



## **COFIL BAN du 30 septembre 2015**

*(merci d'indiquer via l'icône en haut à droite, votre nom pour suivre les contributions de chacun sur ce pad)*

### **Heure et lieu**

10h30, 86 allée de Bercy, bâtiment Sully, salle 319V

### **Ordre du jour**

- Point sur les données: quantité, qualité
- Présentation des tests effectués en Gironde (La Poste), étude qualité sur l'agglomération de Poitiers (par IGN) , 800 000 adresses non appariées: repérage des causes et mesures à prendre
- Intégration des remontées de crowd-sourcing
  
- Proposition d'un séminaire de développeurs et utilisateurs (présentation réciproques des architectures actuelles, prototypage de la base pivot :
  - Réflexion sur une base de données "pivot" permettant d'assurer à la fois les appariements et de dialoguer, via des API, avec les SI des partenaires
  - Fourniture des codes source des travaux IGN sur la BAN et sur le guichet (en lien avec La Poste)
  
- Mise en place d'un gestionnaire de tickets
  
- Logo BAN et dépôt de marque ? (suite comité éditorial)
  
- Position et rôle de la DGFIP

## Présents

- Président: Henri Verdier
- DINSIC/Etalab: Hélène Brisset, Christian Quest, Laure Lucchesi ~~Guillaume Blot~~,
- OSM France: Vincent de Chateau-Thierry
- IGN: François Becirspahic, Olivier Dorie, Stéphane Chataigner, Frédéric Cantat, Michel Ségard
- LA POSTE : Emmanuel HOUDINET, Pascal BARTIER, Christel PAPILLON VIOLETT

## Invités

- Mission THD: Mme Houda Khammari et M. Bertrand Vandeputte
- AMF: Benoit Liénard

## Compte rendu

- **Point sur les données: quantité, qualité**

Etalab applique des scripts de contrôle de qualité sur les livraisons hebdomadaire provenant de l'IGN. Ces contrôles sont actuellement uniquement sémantiques, les scripts sont partagés sur github, les résultats disponibles sur owncloud.

Sur la dernière livraison, 8.9 millions d'adresses comportent au moins une anomalie ce qui représente 34,8 % des adresses.

Au total 16 millions d'anomalies sont détectées, d'une gravité variable.

Un focus est ensuite fait sur l'anomalie des appariements des données Postes dans la BAN, qui approchent les 900.000 cas sur la métropole et 1.3 million en incluant les DOM.

Ces écarts proviennent des adresses non encore géolocalisées (nouvelles adresses) ainsi que d'un défaut d'appariement.

(voir tableau de suivi en annexe)

Test la Poste / IGN effectué sur le département de la Gironde (33) sur la sémantique

Méthode RNVP de la Poste (taux de rapprochement : 93 %) ; méthode IGN (95 %)

22 % adresses BAN non rapprochées sur Hexacle par la Poste

7 % adresses BAN non rapprochées par l'IGN (un correctif dans les chaînes de traitement interne IGN, livré fin de semaine, devrait résoudre en grande partie cet écart)

Enjeux sur la prise en compte des évolutions DGFIP et la synchronisation DGFIP / IGN (voir point à l'ordre du jour) : le cycle d'intégration des données DGFIP est actuellement annuel, celui des données Poste semestriel.

Conclusion du test : fixer des règles communes entre Etalab, la Poste et l'IGN - Enjeux sur l'unicité de la base et la augmentation de la fréquence d'intégration des données et de synchronisation/appariement.

(voir présentation en annexe)

Test géométrique fait sur Poitiers mené par l'équipe IGN "contrôle Qualité"

+3 500 points contrôlés sur un total de +22 000, sur IGN, base de l'agglomération de Poitiers et la BANO. Aucun référentiel, selon les cas, ne détient de façon systématique le meilleur positionnement. Démonstre le caractère hétérogène des plans d'adressage.

L'IGN précise que cette étude indique un résultat valable pour une zone donnée et ne peut pas être généralisé du fait de la grande diversité de cas entre les territoires.

