



Fais-le 21/07/2025

Rapport d'Échange Professionnel

Candidature à l'alternance – Apprenti MOA (Conception & Numérisation) Service de la Documentation Nationale du Cadastre (SDNC)

Poste visé : Apprenti Maîtrise d'Ouvrage – Numérisation & Acquisition de Données

Localisation : Saint-Germain-en-Laye (78)

Échange entre :

Recrutementeurs :

Mme. Sabrina MERABET

M. Michel Feigenbrügel

Mme. Lisa Godart :

Le candidat :

BOUBANDA Lévi-junior

Table des matières

Contexte de l'Échange	2
1.Présentation du candidat	2
2.Présentation de l'application développée	3
3.Objectifs de la maquette	4
4.Fonctionnalités clés	4
5.Compétences mobilisées	5
6.Vision de l'intégration au SDNC	5
Conclusion	6

Contexte de l'Échange

Ce rapport s'inscrit dans le cadre de ma candidature pour une alternance au Service de la Documentation Nationale du Cadastre (SDNC), dans le domaine de la numérisation documentaire et de l'extraction de données.

À travers ce document, je souhaite vous présenter une application technique que j'ai conçue avec Streamlit. Ce projet illustre ma capacité à comprendre des problématiques métier et à concevoir des solutions numériques simples, utiles et évolutives, dans un contexte proche des activités de la DGFIP.

1. Présentation du Candidat

Je m'appelle Lévi Junior Boubanda, actuellement étudiant en 3^e année de passage en 4 - ème année du Programme Grande École à Aivancity School for Technology, Business & Society, en spécialisation Intelligence Artificielle, Data Science et Gestion de Projet Numérique, en vue d'un diplôme visé Bac+5. Mon parcours m'a permis de développer un profil complet, mêlant compétences techniques solides, compréhension fonctionnelle des besoins métiers et réflexion orientée impact concret.

Passionné par la valorisation intelligente de la donnée, je me suis investi dans des projets touchant à l'automatisation de chaînes de traitement, à l'extraction d'informations depuis des documents structurés et non structurés, ou encore à la modélisation prédictive et au traitement du langage naturel (NLP). Chez Fairmat j'ai eu l'occasion de construire des pipelines d'extraction complexes (plus de 16 000 données issues d'APIs ou de pages web), de mettre en œuvre des modèles de machine learning (sur la qualité de vin, la détection de maladies d'abeilles), et de travailler avec des outils avancés comme BERT pour l'analyse de texte.

Techniquement, je maîtrise Python, SQL, Pandas, Scikit-learn, TensorFlow, PyTorch, XGBoost, ainsi que les outils de visualisation Power BI et Tableau, et les environnements cloud tels que AWS, Azure et GCP. Je poursuis également une formation complémentaire sur Dataiku, pour renforcer mes compétences sur les plateformes de data intégrées.

En parallèle de ces expériences techniques, j'ai développé des qualités humaines et opérationnelles précieuses dans des environnements exigeants comme la restauration rapide (McDonald's, Burger King), où j'ai appris la rigueur, l'organisation, la réactivité et la collaboration en équipe. Mon engagement bénévole au Secours Catholique m'a renforcé dans l'écoute, le sens du service et la responsabilité sociale.

Aujourd'hui, je souhaite rejoindre la cellule MOA du SDNC, convaincu que ce poste représente un point d'équilibre stratégique entre technique, métier et service public. Les missions de conception de chaînes de numérisation, de recette fonctionnelle, et

d'acquisition automatisée de données s'inscrivent parfaitement dans mes compétences et aspirations. Je suis prêt à contribuer à des projets structurants en m'appuyant sur ma capacité à comprendre les enjeux métier, traduire les besoins en solutions concrètes et accompagner leur mise en œuvre avec méthode, rigueur et esprit d'équipe.

2. Présentation de l'application développée

Dans le cadre de ma préparation à cet entretien, j'ai développé une application de démonstration en Python avec Streamlit, destinée à illustrer ma compréhension des enjeux métier liés à la numérisation intelligente de documents et à démontrer ma capacité à concevoir une solution fonctionnelle, simple et adaptable.

Objectif de la maquette

Cette application vise à simuler une chaîne complète de traitement documentaire, depuis le téléversement d'un fichier jusqu'à l'export structuré des données extraites, avec un volet de validation manuelle pour garantir la qualité finale des résultats.

