

PROJET BAN COMPTE-RENDU

Objet de la réunion :

Journée BAN Team à Libourne
IGN / La Poste / Etalab

Date 5/11/2015

Participants :

ETALAB : Christian, Yohann.

IGN : Stéphane, Olivier, Rémi, Philippe.

LA POSTE : Isabelle, Arnaud, Christophe, Philippe, Brice, Jérôme C, Didier.

Ordre du jour :

Inventaire de l'existant

Partage sur le code existant pour une prise en main réciproque

Définition de l'organisation, des méthodes pour travailler en collaboration à distance (choix et mise en place d'outils)

Principales conclusions :

- L'architecture cible, avec la BAN en tant que base pivot accessible uniquement via une API, est validée.
- Une quinzaine de lots sont identifiés pour aboutir à cette nouvelle architecture. Les lots API, modélisation ont été décrits plus précisément en séance.
- Les équipes d'EtaLab, LP, IGN vont commencer à co-développer sur le lot API et modélisation. Une journée de prise en main du proto d'API écrit par EtaLab est prévue à Paris le 20/11. Les modèles de données seront échangés sous la forme de bases PostgreSQL. Les codes des prototypes seront mis sur GitHub.
- Sur l'aspect modélisation des données, LP et l'IGN commencent un document de

spécification pour préciser le contenu de la base.

- LP et l'IGN commencent la rédaction du plan directeur du projet qui décrira plus précisément les différents lots du projet, leur enchainement, les étapes clés, les différents livrables ...
- Des points réguliers d'avancement sont à prévoir entre les chefs de projet côté LP, EtaLab, OSM et IGN.
- Calendrier à court terme : une première version prototype de l'API et du modèle BAN sera disponible mi-janvier. Les données seront montées en base dans ce modèle sur un ou 2 départements en utilisant l'API.
- Calendrier à plus long terme : celui-ci sera décrit dans le plan directeur avec l'enchainement des différents lots et les grandes étapes. La première étape est mi-janvier avec un prototype d'API et de modèle. Les étapes suivantes seront l'affinement du module d'initialisation pour obtenir une BAN France entière. Au niveau ouverture à l'extérieur, il y aura 2 grandes étapes dont les dates restent à préciser :
 - o L'ouverture d'une BAN en lecture seule
 - o L'ouverture de la BAN en écriture. Cette ouverture dépend en partie des lots synchronisation SI-IGN et SI-LP et devra être analysée plus finement quant à ses conséquences
- Un OpenLab est prévu le 15 décembre sur Paris avec les utilisateurs pour leur faire part de notre avancement

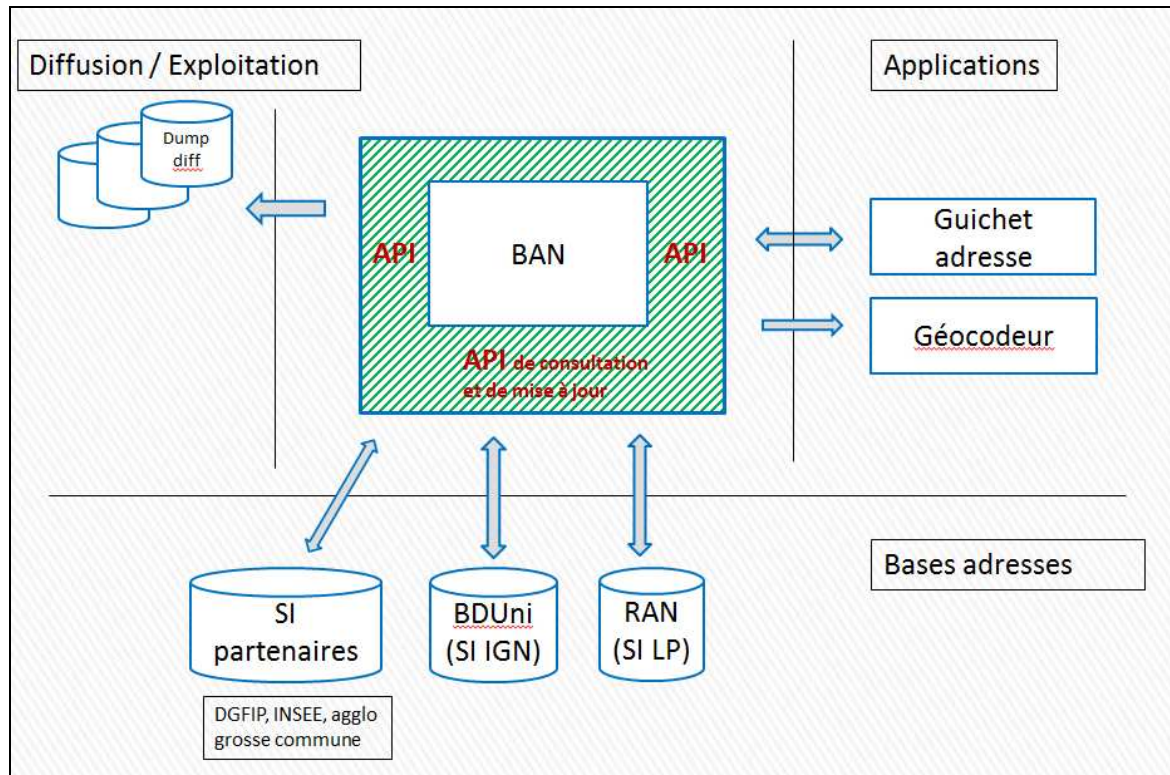
Architecture cible :

