartinique
HALLEMAN	Fionn	IGN
LEOBET	Marc	MEDDE MIG
MOURA	Julien	ISOGEO
PORNON	Henri	AFIGEO - IETI

Sommaire

Introduction.....	2
1. Actualités réglementaires autour du décret « licence »	2
2. Le décret « Service public de la donnée – SPD »	3
3. Le Socle commun de données nationales - Open Data Local	5
4. Perspective pour le projet Base Adresse Nationale – BAN	5
5. PIA.....	7
6. Open Data Locale / OpenDataFrance et socle commun de donnée.....	7
7. Catalogage et INSPIRE	9
8. Calendrier.....	9

Introduction

AFIGEO réalise un bref point des actions conjointement menées par AFIGEO / Etalab depuis 2012, en lien avec le développement du mouvement open data.

Aujourd'hui Jérôme du SGMAP/ ETALAB co anime le GT Open Data de l'AFIGEO et entérine ces relations de longues dates qui favorisent la rencontre entre 2 communautés.

Le mouvement Open Data et le GT AFIGEO



- 2010: 1ers portails open data de collectivités territoriales pionnières
- 2011: Création d'Etalab, service rattaché au Premier Ministre // Asso OpenStreetMap France
- **2012: Colloque AFIGEO « Info géo et open data » à Lyon - GT OGC AFIGEO s'ouvre à l'open data**
- 2013: Lancement V2 du portail data.gouv.fr: **contribution du GT AFIGEO : CoDesign Data.gouv.fr**
- **2013 : Participation d'un représentant AFIGEO au comité d'expert Etalab**
- 2014: Nomination du 1^{er} Administrateur général des données (CDO)
- **2014: Participation Etalab aux DynReg 2014 à Ajaccio: Table-ronde « politiques géo numériques des territoires »**
- **2014: Réunion conjointe AFIGEO / Etalab - équipe data.gouv.fr**
- **2014: Réunion AFIGEO: Intervention Etalab (Simon Chignard) sur les licences d'utilisation**
- **2015: BarCamp AFIGEO/Etalab passerelle Inspire – data.gouv.fr**
- 2015 : AFIGEO participe au Projet européen ENERGIC OD (2015 – 2017)
- 2016: Loi Lemaire République Numérique : **relais appel à consultation sur la mise en œuvre du Service public de la donnée**
- **2016: Com. presse AFIGEO /Etalab la passerelle Inspire/data.gouv.fr: « Fluidifier la diffusion des données géographiques open data » et production de ressources collectives**
- **2017: Publication AFIGEO « L'info Géo et l'Open Data »**
- **2017: Réunion du GT OD AFIGEO chez Etalab**

Les récents décrets « licence » et « Service public de la donnée – SPD » structurent la Loi République Numérique. Cette réunion accueillie chez Etalab est l'occasion de faire bénéficier aux membres du GT d'informations précises et à l'équipe Etalab des retours d'expériences concrets.

1. Actualités réglementaires autour du décret « licence »

Ce décret encadre les licences utilisables par l'administration :

- Licence ouverte 2.0 par défaut
- Licence ODbL 1.0

Le choix de cette dernière licence par une administration doit être justifié car elle limite la réutilisation. Ce choix doit être fait dans un cadre d'intérêt général, proportionné et non discriminatoire. Le partage à l'identique permet d'obtenir un retour pour améliorer les données, si il n'y a pas de retours, alors son usage n'est pas justifié.

Il est aussi possible de faire homologuer une licence supplémentaire, bien que cette démarche se veuille volontairement complexe.

Deux licences ont été homologuées temporairement :

- la licence d'utilisation à titre gratuit des produits du RGE (IGN, homologation pour 12 mois)
- la licence gratuite de repartage (utilisée en plus de l'ODbL pour les données diffusées par la BAN) (IGN, homologation pour 6 mois)

Débats et échanges :

