



**PREMIER
MINISTRE**

Liberté

Égalité

Fraternité

Pseudonymisation des décisions au Conseil d'État

Direction interministérielle du numérique

Plan

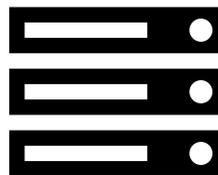
1. Les enjeux pour les juridictions administratives
2. La démarche générale
3. Les spécificités du projet au Conseil d'État
4. Contexte Etalab
5. Réalisations
6. Prochaines étapes

Les enjeux pour les juridictions administratives

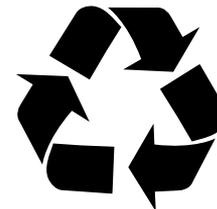
Améliorer les statistiques de pseudonymisation de l'outil actuel



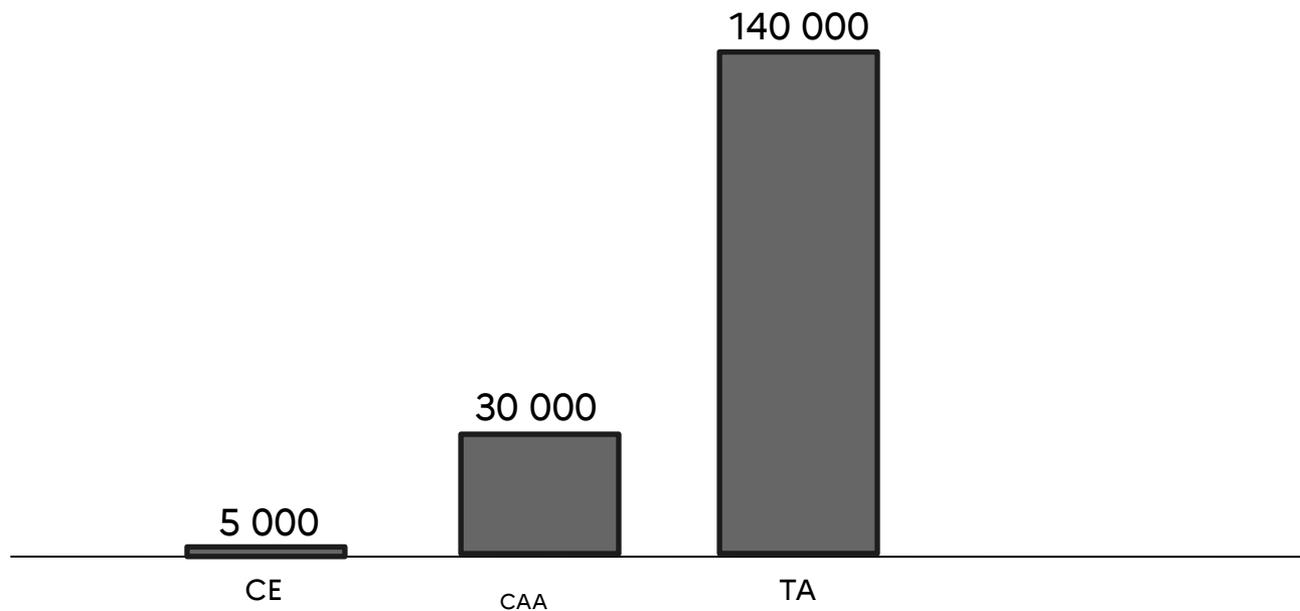
Minimiser les interventions humaines pour la levée de doute



Obtenir un outil plus performant pour la mise à disposition des données numériques dans le nouvel plan Open Data



Flux annuel des décisions*



** sans les ordonnances de rectification et de défaut*

La démarche générale

Partenariat entre
Etalab et le CE



S'appuyer sur les
connaissances déjà
produites à la cour
de Cassation



Réaliser une
première version du
logiciel (MVP)



Les spécificités du projet au Conseil d'État

Le Conseil d'État possède déjà son outil de pseudonymisation

-  Développé sur la base de règles de gestion
-  Comprend un moteur d'anonymisation et une interface de curation
-  Pseudonymise les noms, prénoms et adresses
-  Traite des fichiers Word

Pas d'équipe d'annotateurs

-  Les décisions annotées sont extraites de la base de données

Deux machines pour le système TAL :

-  Moteur physique et une machine virtuelle

Contexte Etalab

Etalab

Département de la DINUM, Etalab coordonne la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée.

Une action tout au long du cycle de vie de la donnée :

-  Ouverture (open data) et partage des données
-  Exploitation des données et algorithmes publics
-  Innovation et ouverture de l'action publique

Le Lab IA

Trois grands axes

-  Accompagner les projets IA de l'administration
-  Développer des outils mutualisés pour l'action publique
-  Animer la communauté IA de l'administration

Et aussi...

