



Offre de CDD 6 mois

Ingénieur en intégration de données

MOT CLES : GENETIQUE, PHENOMIQUE, DATA MANAGER, BASE DE DONNEES, BIG DATA

MODALITES DE CANDIDATURE

Les candidatures (CV + lettre de motivation) doivent être adressées avant le 31 janvier 2019 par courriel à celia.michotey@inra.fr et bruno.fady@inra.fr avec l'objet suivant : [GENTREE-DATA]
Les auditions sont à prévoir en janvier/février et la prise de fonction aura lieu en février/mars.

CONTEXTE

L'URGI, unité de recherche INRA en génomique info dédiée à la génomique et à la génétique des plantes et de ses bioagresseurs, recrute un ingénieur en intégration de données pour un contrat CDD de 6 mois.

La personne intégrera l'équipe « Système d'information et intégration de données » qui suit les méthodologies agiles (*Scrum*, *Extreme Programming*), localisée sur le site INRA de Versailles.

MISSIONS

La personne recrutée participera à l'intégration de données de génomique et phénotypique d'arbres forestiers dans le système d'information de l'URGI : GnpIS [1]. Ces données sont issues du projet européen H2020 GenTree [2] dont l'objectif est de fournir au secteur forestier européen de meilleures connaissances, méthodes et outils pour optimiser la gestion et l'utilisation durable des ressources génétiques forestières.

Les outils d'intégration des données reposent sur des technologies adaptées au défi du Big Data appliqués aux données scientifiques : ETL (Talend), NoSQL orienté document (Elasticsearch), les langages SQL, bash et éventuellement python.

La personne aura à interagir avec le responsable du projet à l'URGI, ainsi que les chercheurs européens producteurs des données. Elle devra être méticuleuse et plonger au cœur des données.

COMPETENCES

- Bonnes capacités relationnelles, goût pour le travail en équipe.
- Curiosité pour la nature des données scientifiques.
- Maîtrise de l'anglais technique du domaine et capacité à bien parler anglais.
- Connaissances en bases de données relationnelles (SQL) et en NoSQL orienté document (ex : Elasticsearch).
- Connaissance d'un outil d'ETL (ex : Talend).
- Connaissance d'un langage de script (bash, perl ou python)
- Maîtrise de l'environnement Linux.

FORMATION

Master2 (bioinformatique, informatique) ou Ecole d'ingénieur.

[1] <https://urgj.versailles.inra.fr/Tools/GnpIS>

[2] <http://www.gentree-h2020.eu/>