Les différents participants s'accordent sur l'architecture cible suivante :

- La BAN est la base pivot qui alimente et est alimentée par différentes bases adresses métiers/partenaires :
 - o Bases adresses des fondateurs (RAN pour La Poste, BDUni de l'IGN)
 - o Bases adresses des autres institutions (INSEE, DGFIP ...)
 - o Bases adresses des collectivités (SDIS, grandes communes ...)
 - o ...
- Sur la BAN viennent se brancher diverses applications de mises à jour/ consultation / service :
 - o Le guichet adresse pour les mairies et autres partenaires
 - o Le géocodeur

○ ...

- L'accès à la BAN se fait exclusivement par l'API.



Les différents lots du projet :

Les lots suivants sont identifiés en séance :

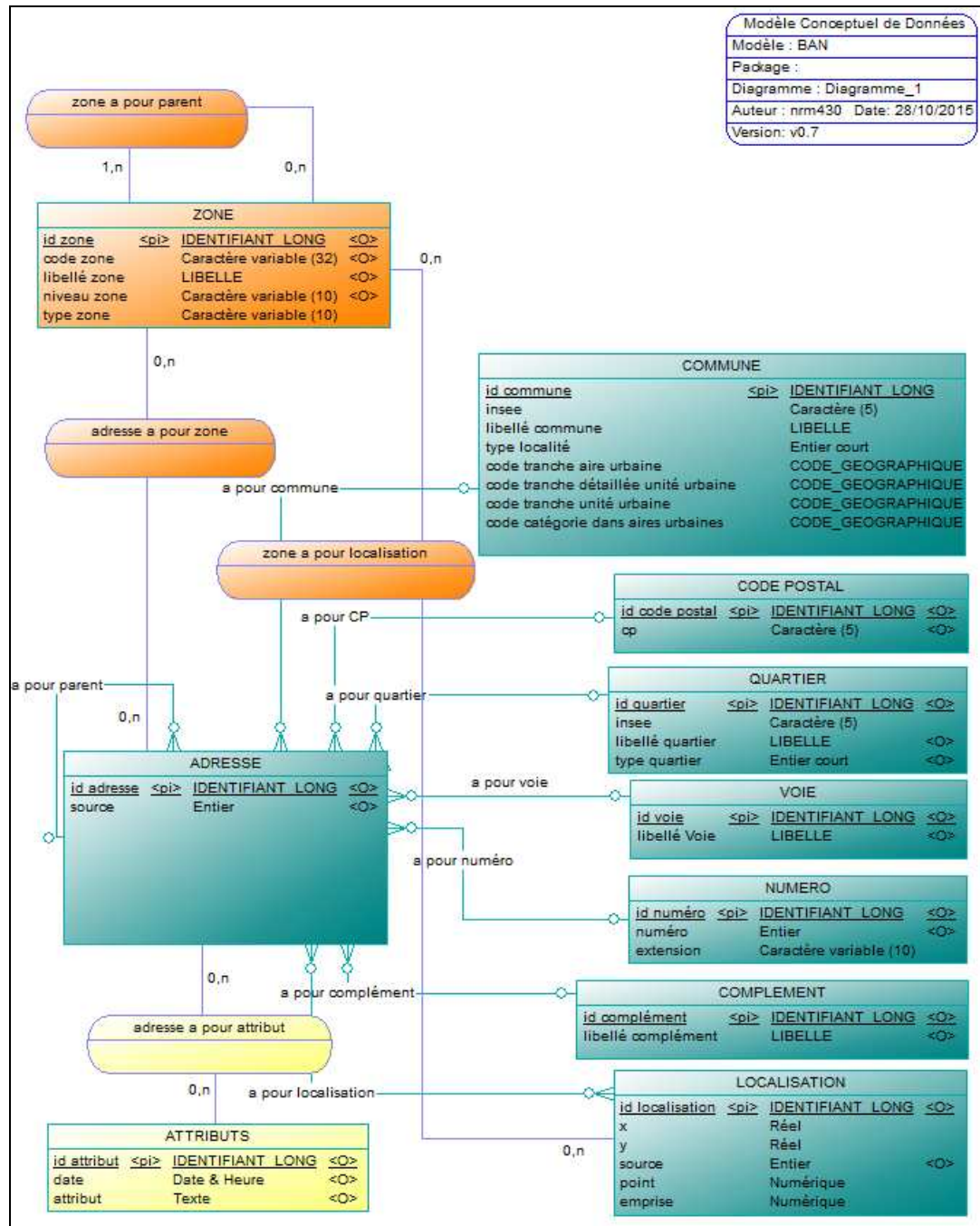
- API
- Modélisation
- Branchement du guichet adresse sur l'API
- Convergence SI IGN
- Convergence SI La Poste
- Initialisation / Chargement initial (IGN, Poste, RIL, DGFIP) + appariement
- Authentification / autorisation
- Infrastructure / hébergement de la BAN
- Extraction / diffusion : mise à disposition régulière des exports complets, par zone ou différentiel de la BAN