-  Développer l'éthique de l'IA dans la sphère publique et la transparence vers les citoyens
-  Construire un partenariat avec l'écosystème de la recherche en IA

Pseudonymisation : un enjeu mutualisé

Un sujet commun aux administrations

ASN, INPI, DIRECCTE

Le Lab IA d'Etalab accompagne et partage ses expertises

Partenariat Etalab – Cour de Cassation

Partenariat Etalab – Éditions Lefebvre Sarrut

Partenariat Etalab – CE

Réalisations

Déroulement des travaux

Systeme NLP

-  Génération d'un dataset d'entraînement
-  Entraînement d'un modèle de Reconnaissance d'Entités Nommées
-  Validation et analyse des erreurs
-  Détection des données personnelles et pseudonymisation

Déploiement

-  Sur deux machines : moteur de calcul (entraînement) + serveur virtuel (inférence)

Transfert de connaissances

-  Onboarding avec la DSI du CE

Comparaison entre les deux logiciels

	Outil actuel	Outil NLP
État du logiciel	Industrialisé avec IHM	Non industrialisé
Fonctionnement	Règles de gestion	IA – NLP
Format de données	.doc → .doc	.doc → .txt
Performance	F-score : 0.88 Taux d'erreur : 76 %	F-score : 0.93 Taux d'erreur : 58 %

Réalisations mutualisées : guide pseudonymisation



guides.etalab.gouv.fr



Sommaire

Télécharger les guides

Contacter Etalab

Code source

Pseudonymiser des documents grâce à l'IA

Introduction

[À quoi sert ce guide ?](#)

[À qui s'adresse ce guide ?](#)

[Sommaire](#)

[Comment contribuer ?](#)

Pourquoi et comment pseudonymiser dans l'administration

Les étapes d'un projet de pseudonymisation grâce à l'IA

La pseudonymisation par l'IA en pratique

Lexique des termes techniques

Introduction

À quoi sert ce guide ?

De nombreuses administrations publiques sont confrontées à des problèmes de pseudonymisation dès lors qu'elles ont à publier des **documents textuels contenant des données à caractère personnel**. Ces documents recouvrent par exemple des décisions de justice, des actes administratifs, des procès-verbaux, des notes, etc.

C'est dans ce cadre qu'Etalab a développé [un outil d'intelligence artificielle de pseudonymisation](#) pour le Conseil d'État, qui publie en open data des décisions de justice administrative. Cet outil est open-source et peut donc être librement réutilisé pour d'autres projets de pseudonymisation.

Pour accompagner la publication de cet outil technique de pseudonymisation, nous pensons qu'il est nécessaire de publier également **un guide qui expose ce qu'est la pseudonymisation de documents textuels et, lorsque c'est possible, l'utilisation de l'intelligence artificielle (IA) pour la mettre en œuvre**.

À qui s'adresse ce guide ?

Ce guide s'adresse principalement **aux organismes publics**, et plus particulièrement **aux personnes**

<https://guides.etalab.gouv.fr/pseudonymisation/>

Réalisations mutualisées : Pseudo App/API



Démo Pseudo^β



Voir sur Github

À propos Pseudonymisez un document Volume de données annotées API Stats

Veillez choisir un fichier à analyser (type .doc, .docx, .txt. Max 100 Ko)

Attention: cette application n'est qu'une démo, aucune donnée n'est conservée.

Veillez à ne pas transmettre d'informations sensibles.

Faire glisser ou cliquer pour charger un fichier

Ou lancez le texte exemple en cliquant ici

Document annoté Document pseudonymisé

Lecture du mercredi 28 février 2018

REPUBLIQUE FRANCAISE

AU NOM DU PEUPLE FRANCAIS

Vu la procédure suivante:

Mme Aleron **PRENOM** Landry **NOM**, demeurant au 123 rue Fausse Ville-Fantastique-Sur-Saône **ADRESSE**, a demandé au juge des référés du tribunal administratif de Poitiers, sur le fondement de l'article L. 521-1 du code de justice administrative, de suspendre l'exécution de la décision du 21 juillet 2012 par laquelle le recteur de l'académie de Bordeaux a rejeté sa demande d'inscription en première année de licence de sciences et techniques des activités physiques et sportives (STAPS) à l'université de Bordeaux pour l'année 2002/2003 et d'enjoindre au recteur de l'inscrire temporairement au sein de cette formation dans un délai de quinze jours, sous une astreinte de 10 euros par jour de retard.

Par une ordonnance n° 1703763 du 21 mars 2011, le juge des référés du tribunal administratif a suspendu l'exécution de cette décision et a enjoint au recteur de l'académie de Bordeaux de procéder à l'inscription de Mme Landry **NOM** en première année de licence de STAPS dans l'attente qu'il soit statué au fond sur sa légalité.

Par un pourvoi, enregistré le 3 janvier 2009 au secrétariat du contentieux du Conseil d'Etat, la ministre de l'enseignement supérieur, de la recherche et de l'innovation demande au Conseil d'Etat d'annuler cette ordonnance.

<https://datascience.etalab.studio/pseudo/>

Prochaines étapes

Travaux de nettoyage et filtrage des décisions de justice

Effectuer un post-traitement pour corriger les erreurs du système

Combiner le logiciel IA avec des règles de gestion voire combiner les deux logiciels

Réfléchir à comment se passer des fichiers .doc

Continuer à explorer des améliorations au modèle

Consolider le dataset d'entraînement

P. ex., SORIANO détecté qu'une fois et pas plus

Étudier la complémentarité des deux approches

Passer à un format de fichier format plus Open Data

Plus de données, plus de paramètres

Cour de Cassation
