

POSTE D'INGÉNIEUR·E ARCHIVES ET TECHNOLOGIES SÉMANTIQUES

(CDD IGE 12 mois– Cat. A)

Héritière de la Bibliothèque de l'Université de Paris ouverte en 1770, la Bibliothèque interuniversitaire de la Sorbonne (BIS) est une bibliothèque de recherche spécialisée en lettres et sciences humaines, rattachée pour sa gestion à l'université Paris 1 Panthéon Sorbonne. Elle accueille un public d'étudiant·e·s (à partir de la troisième année de licence) et de chercheur·se·s et enseignant·e·s-chercheur·se·s. Dotée d'un budget annuel de 3,9 millions d'euros et d'un effectif de 128 agents titulaires renforcés par des contractuel·les, elle conserve dans ses collections et propose à ses lecteurs environ 2 millions de documents, dont 17 750 titres de périodiques et 3 500 manuscrits. La bibliothèque de l'institut de géographie lui est rattachée.

Depuis 2017, elle est bibliothèque délégataire du réseau CollEx-Persée pour l'histoire, la géographie, la philosophie et les sciences de l'Antiquité. Elle propose également de riches collections de littérature.

La BIS développe une politique ambitieuse de valorisation de ses collections et d'accompagnement à la recherche. Elle a créé à cette fin en 2019 un service dédié à ces enjeux, le service de la valorisation numérique des collections et du soutien à la recherche (SERVAL).

Elle porte des projets scientifiques et documentaires en lien avec des chercheurs (constitution de bases et d'entrepôts de données, mise en place de référentiels), et est partenaire de projets d'humanités numériques. Ces projets sont coordonnés ou menés au sein du service transverse de la valorisation numérique des collections.

Contexte :

ORESM (Œuvres et Référentiel des Étudiants, Suppôts et Maîtres de l'Université de Paris au Moyen Âge), est un projet porté conjointement depuis 2020 par la Bibliothèque interuniversitaire de la Sorbonne (BIS) et par le laboratoire de médiévistique occidentale de Paris (LaMOP – UMR 8589). Il implique également de très nombreux partenaires dont notamment trois composantes des Archives nationales de France : deux départements de la direction des Fonds qui sont responsables du traitement et de la valorisation des documents provenant de l'Université de Paris ; et le Lab, qui s'occupe de la R&D numérique de l'institution, qui possède une expertise reconnue dans le domaine de l'ingénierie des métadonnées et des technologies sémantiques appliquées aux archives est donc notamment en charge de ses projets *Linked Data*, et qui est très impliqué dans les aspects numériques du projet. Le projet est soutenu et co-financé par le GIS CollEx-Persée.

Le projet ORESM vise à collecter, créer et valoriser des données relatives à l'histoire de l'Université de Paris, des collèges et écoles parisiens entre le XII^e et le XVI^e siècle, et à les rendre consultables sur un portail web dédié.

Ce portail ambitionne entre autres :

- de réunir au sein d'un inventaire virtuel les documents qui relèvent du fonds d'archives de l'université et qui sont dispersés depuis la Révolution entre différentes institutions partenaires du projet, notamment la BIS, les Archives

- nationales de France (AnF) et la Bibliothèque interuniversitaire de Santé (BIU Santé) ;
- de proposer un référentiel ORESM des personnes liées à l'université de Paris au Moyen Âge en s'appuyant sur les riches données de la base *Studium Parisiense* (gérée par le LaMOP) et aligné sur les référentiels documentaires existants, tout en proposant aux chercheurs une évaluation du degré d'incertitude basée sur des critères scientifiques ;
 - de proposer un répertoire des manuscrits liés aux personnes ayant gravité autour de l'université de Paris.

À l'heure actuelle, plusieurs actions ont été menées :

- un inventaire virtuel a été constitué en XML/EAD. Il sera prochainement mis en ligne via l'outil ArchivesSpace ;
- des dépouillements fins d'archives ont été réalisés sur les fonds de 14 collèges parisiens conservés dans la série M des AnF ;
- les données de dépouillement ont fait l'objet de traitements afin de produire un graphe de connaissances : modélisation du graphe à l'aide d'une extension de l'ontologie générique [Records in Contexts Ontology \(RiC-O\) dans sa version 0.2](#), conversion des données de dépouillement en triplets RDF conformes à cette ontologie via un script XSLT ; import du graphe obtenu dans une instance du logiciel GraphDB Free, installée sur un serveur de la BIS. Une première interface de recherche a été créée via l'outil [Sparnatural](#) afin de tester une interrogation plus intuitive du graphe via le langage SPARQL.
- Une fiche d'expression des besoins a été préparée pour réaliser la plate-forme Buridan, qui permettra de réaliser les alignements de noms de personnes issues des inventaires et dépouillements d'archives et des bases Studium parisiense et IdRef en intégrant dans les résultats un facteur d'incertitude.
- Un projet d'édition numérique du *Chartularium universitatis parisiensis* a été financé par l'équipEX Biblissima+, qui sera lié au projet ORESM.

Missions du service :

Le Service de la valorisation numérique des collections et du soutien à la recherche (SERVAL) a pour mission de définir, coordonner et opérer des actions de valorisation des collections sur des outils numériques, promouvoir et porter des partenariats scientifiques en particulier, en lien avec des projets de recherche.

Au 13/12/2023, l'équipe comporte 2 conservatrices titulaires, 1 conservatrice contractuelle, 2 chargé·e·s de projets, trois archivistes, deux ingénieurs informatiques et un gestionnaire de bibliothèque numérique (cat.B).

