L'Abes recrute

Un(e) stagiaire

Ingénieur(e) informatique - gestion de données

- Début du stage : février 2019
- Durée : 5 mois
- Stage localisé à Montpellier
- Gratification selon tarif officiel : 535€ / mois

Contexte du stage

Depuis 1994, l'Abes met en œuvre plusieurs outils documentaires nationaux parmi lesquels **le Sudoc,** catalogue collectif des bibliothèques universitaires françaises, riche de plus de 17 millions de notices décrivant 40 millions de documents localisés dans plus de 3 500 bibliothèques, et **theses.fr,** catalogue des thèses de doctorat soutenues en France depuis 1985, décrivant plus de 400 000 thèses.

Le projet auquel vous serez intégré(e) consiste à étudier le remplacement du **cœur du système de stockage et de manipulation des métadonnées** de l'Abes. Ces métadonnées, riches, structurées, de qualité et massives, sont aujourd'hui gérées dans un système propriétaire fourni par un prestataire. Ce système, très robuste, est difficile à paramétrer et à adapter aux nouveaux usages.
Au fil des années, l'Abes a développé différents outils autour de cette base de production, afin de mieux exploiter les métadonnées pour répondre aux demandes de services de son réseau : une base Oracle pour convertir les données en XML ; plusieurs bases RDF (Virtuoso) pour manipuler et enrichir les notices à partir de sources externes ; des moteurs de recherche SolR ; un outil de raisonnement développé en collaboration avec l'équipe GraphIK du LIRMM (projet ANR Qualinca). Il en résulte un système complexe, lourd à maintenir, qui nécessite de nombreuses synchronisations entre les différentes bases, lesquelles contiennent des données dans des formats différents (MARC, XML-EAD, TEF, KBART, ONIX, etc)

Une vision du produit cible a été définie : il s'agit de mettre à disposition de l'Abes et des établissements membres des réseaux un système permettant, tout comme le système actuel, l'ingestion, le traitement, le stockage et la fourniture de données bibliographiques, mais en le faisant désormais de manière unifiée, dans un environnement technique ouvert, adaptable, évolutif donc maîtrisable par l'Abes, tout en s'appuyant sur l'exhaustivité des données utiles au métier, structurées selon un modèle de données unique compatible IFLA-LRM (modèle fondé sur les tâches utilisateurs).

La 1ère phase du projet a débuté en avril 2019 et prendra fin en juin 2020. Menée en méthode Agile, elle consiste à réaliser des preuves de concept (PoC) à partir d'outils, technologies ou frameworks. L'équipe projet est constituée de 9 membres, bibliothécaires et informaticiens.

Qui sommes-nous ?

L'Abes - Agence Bibliographique de l'Enseignement Supérieur est un établissement public administratif (EPA) employant 80 agents - titulaires et contractuels - dans les filières bibliothèque, informatique et administrative.

En tant qu'agence nationale, l'Abes administre une gamme d'outils et services mutualisés au profit de l'ensemble des établissements documentaires de l'Enseignement Supérieur et de la Recherche :

- **Sudoc** : bases de données bibliographiques pour faciliter l'accès aux collections des bibliothèques universitaires et de recherche
- **theses.fr - thèses de doctorat :** gestion des métadonnées et de l'archivage des thèses numériques soutenues en France
- **Calames - archives et manuscrits de l'ESR** : description et accès aux collections d'archives et manuscrits de l'ESR
- **IdRef - identifiants et référentiels** : notices d'autorités produites par les bibliothèques de l'ESR pour une meilleure identification des auteurs
- **Bacon - Base de Connaissance Nationale** : métadonnées des ressources numériques à la disposition des établissements de l'ESR
- **Portage financier** : achat de ressources numériques ; accord-cadre pour l'acquisition d'un Système de Gestion de Bibliothèques mutualisé

En savoir plus…

Gardons le contact !

Sujet du stage

Il s'agira, au sein de l'équipe, d'analyser, évaluer et tester un éventail de solutions possibles pour répondre aux différentes problématiques :

- La modélisation des données bibliographiques en différents formats selon le modèle IFLA-LRM
- L'intégration de données venant de différentes bases : Sudoc, theses.fr, Bacon, etc,
- La réalisation d'expérimentations sur différents outils et sur différents types de *framework* :
    - Bases relationnelles
    - Bases RDF
    - Graphes de propriété (Oracle, Neo4J, TigerGraph…)
    - Wikibase
- La réalisation de *benchmark* sur les problématiques de volumétrie des données et de montée en charge,
- Les données de gestion sur les métadonnées : traçabilité des modifications, réversibilité, *versioning*, granularité, *sourcing*, données pour statistiques.
- L'import et l'export en masse de données
- Le contrôle et la validation des données issues du catalogage partagé coordonnée par l'Abes : expression des règles métier (dont règle de similarité doublon)
- L'utilisation de référentiels externes

Vous participerez pleinement à la construction des prototypes sur lesquels le choix de la solution s'appuiera.

Qualités attendues

Vous appréciez travailler en équipe projet.
Vous connaissez une ou plusieurs bases de données RDF et NoSQL.
Vous disposez de connaissances en web de données et web sémantique.
Vous aimez modéliser et manipuler de grandes masses de données très structurées.

Informations complémentaires

- Accès parking fermé, local à vélos, proximité ligne de Tram
- Tickets restaurant
- Remboursement partiel des abonnements de transport domicile-travail

CV et lettre de motivation sont à envoyer
à stage@abes.fr ou par courrier postal à :

Abes - Agence Bibliographique de l'Enseignement Supérieur
à l'attention de M. le Directeur
227 avenue du professeur Jean-Louis Viala - CS 84308
34193 Montpellier Cedex 5

Contact pour informations :
Sandra Beffaral, chef du service RH - beffaral@abes.fr
Stéphane Rey, responsable LABO - rey@abes.fr
Standard téléphonique : 04 67 54 84 10

Les éventuels entretiens se déroulent à Montpellier