Réalisation 2024

# La Base d’Adresse Nationale

Contexte :

Objectif : Utiliser la BAN comme référentiel pour identifier l’adresse d’un point de collecte ou livraison

## Importation et suivi de la Base d’adresse Nationale

### L’équipe :

* Développement des outils  : Alexander Dalziel
* Réalisation et analyse des importations : Alexander Dalziel, Stéphane Millot, El Moundir Faraoun
* Rédacteur : Stéphane Millot

### Le planning

### Les travaux réalisés

**Rappel :** Nous utilisons les fichiers CSV mis à disposition sur le site [Données nationales | adresse.data.gouv.fr](https://adresse.data.gouv.fr/donnees-nationales) pour les importer dans une table SQL avec l’intégralité des colonnes mise à disposition.

**Lors d’une mise à jour :**

* soit l’identifiant dans la BAN (id) est retrouvé et on met à jour la fiche à partir des informations (avec un éventuel réveil si la fiche était en sommeil)
* soit l’identifiant n’est pas trouvé et on insert la fiche
* et on met en sommeil les fiches non présentes dans le nouveau fichier mis à disposition

**Identification d’une nouvelle adresse annulant et remplaçant une adresse**

Certaines adresses mises en sommeil sont remplacées par d’autres avec un nouvel identifiant.

*Exemple :*

|  |  |  |  |
| --- | --- | --- | --- |
| Ancienne ID BAN | Nouvelle ID BAN | Ancienne Adresse | Nouvelle adresse |
| 89086\_3xnnc6\_00003 | 89086\_hf25o3\_00003 | 3 RUE DE LA CROIX SAINT PIERRE | 3 RUE DES MARNES |
| 89046\_h9b9hu\_00143 | 89046\_0060\_00143 | 143 RUELLE DU CHATEAU | 143 RUE DU CHATEAU |

Pour les relier :

Identification des adresses actives ayant les mêmes coordonnées GPS qu’une adresse en sommeil

### Suivi de la base d’adresse nationale

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date d’importation | Nb fiches actives | Nb fiches en sommeil | Nb fiches totales | Nb ligne dans le fichier CSV |
| 07/04/2023 | 68 749 | 0 | 68 749 | 210 022 |
| 23/04/2024 | 206 144 | 2 224 | 208 368 | 206 144 |
| 06/06/2024 | 205 787 | 11 169 | 216 956 | 205 787 |
| 27/06/2024 | 205 347 | 12 052 | 217 399 | 205 347 |
| 01/07/2024 | 205 133 | 12 391 | 217 524 | 205 133 |
| 03/07/2024 | 205 134 | 12 405 | 217 539 | 205 134 |
|  |  |  |  |  |

### Les problèmes rencontrés

Les adresses étant en cours de validation nous constatons les problèmes suivants :

* La BAN est incomplète car des bâtis n’ont pas d’adresse

Exemple : la rue gustave cotteau à Auxerre au 03/07/2024 n’est pas référencée



Les écuries du bas four 89310 Fresnes n’ont pas d’adresse



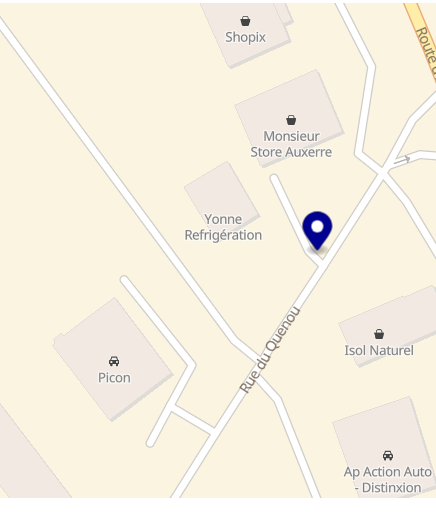
* La BAN n’est pas précise au moins sur les adresses non certifiées

Exemple : Garage Roizot 12 rue du quenou 89380 Appoigny

Selon google :



Selon la BAN au 04/07/2024

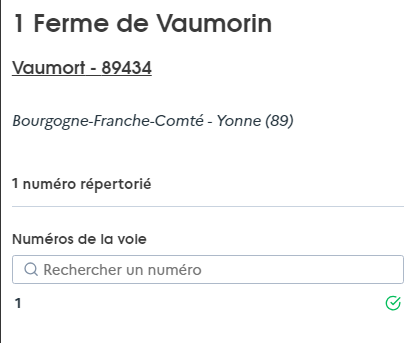


* La validation des adresses par les communes entraine des problèmes de mises à jour car :
* Une adresse peut disparaitre dans un certain laps de temps. Des adresses avec des bâtis (dans l'exemple plusieurs immeubles) sont supprimées sans nouvelles adresses sur une commune qui annonce 100% de réalisation de la BAN.

exemple : ID BAN : 89418\_zkk3od\_00015 à TONNERRE (89700) 15 rue des vals profondes qui n'existe plus Au moins 4 immeubles sont sans adresses



* Des noms de voies sont incorrects car le numéro a été mis dans le nom de la voie :



* Nous n’avons pas trouvé comment relier la nouvelle adresse avec l’ancienne car les id sont différentes.