- Concernant les données dites « sensibles » non communicables : la notion de « secret par défaut » n'existe plus, et il faut justifier la sensibilité de la donnée. Les exceptions à la diffusion (code de l'environnement) sont décrites dans le guide de rédaction des métadonnées Inspire du CNIG. Ces exceptions ne doivent cependant pas empêcher leur échange entre autorités publiques. [Ces contraintes sur la diffusion ne sont pas explicités par la Licence Ouverte ; c'est le droit qui s'applique].
- Des blocages sur la diffusion des **servitudes telles que les PPRT (plan de prévention de risques technologiques)** sont constatés en local (et parfois même entre administrations). Les fondements des instructions en cours ne sont pas clairs et dépendent d'une Préfecture à l'autre. Les échanges actuels dans les travaux relatifs aux liens entre la Passerelle et GEOIDE, et dans le cadre du Géoportail de l'urbanisme traitent ce sujet, mais semblent décalés pour des données qui ont vocation à être publiées sur internet. **Il faudrait lister les exceptions et les cas abusifs dans un document de référence, afin de bâtir un argumentaire vis-à-vis des administrations.**
- Les **conventions entre administrations** n'ont plus lieu d'être dans la plupart des cas mais cette pratique se poursuit. Par ex. les conventions relatives à l'utilisation des fichiers fonciers PCI : la matrice cadastrale (qui comportent des info personnelles, un secret fiscal) ne peut pas être utilisée librement : les conventions sur les usages par les administrations restent donc d'actualité. Ce qui va changer c'est pour le PCI vecteur (voir 2.1.)
- Les API - pouvant être perçues comme un frein à l'accès aux données avec la limitation d'accès à l'ensemble de la donnée, la complexité de certains jeux de données...- constituent un confort complémentaire mais non obligatoire. Des téléchargements automatisables sont une alternatives par exemple. On constate sur ce point que les 2 démarches, open data et Inspire s'opposent sur l'emploi des API : la seconde recommande plutôt les API (WMS, WFS...) alors que la première voit les API comme un complément intéressant.

2. Le décret « Service public de la donnée – SPD »

Publié début avril, ce décret doit encore être précisé via un arrêté pour définir les conditions de performance et de mise à disposition des 9 jeux données qui bénéficient de garanties de dispo, de mises à jour, d'engagements du producteur et du diffuseur.

Echanges sur les grands principes et exceptions :

- Garantir la **disponibilité** des fichiers au-delà de 99%
- Pour être considérées comme des données de référence, elles doivent être **fraîches, avec un engagement plus fréquent sur la diffusion des mises à jour** (trimestrielle ? avec à terme une diffusion fraîche comme sur cadastre.gouv.fr).

- Proposer **des modes de diffusion** accessibles non contraints. Pour cela, le SPD rend obligatoire le téléchargement.
- Tous les référentiels identifiés aujourd’hui dans le périmètre du SPD seront **disponibles en open data** sauf le RGE. En effet, sur ce point ; le décret n’impose pas de licence spécifique au RGE qui reste payant pour les organismes non éligibles à la DCM. Le SPD n’est pas qu’un principe d’open data mais un engagement plus global. La loi VALTER précise les 3 organismes qui peuvent mettre des exceptions à la gratuité (la loi Lemaire précise le volet licence) :
 - IGN
 - SHOM
 - Météo France
- **le PCI Vecteur** devrait déjà être publié depuis avril. La licence applicable au PCI et son mode de diffusion font encore débat avec la DGFIP. En effet, un décret récent impliquant une redevance pour la mise à disposition a été publié. A terme, Jérôme Desboeufs estime que les données téléchargeables du PCI devraient être mise à jour à la même fréquence que les données consultables sur cadastre.gouv.fr. A noter qu’une mise à jour à jour aussi fréquente n’est pas toujours souhaitable pour tous les utilisateurs en raison du besoin de cohérence entre la partie graphique et la matrice cadastrale (qui ne figure pas dans le SPD). Il est donc utile de mettre à disposition une base millésimée cohérente avec la matrice en plus des données fraîches.
- Quels autres référentiels pourraient intégrés le SPD ? Comme cette liste peut être enrichie par Décret au fur et à mesure, **le GT propose d’ouvrir un fichier collaboratif pour recueillir des propositions** (Routes, monuments historiques, qui ont un ID, et gérées par des autorités...)
- Comment les diffuseurs sont-ils désignés? Les producteurs choisissent qui diffuse leur donnée : l’INSEE a confié à Etalab le soin de diffuser la base SIRENE, l’IGN diffusera ses données, le RNA - registre national des associations - est diffusé par le Ministère de l’Intérieur etc... data.gouv.fr est un point unique de recensement de tous les jeux de données SPD ([rubrique SPD](#))

Débat sur les moyens d’informations sur ces évolutions : Sur les licences et SPD, une explication pédagogique est nécessaire :

- Les messages issus de ce travail devra-t-il passer via l’AMF ? La FNCCR ?
- Via la réalisation de courtes vidéos en complément de fiches pratiques utilisables dans les territoires ?
- Il faut mettre par écrit les grands principes généraux et réfléchir à la façon dont on peut illustrer concrètement les impacts réglementaires pour le quotidien (ex. concrets sur des jeux de données emblématiques).

