# Modification base de donnée projet IA et destinataire

## Données entreprise destinataire

Il faudrait pouvoir consolider les données relatives aux entreprises destinataire. L’une des principales donnée à récupérer et le SIRET de l’établissement destinataire. Cela permettra de s’avoir si le destinataire est toujours en activité ou non, permettant de le mettre en sommeil dans la base de données.

Ajout de destinataire.SIRET (champs visible uniquement si destinataire.type = 1) ou pour recueillir plus d’information une table destinataireEntreprise (avec pour colonnes les différentes informations concernant l’établissement et une clé destinataire)

Pour récupérer les informations il est envisageable d’utiliser l’API du gouvernement : <https://recherche-entreprises.api.gouv.fr/>

(Documentation : [ReDoc (api.gouv.fr)](https://recherche-entreprises.api.gouv.fr/docs/))

Depuis l’api 2 type de recherche :

* Recherche textuelle : Cet endpoint permet de récupérer les unités légales et les établissements correspondantes à la recherche textuelle effectuée sur la dénomination, l'adresse, les dirigeants et les élus.
* Recherche géographique : Cet endpoint prend en paramètre une latitude (:lat) et une longitude (:long) et renvoie les unités légales et leurs établissements autour de ces coordonnées. Vous pouvez également préciser un paramètre radius en km(défaut: 5 km).

Il est donc possible de croiser les données des deux requêtes pour être sur d’identifier le bon établissement.

### Exemple de réponse



Faire une table SIRET

* Cle
* Cledestinataire
* SIRET
* CodeAPE (activite\_principale)
* DateCreationEtablissement (date\_creation)
* DateFermetureEtablissement (date\_fermeture)
* GPSX
* GPXY
* Numerovoie
* Typevoie
* Libellevoie
* Codepostal
* Ville

Table SIRENEDIRIGEANTS

* cle
* cleSIREN
* nom
* prenom
* qualité
* typedirigeant

## Labélisation des erreurs

Pour entrainer l’IA dans l’identification et la correction des erreurs depuis les requêtes EDI (Nom dest + Adresse), nous devons labéliser manuellement les types d’erreur que nous rencontrons.

Pour cela nous pouvons envisager une table TypeErreur (cle, typeerreur en tinyint) contenant :

|  |  |
| --- | --- |
| 1 | Adresse imprécise |
| 2 | Création |
| 3 | Test |

Ainsi qu’une table Erreur (cle, cleType, sstypeerreur en tinyint, erreur varchar(max)) contenant :

|  |  |  |
| --- | --- | --- |
| sstypeerreur | Libelle | Cletype |
| 1 | Zone (ZI ZA ZAC …) | 1 |
| 2 | Mauvais numéro | 1 |
| 3 | Mauvais type de voie | 1 |
| 4 | mauvais nom de voie | 1 |
| 5 | Commune regroupement de commune | 1 |
| 6 | Commune approximative | 1 |
| 7 | Commune inattendue | 1 |
| 8 | Numéro inexistant dans le BAN | 1 |
| 9 | Nom d'un dirigeant | 1 |
| 10 | Nom d'un salarié | 1 |
| 11 | Utilisation du sigle sncf … | 1 |
| 12 | Adresse mal ordonée | 1 |
| 13 | Adresse différente | 1 |
| 14 | Création | 2 |
| 15 | Entreprise dans une entreprise (entrepôt logistique, …) | 1 |
| 16 | Intervention d'une entreprise à l'adresse de son client | 1 |
| 17 | Test EDI | 3 |

Et pour lié les erreurs aux livraisons une table LivraisonErreur (cle, cleLivraison, cleErreur)

## Encodage vectorielle des données

Il faudrait ajouter un champs ‘Encode’ en DateTime au table destinataire et destAlias pour pouvoir encoder les données en temps réelle.  Il serait même intéressant de rajouter un champs ‘DateModif’ en DateTime et si Encode – DateModif < 0 de nouveau Encode le destinataire. Pour alimenter le champs DateModif il serait envisageable de créer un Trigger.