L’identification est réalisée par l’ajout d’une comparaison sur les coordonnées gps, les informations textuelles (lieu-dit, …)

### Travaux à réaliser

Ajouter un champ dans la table BAN pour relier une adresse mise en sommeil avec une adresse active en se basant sur des stricts égalités sur l’adresse textuelle puis pour celle non trouvées sur l’adresses exactes sur les coordonnées GPS.

Ré entrainer le modèle pour qu’il apprenne à reconnaitre l’ancienne adresse et la nouvelle.

## Les destinataires et l’identification dans la Base d’Adresse Nationale

### L’équipe

* Développement des outils  : Alexander Dalziel
* Réalisation et analyse des importations : Alexander Dalziel, Stéphane Millot, El Moundir Faraoun
* Rédacteur : Stéphane Millot

### Le planning

### L’algorithme

L’identification de la voie de la BAN à partir de l’adresse du destinataire nécessite :

* que l’adresse du destinataire soit suffisamment précise (présence du nom de la voie, du numéro, commune correcte,…)
* qu’elle existe dans la BAN

### La réaffectation des id ban en sommeil sur un id ban actif

Voir la requête d’Alex pour la décrire

### Les résultats

Les résultats n’étant pas satisfaisant nous souhaitons reprendre les travaux sur la détection d’entité nommé pour améliorer le modèle :

* Apporter plus de contexte au modèle
* Améliorer l’entrainement
* Identifier le meilleur prétendant ou l’impossibilité de le trouver

# La détection d’entité nommé

## L’équipe :

## Développement des outils :

* Réalisation et analyse du modèle NER : El moundir FARAOUN

## Le planning

## Les travaux réalisés

Nous avons repris les travaux de l’expérimentation préliminaire de l’année dernière pour améliorer le modèle.

L’objectif est de pouvoir l’utiliser pour identifier le numéro qui n’est pas connu dans l’adresse référencé en cherchant si des numéros sont présents dans les alias.

Nous avons repris les travaux de l’expérimentation préliminaire de l’année dernière concernant la détection d’entités nommées (NER) pour :

* Améliorer le modèle de NER.
* Réfléchir sur comment une détection d’entités postales dans une adresse erronée pourrait être intégrée dans le processus global d’identification des destinataires.

Après réflexion, nous avons conclu que le modèle NER serait plus bénéfique en l’intégrant dans un module de correction et de complétion de la base d’adresses des destinataires de référence en Offline. L’objectif est de pouvoir récupérer les attributs manquants dans une adresse destinataire correcte via la détection nommée de ces attributs dans les adresses alias associées. Par exemple, un numéro manquant dans l’adresse de référence pourra être détecté dans une des adresses alias et récupéré pour compléter l’adresse de référence.

Cette complétion nous aidera d’autant plus dans l’identification des adresses BAN de certains destinataires qui jusqu’à présent été impossible à cause du manque d’attributs dans les adresses correctes des destinataires.

Nous croyons que cela améliorait la qualité d’identification globale des destinataires dans les points suivants :

* Une base d’adresses complète et correcte des destinataires nous permettra de corriger la précision du géocodage de certains destinataires ce qui améliorai l’affectation correcte des point de livraisons.
* Le dédoublonnage de certains destinataires via la comparaison de leurs adresses complétées.

Etapes de réalisation du modèle :

* Mettre au point une typologie d’erreurs réaliste qui servira dans la génération d’adresses erronées.
* Etiquetage des adresses générées par des tags d’attributs et tags d’erreurs ou bruits ajoutés.
* Finetuning d’un modèle Transformers pour la tâche de NER.

## Les résultats

# Le modèle d’identification du destinataire

## L’équipe :

## Le planning

## Les travaux réalisés

## Les résultats

# Le choix du meilleur prétendant

## L’équipe :

## Le planning

## Les travaux réalisés

## Les résultats

# Dé doublonnage et contrôle de doublon lors de la création

# Décafouillage

# Test et suivi des améliorations

## L’équipe :

## Le planning

## Les travaux réalisés

## Les résultats

A résoudre