Fonctionnalité principale

Étape	Fonctionnalité intégrée	Technologie utilisée
Téléversement	Chargement de fichiers PDF ou images	Streamlit
OCR	Reconnaissance de texte automatique	Tesseract
Extraction de données	Identification de champs clés via expressions régulières (regex)	Regex + Python
Interface de correction	Modification manuelle et validation par l'utilisateur	Streamlit (form UI)
Export structuré	Données enregistrées aux formats JSON et CSV	Pandas / JSON writer

Exemples de champs extraits automatiquement :

Référence de document

Nom / Prénom

Date

Montant

SIRET

Adresse électronique

Cas d'usage simulés

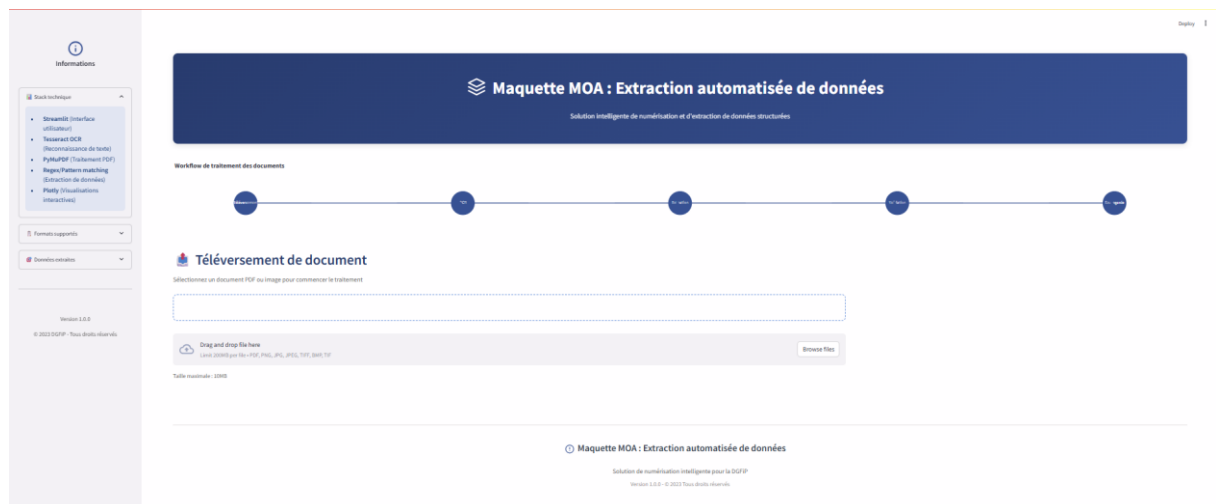
Cette maquette s'inspire de situations réelles rencontrées au sein de structures comme

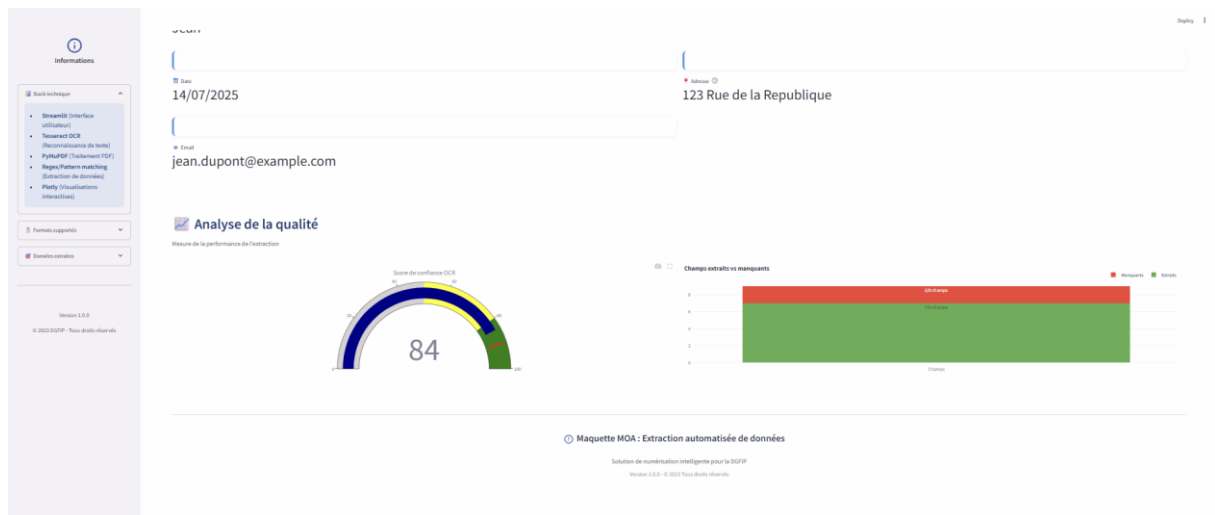
Le SDNC (traitement de formulaires cadastraux ou fiscaux)

Le BNIC (bureau national d'identification cadastrale)

Les SAPF (services d'appui à la publicité foncière)

L'objectif est de proposer une approche générique, légère et personnalisable, facilement intégrable dans une chaîne de numérisation existante, avec une attention particulière portée à la simplicité d'usage pour les agents métier.