Article 12 de la loi Lemaire : pour une demande de communication d’une info publique, il y a une obligation de diffuser l’info, la donnée associée ainsi que leur mise à jour future. Ce principe implique mécaniquement une publication des mises à jour, cela lie offre et demande.

Sur les données qui contiennent des infos personnelles : Les données doivent être diffusées après anonymisation si le processus d’anonymisation n’est pas trop complexe ni trop coûteux. Cela n’empêche pas de demander des informations non anonymes qui nous concernent personnellement.

3. Le Socle commun de données nationales - Open Data Local

Pour répondre à cet [axe](#) de travail mené dans le cadre d'ODL (qui vise à favoriser l'accès en local à des jeux de données nationaux de façon massif) il faut disposer de référentiels sur les découpages administratifs. Un travail est réalisé par Etalab (expérimentation menée avec IdéoBFC et GéoRhones Alpes) via le projet territoire.data.gouv.fr, afin de mettre à disposition :

- un découpage territorial via une API
- une interface graphique,
- La projection d'un CSV à la volée.

Pour extraire d'un jeu de données global les données locales qui concerne un territoire il est indispensable de se mettre d'accord sur l'identification de géo zones. Les référentiels actuels ne sont pas suffisants. En effet, pour les communes par exemple, le seul code INSEE ne permet de connaître le périmètre géographique associé car un même code INSEE peut correspondre à des étendues différentes au cours des ans en raison des aléas des fusions et scissions de communes.

Une base de données [GeoHisto](#) est en cours de constitution : elle est alimentée par les fichiers historiques Insee et suit l'évolution des identifications des périmètres des territoires.

Un autre projet, [GeoZones](#), est en cours de réalisation. Il a pour but de produire des couches de données géographiques ouvertes pour les découpages administratifs (en utilisant des sources de données ouvertes comme OSM, Admin Express et GeoHisto).

GeoZones permettra de connaître l'emprise géographique à partir d'un identifiant de territoire. GeoZones est avant tout un outil et il peut s'appliquer à d'autres territoires que la France. Ainsi, en l'intégrant à Udata (le logiciel qui motorise data.gouv.fr) il est possible de reproduire cette démarche dans d'autres pays.

A terme les ID définis pour GeoHisto et utilisés par GeoZones faciliteront le découpage des données nationales à l'échelon local avec une gestion temporelle. Il serait également utile que ces identifiants soient utilisés par les plateformes Géoclip, par le Ministère de l'intérieur et la gestion de l'état civil...

Ces identifiants seront aussi utilisés dans la v2 de l'[API Géo](#) d'Etalab, prévue pour la fin de l'été.

4. Perspective pour le projet Base Adresse Nationale – BAN

Licences de la BAN

La BAN actuelle est constituée des adresses exportées de la BD UNI de l'IGN. Actuellement, la BAN est disponible sous 2 licences :

- Licence gratuite de repartage
- Licence ODbL

La base de données disponible selon la seconde licence a les spécificités suivantes :

- le libellé à la norme AFNOR et le libellé d'acheminement ne sont pas disponibles ;
- les données subissent des traitements qualité supplémentaires.

Les différences précises entre données et licences sont expliquées sur <https://adresse.data.gouv.fr/download>.

Si ces 2 licences ne conviennent pas, chacun est libre de se tourner vers les produits commerciaux proposés par les producteurs.

- une doctrine doit être rédigée pour encadrer les usages selon les licences et concédants
- L'objectif est l'enrichissement collectif de la base adresse via des remontées, mais le cas particulier c'est que les partenaires figurent parmi les exceptions de l'open data.
- La clause de repartage à l'identique sur le périmètre « France » doit être précisé.