OSM-France fait remarquer qu'il est dommage d'avoir utilisé des données anciennes (septembre 2014) pour le comparatif BANO alors que l'étude a été faite en juillet 2015.

(voir présentation et rapport complet en annexe)

A la suite de ces présentations, un consensus émerge sur les points suivants :

- la forte complémentarité des connaissances et des données des uns et des autres pour optimiser l'ensemble (personne seul n'a la vérité), associé à un besoin fort de travailler sur des règles d'appariement
- le besoin partagé de partir d'une nouvelle base, et pas d'une base existante pour fiabiliser les données. Faire reposer la BAN sur les bases métiers actuelles (BDUni ou RNA) a mené à une impasse, celles-ci ayant des contraintes de production quotidiennes trop fortes pour évoluer à court terme.
- le rôle clé de la BAN à cet effet et le fait qu'elle doit devenir le point d'appui pour de nombreuses API métier.
- les SI actuels doivent être adaptés à l'accélération nécessaire des cycles de mise à jour des données et à l'intégration des données tierces.

- **Proposition d'un séminaire de développeur (présentation réciproques des architectures actuelles, prototypage de la base pivot :**

Réflexion sur une base de données "pivot" permettant d'assurer à la fois les appariements et de dialoguer, via des API, avec les SI des partenaires. Cette base est le « pot commun » de la BAN.

L'objectif du séminaire est de converger rapidement sur la construction d'une base commune, partagée, qui ne soit pas assujettie aux contraintes des uns ou des autres et qui pour se développer nécessite un espace de travail et de test commun et collaboratif.

**Proposition de décision :** organiser un séminaire début novembre, avec des utilisateurs (en s'appuyant, si possible, sur des instances représentatives)

La Poste propose de l'organiser, à Libourne (33), première semaine de novembre (entre le 2 et le 6 novembre). L'équipe technique des fondateurs est chargée de le préparer, de façon utile et intense. Si possible sur deux jours, l'un avec des utilisateurs et/ou contributeurs (DGFIP, INSEE, etc.), l'autre sans les utilisateurs/contributeurs.

Henri Verdier fait part de son inquiétude sur l'agilité nécessaire pour coller aux attentes des utilisateurs, en particulier les services de l'État.

Témoignage de la mission THD sur le groupe Interop constitué par les opérateurs et les organismes intervenant sur le terrain pour déployer le réseau fibre optique (besoin métier : référence exhaustive et à jour - opportunité pour améliorer la BAN). La mission indique l'urgence du besoin de ce référentiel.

L'AMF rappelle aussi la nécessité de s'adresser à des représentants des collectivités pour faire l'interface vu la diversité des collectivités plutôt que de multiples contacts ponctuels.

- **Intégration des remontées de crowd-sourcing**

Le guichet Citoyen n'est pas encore activé mais une première version transmettant les signalements par email vers La Poste est disponible. Demande transmise par Etalab à Identité numérique (SSO La Poste) non encore instruite.

Les signalements sont aussi récupérés actuellement par OSM-FR sur son outil d'analyse qualité.

L'IGN de son côté est en train d'étudier l'intégration et le traitement possible sur ces signalements.

- **Fourniture des codes source des travaux IGN sur la BAN et sur le guichet (en lien avec La Poste)**

L'IGN a mis à disposition les codes source des travaux sur la BAN et le guichet (FTP pour le moment, références à transmettre aux participants) -> à mettre sur le owncloud partagé comme tout le reste est déjà partagé là actuellement

- **Sujets pour le séminaire:**

- déploiement d'un outil de développement collaboratif (gitlab ou équivalent) pour partage de code, gestion des tickets, etc
- plateforme de développement et de test à mettre en place
- appariement de données...
- modèle de données interne et externe, modèle simple d'échange
- méthodes et outils pour travailler en collaboratif
- définition du périmètre, des API
- le guichet mairie est à traiter séparément et n'est pas l'objet du séminaire

2 jours: journée mixte (développeurs/utilisateurs/contributeurs) + une journée tech (dev)

**Prochain Copil BAN** : le mercredi 25 novembre à 10h00 - 12h00 (Mirabeau)