



Offre n°2020-02913

Ingénieur(e) data science collaboration Inria - AP-HP Entrepôt de données de Santé (EDS)

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Autre diplôme apprécié : Thèse

Fonction : Ingénieur scientifique contractuel

A propos du centre ou de la direction fonctionnelle

Le centre de recherche Inria Saclay – Île-de-France, créé en 2008, accueille 450 scientifiques et 60 membres des services d'appui à la recherche. Les scientifiques sont organisés en 28 équipes de recherche dont 23 sont communes avec des partenaires du plateau de Saclay.

Le centre Inria Saclay - Île-de-France est un acteur essentiel de la recherche en sciences du numérique sur le plateau de Saclay. Il porte les valeurs et les projets qui font l'originalité d'Inria dans le paysage de la recherche : l'excellence scientifique, le transfert technologique, les partenariats pluridisciplinaires avec des établissements aux compétences complémentaires aux nôtres, afin de maximiser l'impact scientifique, économique et sociétal d'Inria.

Contexte et atouts du poste

L'équipe a été fortement impliquée dans le pilotage de la crise sanitaire du Covid-19 au sein de l'AP-HP. Cette implication a été matérialisée par un travail intense sur plus de 2 mois avec l'Entrepôt de Données de Santé (EDS) de l'AP-HP. L'équipe a mis en place des outils logiciels en Python servant à la réalisation d'un dashboard web qui était quotidiennement transmis au sein de l'AP-HP. L'objectif de ce poste d'ingénieur à l'interface entre data science et recherche clinique est de poursuivre ce travail entre Inria et l'EDS de l'AP-HP.

Cette vidéo parle du travail réalisé et donne les bases du projet : <https://www.youtube.com/watch?v=9uuFGBYFSok>

Voici une autre communication sur le sujet : <https://www.inria.fr/fr/scikiteds>

Le poste sera localisé à mi-temps entre Inria Saclay sur le campus de l'école Polytechnique et le site de l'EDS de l'AP-HP dans le 12^{ème} à Paris.

Mission confiée

La librairie Python développée durant la crise Covid s'appuie sur la base SQL "CoronaOmop" (de type postgres ou pyspark/HIVE) au format standard OMOP. Ce schéma de base de données est un standard du monde médical permettant la potentielle réutilisation du code à grande échelle, y compris sur le Health Data Hub. La librairie et le dashboard sont basés sur le stack de Python scientifique Numpy, Scipy, Pandas, Matplotlib, Plotly et scikit-learn.

La librairie est testée avec pytest et l'intégration continue se fait avec Gitlab-CI.

La librairie est documentée avec sphinx.

Objectifs:

- > Reprise et généralisation sur la base OMOP des librairies développées pour CoronaOmop. L'objectif est de mettre le code en open-source (utilisation par les hôpitaux ayant standardisé leurs données au format OMOP)
- > Création de briques logicielles transversales pour la data-quality / data management / enrichissement de données (ex : pipelines NLP/validateur des extractions, outil d'annotation de données)
- > Développer des briques de visualisation de données pour Cohort360 sur la base OMOP (sankey plot, pyramide des âges...). Ce travail sera orienté recherche et soin.
- > Développer des briques d'exploration de données pour le pilotage

En termes de code déjà publique qui servira de base:

> le code de la librairie covid19-python développée par l'équipe Inria

> le framagit « inter-CHU » très animé par Nicolas Paris (mais pas que) <https://framagit.org/interchu> (Suzanna ou Omop Mapper : <https://framagit.org/interchu/omop-omop-back>)

> le GitHub EDS qu'il conviendra de consolider : <https://github.com/EDS-APHP>

Principales activités

Principales activités (5 maximum) :

- Analyser les besoins de l'AP-HP
- Proposer des solutions logiciels
- Développer des programmes en Python et possiblement web
- Tester, modifier jusqu'à valider les outils développés
- Animer une communauté d'utilisateurs médecins

Compétences

Compétences techniques et niveau requis :

- Python
- Git
- Numpy, Scipy, Scikit-Learn, Matplotlib, Pandas
- SQL

Langues : Français et Anglais courant

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Rémunération

Rémunération selon expérience

Informations générales

- **Thème/Domaine** : Neurosciences et médecine numériques
Ingénierie logicielle (BAP E)
- **Ville** : Palaiseau
- **Centre Inria** : [Centre Inria de Saclay](#)
- **Date de prise de fonction souhaitée** : 2022-06-01
- **Durée de contrat** : 1 an, 10 mois
- **Date limite pour postuler** : 2020-09-30

Contacts

- **Équipe Inria** : [SODA](#)
- **Recruteur** :
Varoquaux Gael / Gael.Varoquaux@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

- Goût pour le développement logiciel (expérience de l'open source idéale)
- Connaissance des librairies standards de data science en Python (Numpy, Pandas, Scikit-Learn)
- Connaissance d'un gestionnaire de suivi de version tel que git ou équivalent
- Goût pour le travail en équipe
- Intérêt pour le domaine de la santé / médecine
- Avoir envie de faire avancer les choses avec autonomie et motivation personnelle

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.