Pour rappel, la BANO n'est pas une version de la BAN mais est une base OSM indépendante. Elle possède l'avantage pour certains usages de disposer de lieux-dits.

Évolutions de la BAN

Une nouvelle version de la BAN doit sortir en beta fin juin 2017 :

- Evolution de la plateforme qui proposera une base de données initiale produite à partir des différentes sources suivantes : la BD UNI, associées à des données exportées du référentiel adresse de la Poste, du fichier Fantoir, du code officiel géo, du PCI traité par OSM...
- Par-dessus il y aura une API, et divers outils de gestion qui permettront aux acteurs de contribuer :
 - o Guichet Adresse de La Poste et de l'IGN
 - o Un outil de contribution citoyenne développé par Etalab
 - o La possibilité de déposer des fichiers de données Adresse par les CT via le format de base adresse défini par l'[AITF](#) (courant juillet)

A noter, que le libellé AFNOR aujourd'hui disponible dans la version ODbL de la BAN ne sera sans doute plus disponible dans les futures versions. La structure de donnée actuelle de la BAN est réputée stable (à l'exception du libellé AFNOR). Des nouveaux champs pourront sans doute néanmoins être ajoutés à l'avenir.

Évolutions du site internet de la BAN

Le site <https://adresse.data.gouv.fr> est en cours d'évolution et a déjà été revu sur certains points :

- Explication synthétique des licences ODBL, etc...
- Précision sur les données disponibles.
- Traitement qualité.

Quel fonctionnement de la BAN avec les Hexaclé / Hexavia de la Poste ?

- Ce sont des données Médiapost vendues par La Poste que La Poste n'a pas l'intention de reverser dans la BAN.
- Pour ceux qui utilisent ces données, les outils de géocodage disponibles permettent de retomber sur des ID BAN. Cet usage concerne précisément des grands groupes qui commercialisent la fibre optique : il faut faire évoluer le problème d'où les échanges entre Etalab et Mission Très Haut Débit afin de proposer un ID de zone géographique (1 à 8 caractères selon la précision souhaitée), ou d'utiliser des outils cartographique au niveau commercialisation (sélection d'un point sur une carte).
- ces clés de la Poste se retrouvent dans la BD Adresse Premium de l'IGN : cela relève d'accords commerciaux indépendants des projets de la BAN et rentre dans les activités

commerciales et industrielles de l'institut.

On fait face à 7 types de position possibles pour la localisation des points Adresse : (boite au lettre, pied de porte, ...):

- Chaque adresse peut être présente plusieurs fois (une fois pour chaque type)
- Actuellement, lorsque l'on télécharge la base de données depuis le site adresse.data.gouv.fr, un seul point est exporté par adresse (choix d'usage général réalisé par l'équipe de la BAN, il s'agit prioritairement de l'entrée).
- Idéalement, un export de l'ensemble des instances des points adresse devrait être possible afin que la personne qui réalise le téléchargement puisse appliquer ses propres filtres en fonction de ses besoins [demande adressée en réunion à Jérôme Desboeufs]. Par exemple, pour les SDIS qui pourraient être intéressés en priorité par les « accès secours ». Les SDIS sont invités à manifester leur intérêt pour cette approche.

Quelle stratégie de communication sur les outils de remontée et articulation avec les PF ?

- Tout sera communiqué sur le site et via les différents canaux.
- La question de la gouvernance / de la communication autour du projet en termes de pédagogie, d'animation, sensibilisation, accompagnement sur les outils etc... semble nécessaire pour que la remontée des contributions soient effectives, cela garantira la réussite de la démarche, d'autant qu'il est rappelé qu'un important travail d'animation avaient été engagés par les plateformes et régions autour des BAN TOUR.

Il faudra clarifier et proposer un message clair (refonte prévu du site pour mieux expliciter ces points):

1. comment alimente-t-on la BAN ?
 2. comment l'utilise t-on ?
- Le **géocodeur disponible sur [adresse.data.gouv](http://adresse.data.gouv.fr)** a été amélioré avec la prise en compte de nouvelles données, recherche via code Insee des communes, intégration des X et Y légaux en plus du WGS84... Pour rappel les données géocodées sont nécessairement sous licence ODbL en sortie, sauf lorsqu'elles sont pour usage interne ou privé.

5. PIA

Rapide point d'info, et échange sur la nécessaire évaluation des PIA deux ans après.