- Contrôles : mise en évidence des anomalies dans la BAN (contrôle serveur). Contrôle des données en entrée de la BAN (contrôle au niveau de l'API)
- Tableau de bord / monitoring
- Services : géocodeur, certification d'adresses, normalisation d'adresses, appariement
- Exploitation : outils d'administration pour lancer les imports massifs ou traiter les fusions de communes par exemple
- Synchronisation transitoire : pour synchroniser les nouveautés du RAN dans la BAN (tant que les mises à jour sont effectuées dans le RAN et pas dans la BAN). Le même processus est envisageable pour le SI IGN.

Les lots suivants sont discutés plus en détail

Lot modélisation :

Le SNA propose le modèle suivant pour la BAN :

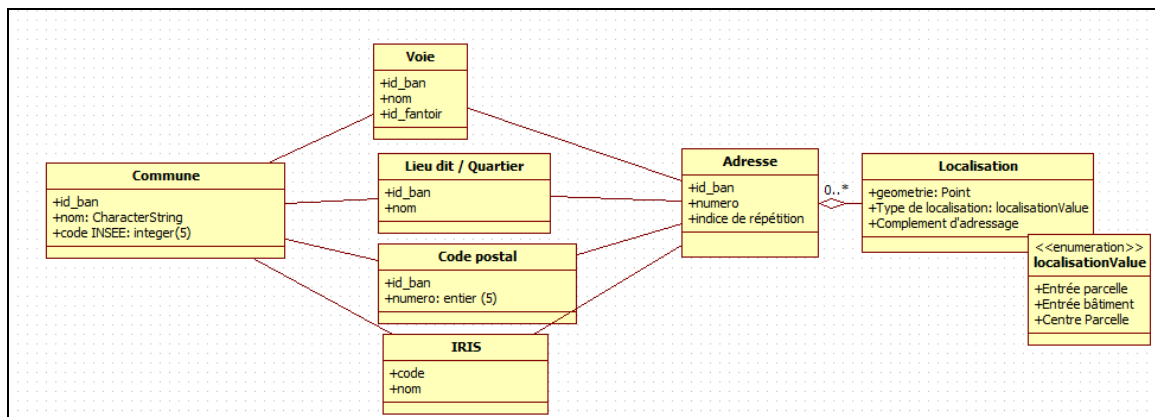


Les points suivants sont discutés :

- Dans le modèle du SNA, l'objet « Voie » est en fait une ligne dans un dictionnaire. C'est-à-dire que toutes les avenues Pasteur de France sont représentées par la même entité. Pour EtaLab et l'IGN, une voie est un objet à part entière et dans le cas de l'avenue Pasteur, il faudrait une entité par commune. Ces entités pourraient être reliées à un dictionnaire cependant ...

- Les points suivants sont à préciser :
 - o La notion de Quartier/ Lieux-dits
 - o Le complément d'adressage
 - o La gestion des métadonnées et des sources.

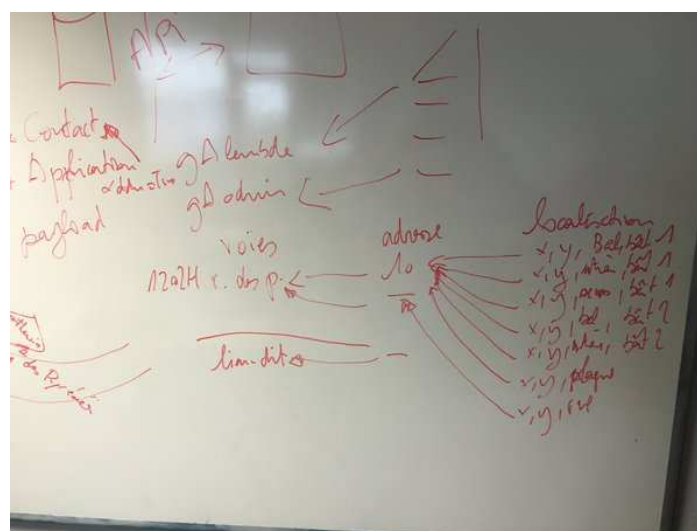
Le modèle suivant est proposé par C. Quest en séance :



Les attributs sont donnés à titre indicatif (il en manque un certain nombre).