Résultats de l'extraction

Vérifiez et validez les données extraites

Données structurées Texte brut Validation

Texte brut extrait par OCR

DOCUMENT DE TEST POUR OCR
Références:
REF-0000
Date:
14/07/2025
Nom:
DUPONT
Prénoms:
Jean
Adresse:
123 Rue de la République
75001 PARIS
Téléphone:

Statistiques du texte

Nombre de caractères
424

Nombre de mots
70

Nombre de lignes
39

Résultats de l'extraction

Vérifiez et validez les données extraites

Données structurées Texte brut Validation

Validation et correction manuelle

Numéro de référence	Montant
entree02	5860
Nom	SIRET
DUPONTPrénom	
Prénoms	Téléphone
Jean	
Date	Adresse
14/07/2025	123 Rue de la République
Email	
jean.dupont@example.com	

Sauvegarder JSON Sauvegarder CSV Réinitialiser

Sauvegarder JSON Sauvegarder CSV Réinitialiser

3. Objectifs de la maquette

Cette application a été conçue dans une logique démonstrative, afin de répondre à plusieurs objectifs en lien direct avec les missions d'une cellule MOA spécialisée dans la numérisation intelligente et l'acquisition de données structurées.

Objectifs principaux :

Compréhension des enjeux métier

Montrer ma capacité à analyser un besoin fonctionnel dans le contexte de la digitalisation documentaire (cadastre, formulaires fiscaux, pièces justificatives...).

Intégrer les contraintes métier (lisibilité, qualité des extractions, corrections manuelles, formats exploitables).

Approche MOA centrée utilisateur

Illustrer une posture de Maîtrise d'Ouvrage, à travers une vision fonctionnelle claire, la structuration des étapes de traitement et la prise en compte de l'expérience utilisateur.

Proposer une interface simple et exploitable par un agent non technique.

Prototype léger, reproductible et adaptable

Développer un outil léger, portable, open source (basé sur Python et Streamlit), facile à adapter à différents cas d'usage (SAPF, BNIC, SDNC).

Créer une solution réutilisable en environnement de test, formation, ou dans un prototype de preuve de concept.

4. Fonctionnalités Clés

L'application a été conçue pour proposer une chaîne de traitement complète, depuis l'import du document jusqu'à l'export structuré des données extraites. Elle intègre à la fois des fonctions techniques robustes et une interface simplifiée pour les agents métier.

Fonctionnalités principales :

1. Support multi-format

Téléversement de documents PDF ou images scannées (PNG, JPEG, TIFF)

Taille maximale prise en charge : 200 Mo

2. Reconnaissance de texte (OCR)

Intégration de Tesseract, moteur open-source reconnu

Traitement performant sur texte imprimé ou scanné

3. Extraction automatique des données

Utilisation d'expressions régulières (Regex) pour identifier les champs :

- Nom / Prénom
- Référence de document
- Date
- Montant
- SIRET
- Adresse email

4. Interface utilisateur accessible

Prévisualisation du document original

Correction manuelle des champs extraits

Validation des données avant export

5. Export structuré et interopérable

Sauvegarde des résultats au format CSV ou JSON

Structure adaptée aux outils métier (Power BI, Dataiku, etc.)

6. Workflow visuel avec indicateurs

Affichage graphique des étapes de traitement

Indication des taux de confiance OCR et champs à vérifier

5. Compétences mobilisées

- Python : traitement des fichiers, structuration des données
- Streamlit : développement de l'interface utilisateur
- Tesseract OCR : reconnaissance de texte à partir d'images ou PDF
- Regex : extraction automatisée d'informations structurées
- Pandas : transformation de données
- Structuration logicielle dans une logique MOA / métier

6. Vision de l'intégration au SDNC

Je me projette pleinement dans une logique d'apprentissage en alternance au sein du SDNC, avec une montée en compétences progressive autour de :

- L'amélioration des chaînes de numérisation en production
- La documentation fonctionnelle
- L'accompagnement des utilisateurs
- Le lien entre les données physiques et numériques

Travailler sur des projets d'envergure nationale avec impact concret sur la modernisation des services publics est un véritable moteur pour moi.

Conclusion

Ce projet Streamlit incarne ma volonté de lier technique et sens du service public. Il démontre ma capacité à concevoir une solution fonctionnelle, orientée métier, et à contribuer efficacement à des équipes comme celle du SDNC.

Je suis prêt à m'investir dans une alternance enrichissante, et je reste disponible pour une démonstration de l'application, un entretien complémentaire, ou toute précision utile.