L'analyse des dossiers de cette dernière vague de PIA open data sera rigoureuse.

- 1. Analyse technique
- 2. Délibération du jury

AFIGEO se fait le relais d'une forte demande de ses membres dont le réseau des CRIGEs de pouvoir être informé au plus tôt. Etalab n'a pas de visibilité sur le calendrier de lancement des appels à projet.

6. Open Data Locale / OpenDataFrance et socle commun de donnée

Des membres AFIGEO ont participé à la restitution de l'avancement des expérimentations [OpenData Locale \(journée du 27 avril\)](#). Des axes de travail ont été identifiés notamment sur le projet impulsé par Thomas et l'axe de travail "Constitution de modèles minimaux de données" - GT éditeurs d'ODL.

Christian précise que la démarche (présentée le 27/04 dans l'atelier « socle commun local ») vise à disposer d'un minimum de modèle commun sur les données locales pour faciliter leur agrégations à un échelon supérieur... sans aller dans la très grande richesse métier des données (ex : borne de recharge de véhicules électrique, établissement recevant du public, place de parking handicapé, hydrants, etc).

Travail lancé sur les marchés publics, les « bâtiments publics ».

Etalab travaille sur des outils pour faciliter la détection de la conformité ou non d'un modèle de donnée, et voir s'il peut être rattaché à d'autres modèles de données.

Les échanges entre ODF et AFIGEO dévoilent :

- Intérêt ODF pour travailler collectivement sur ces sujets : identifier les données à ouvrir en priorité, puis leur structuration et aussi sur le volet « sensibilisation » : proposer des fiches pratiques afin de contribuer au kit pédagogique
- Le GT SIG / TOPO de l'AITF a travaillé sur les Adresse, il serait intéressant de les associer. Avec le CNIG, la COVADIS,... le jeu d'acteurs est complexe.
- Le groupe de travail « éditeurs » d'ODL travaille sur ces modèles de données, pas toujours de lien avec les expérimentations locales : quel emboîtement de ces actions ? d'autre part, le temps des éditeurs, du niveau national, local n'est pas le même : d'où la complexité en matière de coordination de cette démarche.
- Sur le **socle commun des données** les choses sont plus complexes : interactions entre usages, techniques, et juridiques. Le nombre restreint de données permet de tester petit à petit sur des jeux de données territoire par territoire
- Jérôme Boutet alerte le GT sur le fait qu'ODF a planifié l'ensemble de son travail en 2017 et qu'il serait nécessaire de s'y coller. Les appels à commentaires sur les modèles de données sont prévus sur une quinzaine de jour, ce qui semble trop court.

L'expérimentation permet de tester des outils, des standards et ainsi d'évaluer leur intérêt.

Thomas souhaite que le GT se mette d'accord sur le lancement de la **définition de modèles de données** à partir de la méthodologie de travail et solliciter les différents réseaux et identifier les territoires.

Sur le cas des HYDRANTS :

- difficulté pour identifier les bons partenaires : AITF ? SDIS ? ODL ? Plateforme régionale ?
- difficulté sur le positionnement de l'AFIGEO vis-à-vis d'entités administratives
- le lien avec les éditeurs de logiciel peut poser question...
- il serait opportun d'intégrer cette démarche dans l'expérimentation ODL

Cependant le GT acte le lancement d'un appel à manifestation...

- les usages / clients : besoin d'interopérabilité,
- les producteurs : type collectivité, SDIS 33, SDIS 59, GIP ATGERI, réseau géomaticiens des SDIS

7. Catalogage et INSPIRE

Benjamin Chartier fait état des évolutions des recommandations techniques européennes de renseignement des métadonnées qui peuvent potentiellement impacter le guide de rédaction de métadonnées du CNIG (contraintes légales, licences...). Une réunion de la commission RMO aura probablement lieu en octobre / novembre 2017.

Voir les ressources :

- Article publié en décembre 2016 : <http://inspire.ec.europa.eu/news/new-inspire-technical-guidance-published-metadata-coverage-and-observation-data>
- Dernière version des recommandations :
<http://inspire.ec.europa.eu/id/document/tg/metadata-iso19139>

8. Calendrier

Pas de date définie mais il est proposé de tenir des réunions régulières dédiée à un sujet en particulier.