Les classes n'ont pas de géométries sauf l'adresse qui peut avoir 0 ou n localisations géométriques. Chaque localisation géométrique est accompagnée du type de localisation (par exemple « Entrée de la parcelle » ...) et éventuellement d'un complément d'adressage.

Ce modèle est soumis à quelques exemples d'adressage. Dans le cas ci-dessous, pour un numéro particulier, on se retrouverait avec une adresse et 8 localisations.



Des questions se posent sur la signification du complément d'adressage.

L'IGN se pose aussi la question de la pertinence du stockage vers les identifiants des parcelles du cadastre.

Des questions sont aussi posées sur le périmètre de la BAN :

- Question des données opendata /hors opendata
- S'assurer que le modèle est compatible avec l'offre

Le modèle devra aussi gérer des métadonnées et l'historisation.

Pour progresser sur ce lot, il est proposé que :

- Nous échangions nos propositions de modèle physique (dump de schéma PostgreSQL)
- Nous essayions de charger nos données dans ce modèle
- Nous commençons de recenser « les cas tordus » d'adressage pour les soumettre au modèle.
- LP et l'IGN commencent un document de spécification et le soumettent au groupe pour commentaire/remarque jusqu'à convergence.

Lot API :

EtaLab a commencé à prototyper une API en python (avec les frameworks falcon et django).

Une démonstration est faite en séance. Les fonctionnalités suivantes existent :

- Lecture : récupération des objets par identifiant (sur une commune / voie / adresse)
- Ecriture : il est possible d'ajouter les objets un par un en passant tous leurs champs.

Il est possible de récupérer les différentes versions d'un même objet.

Le prototype gère aussi les conflits (quand 2 personnes mettent à jour la même version du même objet).

Il n'y a pas de récupération par boîte englobante. L'authentification et les autorisations ne sont pas gérées actuellement.

EtaLab propose que :

- l'API gère l'authentification et les autorisations par applicatif et pas pour chaque utilisateur de chaque applicatif.
- Au moment de son appel à l'API, l'applicatif pourrait fournir des infos supplémentaires sur l'utilisateur (de l'applicatif) à stocker dans les métadonnées de la BAN

- L'API gère les niveaux d'autorisation suivants: admin, écriture sur les voies, écriture sur les adresses. Par défaut, l'accès en lecture seule serait autorisé.
- Les accès par zonages géographiques ne seraient pas gérés par l'API mais par l'applicatif appelant l'API (par exemple le guichet adresse)

Pour progresser sur ce lot, il est proposé que :

- EtaLab continue à prototyper l'API. LP et l'IGN peuvent aussi prototyper de leur côté.
- LP et l'IGN prennent en main cette API et commencent à l'utiliser en chargeant des bases tests (réunion de prise en main et d'échange à Paris le 20/11)
- Les codes seront mis sur github et coadministrés par EtaLab, LP, IGN

Lot Import des données / appariement

Quelques discussions ont porté sur ce lot :

- Faut-il mettre de l'appariement dans l'API ? → Oui, mais un minimum
- Il est proposé de faire l'appariement en amont puis de transmettre les données « propres » à l'API

Choix des outils :

- Base de données : PostgreSQL. Lors de la phase de prototypage, chaque partenaire chargera une instance test de la BAN sur une petite zone dans son organisme.
- Outils collaboratif de développement : les prototypes seront mis et échangés sur github. Chaque partenaire pourra commiter et le github sera coadministré.
- Gestion de projet, suivi des tâches et avancement : outil collaboratif trello
- Echange de document : les documents seront mis sur owncloud. Par contre, il faudrait trouver un moyen de faire de l'édition collaborative notamment pour les documents sur le modèle et le plan directeur.

Organisation du travail :

- LP et l'IGN commencent à rédiger le plan directeur du projet et le soumettent au groupe pour commentaires/remarques jusqu'à convergence. Il contiendra notamment les différents lots du projet, leur enchainement, les étapes clés, les différents livrables ...
- Des réunions régulières seront organisées entre les différentes équipes et les « chefs de projet » LB, EtaLab et LP pour organiser le travail et faire remonter les avancées/difficultés aux comités de suivi/pilotage.