Missions et activités du poste :

1. À partir du travail déjà fourni et selon les préconisations établies par le comité scientifique, poursuivre la constitution du graphe de connaissances (8 mois), notamment :
 - Mettre à jour et améliorer le script de conversion RDF/RiC-O, procéder à la sémantisation de données complémentaires issues de l'inventaire virtuel cité ci-dessus ;
 - Enrichir le graphe des données ;
 - Mettre en place un workflow le plus automatisé possible pour opérer toutes les opérations de sémantisation et de traitement, afin de pouvoir facilement mettre à jour régulièrement le graphe au fur et à mesure que les données de dépouillement et données complémentaires évolueront ;
 - Veiller à l'import régulier du graphe mis à jour dans l'instance GraphDB, à sa cohérence interne et à sa qualité en assurant un contrôle final

(réalisation de statistiques, élaboration et exécution de règles en SHACL, réalisation des corrections utiles).

2. Concevoir et développer l'application web constituant le portail ORESM (4 mois), intégrant en particulier un SPARQL endpoint et une interface de recherche « intuitive ».
3. Rédiger la documentation technique des développements mis en œuvre, rendre le code compréhensible (commentaires) et accessible (logique d'Open Access) sur la plateforme GitHub.
4. Participer aux autres aspects du projet en fonction de son expertise.
5. Présenter le travail en cours aux partenaires, au sein des instances du projet (comité scientifique et groupes de travail) et lors de journées d'études.

La personne recrutée mènera à bien ces missions en concertation avec l'ensemble des personnes participant au projet, en particulier pour ce qui concerne la conception des interfaces de recherche et de consultation. Ces missions seront co-pilotées par la BIS et le Lab des AnF.

Rattachement hiérarchique : sous la responsabilité de l'adjointe à la cheffe du SERVAL, au sein de la Bibliothèque interuniversitaire de la Sorbonne, Université Paris 1 Panthéon-Sorbonne.

Liens fonctionnels :

En interne :

- Coordinatrice scientifique du projet à la BIS
- Département des manuscrits et livres anciens (DMLA)

En externe :

- Coordinateur scientifique du projet à l'université de Franche-Comté
- Archives nationales de France : Lab ; départements du Moyen Âge et de l'Ancien Régime et du Minutier central des notaires de Paris
- Laboratoire de médiévistique occidentale de Paris (LaMOP)
- Chercheur·se·s et professionnel·le·s des archives et des bibliothèques participant au comité scientifique et aux groupes de travail du projet

Compétences :

- Méthodologie de conduite de projet appuyée sur l'expérience utilisateur.
- Maîtrise des concepts du web sémantique et des *Linked Open Data* ; expérience solide de leur mise en œuvre concrète : production, sérialisation, administration et gestion de la qualité de données RDF, utilisation des langages RDFS, OWL (pour éditer une ontologie et tirer parti de sa logique formelle) SPARQL (protocole SPARQL, SPARQL Query, SPARQL Update) et si possible SHACL. Réalisation d'application web de recherche et de consultation de graphe RDF. Connaissance concrète des logiciels de bases de graphe tels que GraphDB.
- Maîtrise de Git et GitHub.
- Formation souhaitée en histoire médiévale, archivistique et humanités numériques. Une expérience de participation à un projet d'humanités numériques serait un plus.
- Anglais lu et écrit.
- Connaissance souhaitée, au moins théorique des principes et normes de description archivistique (ISAD(G), ISAAR(CPF), et du nouveau modèle conceptuel Records in Contexts) ;
- Connaissance des formats XML/EAD, TEI, maîtrise des technologies XML (notamment, langage XPath, transformation de documents XML avec le langage XSLT 2.0).

- Expérience souhaitée de méthodes et outils de réconciliation et d'alignement de données ; maîtrise d'un langage de programmation tel que Python. Expérience requise en développement d'interface Web (maîtrise des langages HTML, CSS, JavaScript, JSON).

Savoir-être :

- Capacités d'adaptation à l'environnement scientifique du projet et d'apprentissage
- Capacités d'analyse et de synthèse
- Écoute et sens du dialogue
- Aptitudes à la pédagogie et à la médiation
- Excellentes capacités d'expression écrite et orale
- Aptitude à travailler en équipe
- Rigueur et organisation
- Capacités d'analyse, d'abstraction et de synthèse ;
- Autonomie et capacité d'initiative, capacité à évaluer son activité et à en rendre compte

Contraintes liées au poste :

35h hebdomadaires. Lieu de travail principal situé à Paris 5^e. Pendant la mission 1, déplacements fréquents à prévoir aux Archives nationales, site de Pierrefitte-sur-Seine, pour des réunions de travail et ateliers (en moyenne 2 jours par semaine).

Statut : CDD de 12 mois, rémunération selon expérience sur la base de la grille des ingénieurs d'étude.

Prise poste souhaitée : 01/02/2024

Contacts pour information et candidature :

Par courriel à :

- Laurence BOBIS, directrice de la bibliothèque, coordinatrice scientifique du projet ORESM : direction-bis@univ-paris1.fr
- Thierry KOUAME, professeur d'histoire médiévale à l'université de Franche-Comté, coordinateur scientifique du projet ORESM : thierry.kouame@univ-fcomte.fr
- Cécile OBLIGI, cheffe du SERVAL : cecile.obligi@bis-sorbonne.fr
- Laurie AOUSTET, adjointe à la cheffe du SERVAL : laurie.aoustet@bis-sorbonne.fr
- Florence CLAVAUD, responsable du Lab des Archives nationales de France : florence.clavaud@culture.gouv.fr