

## FICHE DE POSTE

Libellé de Fonctions :

### Ingénieur(e) d'études en humanités numériques

**Catégorie : A**

**Corps : IGE**

**BAP (le cas échéant) : D**

**Emploi-type (le cas échéant) : D2A41 - Ingénieur-e d'études en production, traitement, analyse de données et enquêtes**

*Les activités qui composent la fiche de poste sont appelées à évoluer en fonction des connaissances du métier et des nécessités de service*

### Qui sommes-nous ?

Sorbonne Université est une université pluridisciplinaire et de recherche intensive. Poursuivant la tradition humaniste de la Sorbonne, elle s'attache à répondre aux enjeux scientifiques du 21<sup>e</sup> siècle et à transmettre les connaissances issues de ses laboratoires et de ses équipes de recherche à ses étudiantes et étudiants et à la société tout entière.

Déployant ses formations auprès de 54 000 étudiantes et étudiants dont 4 700 doctorantes et doctorants et 10 200 étudiantes et étudiants étrangers, elle emploie 6 300 enseignantes et enseignants, enseignantes-chercheuses et enseignants-chercheurs, chercheuses et chercheurs et 4 900 personnels de bibliothèque, administratifs, technique, sociaux et de santé. Son budget est de 670 M€.

Sorbonne Université, principalement située au cœur de Paris, dispose d'un potentiel de premier plan et étend sa présence dans plus de vingt sites en Ile-de-France et en régions.

### Présentation de la structure

#### Le CELLF :

Laboratoire de recherche en littérature, philologie et science des textes, allant du XVI<sup>e</sup> au XXI<sup>e</sup> siècle et incluant la francophonie, spécialisé notamment dans l'histoire littéraire et l'édition critique (papier et numérique), le Centre d'étude de la langue et des littératures françaises (UMR 8599 CELLF, CNRS - Sorbonne Université) compte 80 chercheurs et enseignants-chercheurs, 6 IT-BIATSS, et 150 à 200 doctorants. Des membres non-statutaires lui sont aussi rattachés en tant que « chercheurs associés ». Hébergé par Sorbonne Université, il est multi-site (rue Victor Cousin, rue Serpente). Le CELLF encourage les projets transdisciplinaires et trans-séculaires. Les humanités numériques sont au cœur de son programme scientifique.

#### L'ObTIC :

L'Observatoire des textes, des idées et des corpus (ObTIC) est une équipe-projet de la Faculté des Lettres à Sorbonne Université, dédiée aux humanités numériques et rattachée administrativement au CELLF UMR 8599. Faisant suite à l'ancien LabEx OBVIL, l'ObTIC s'appuie sur une expertise acquise dans la production et l'édition des données, dans la conception et l'expérimentation des logiciels et des algorithmes pour les chercheurs en lettres et sciences humaines et sociales, et dans le développement de nouvelles méthodologies de recherche, fondées sur le numérique, au sein de ces disciplines.

**Localisation :** Maison de la Recherche, 28 rue Serpente, 75006 Paris ; SCAI, campus Jussieu, 4 place Jussieu, 75005 ; BnF DataLab, Bibliothèque nationale de France, Bibliothèque François-Mitterrand Quai François Mauriac, 75013.

## Missions et activités principales

### Mission :

Co-financé par le projet européen ERC ModERN et l'ObTIC, ce poste a pour objectif de renforcer le déploiement de la politique numérique des deux équipes. Au sein du projet ModERN, la personne recrutée travaillera en étroite collaboration avec le PI et les autres membres de l'équipe. Son travail l'amènera aussi à collaborer avec divers services et partenaires de Sorbonne Université. La personne recrutée aura pour mission principale de concevoir et administrer une base de données avec des réemplois textuels dans des textes littéraires pour le projet ModERN, ainsi que de développer la politique de traitement et d'exploitation des données de cette BDD. Les exploitations visées sont principalement la conception et la gestion de la BDD du projet ModERN, l'analyse textométrique et la visualisation de données textuelles, mais aussi la gestion générale des serveurs (sécurité, installation du logiciel, configuration).

Il/elle fera des propositions en matière d'outils numériques pertinents et envisagera si besoin des développements spécifiques. Il/elle saura également accompagner techniquement l'équipe ObTIC dans ses projets de recherche et d'analyse et de fouille textuelle, et de réaliser le développement et la maintenance de solutions autour des technologies informatiques pour les humanités numériques. La personne recrutée aura un rôle de formation vis-à-vis de l'ensemble des acteurs ModERN et ObTIC en matière des outils mis en place.

Cette mission sera décomposée de la manière suivante :

### Activités principales :

- Concevoir, modéliser et construire la base de données principale du projet ModERN.
- Développer et implémenter la politique de traitement et d'exploitation des données contenues dans cette base textuelle pour l'alignement de corpus, la fouille de données et l'analyse de réseaux.
- Établir le modèle de données permettant l'interrogation croisée de la totalité du corpus tout en respectant son hétérogénéité et la spécificité des différents sous-ensembles, afin de permettre des recherches croisées
- Encoder les données et métadonnées selon le modèle de données établi.
- Proposer des solutions informatiques en matière de logiciels et d'outils numériques pour et envisager des développements ponctuels si besoin.
- Accompagner les projets de recherche soutenus par l'équipe ObTIC, notamment par le développement ponctuel de solutions d'analyse textuelle et de fouille de textes.
- Accompagner le développement des outils numériques et de la production documentaire : environnement, données, outils et procédures.
- Participer à la rédaction des spécifications de projets de fouille et analyse textuelle.
- Participer à la vie scientifique des deux équipes : formations, séminaires et ateliers.

**Conduite de projet : OUI**

**Encadrement : NON**

*Dans le cadre de vos fonctions, vous pourrez être amené à dispenser des formations internes en lien avec votre expertise métier.*

## Connaissances et compétences\*

### Savoir-faire transversaux :

Nous recherchons une personne de formation supérieure (Ingénieur, Master...), avec ce type de compétences :

- Maîtrise de la conception et de l'administration des bases de données relationnelles et/ou orientée graphe
- Connaissance de langages de programmation (Python) et plus spécifiquement de langages de traitement de données (SQL, XML, RDF, SPARQL).
- Maîtrise d'un langage de script (Python).
- Connaissance des standards patrimoniaux (XML, TEI, RDF...), des langages d'interrogation et de transformation (Xpath, XSLT), des métadonnées, des normes d'interopérabilité et de la science ouverte.
- Connaissance des outils de traitement automatique du langage, de l'apprentissage automatique et des modèles de langue (BERT, CamemBERT).

- Expérience de gestion de serveurs : installation de logiciel, sécurité.
- Autres connaissances appréciées : moteurs de recherche (Solr ou Elasticsearch), édition d'archives numériques (Omeka), outils de visualisation.
- Expérience basique de la gestion des serveurs.

**Connaissances transversales requises :**

- Organisation et fonctionnement de la recherche et de l'enseignement supérieur.
- Organisation et fonctionnement de Sorbonne Université.
- Réglementation applicable à son domaine d'activité professionnelle.
- Statuts, règlements et procédures de gestion des personnels des établissements publics d'enseignement supérieur et de recherche.

**Savoir-faire :**

- Capacité à proposer et à élaborer des tâches en lien avec les projets en cours
- Capacité à définir et à implémenter des solutions informatiques agiles et sur mesure
- Capacité à interagir avec des interlocuteurs de différentes disciplines
- Capacité à hiérarchiser et à prioriser

**Savoir être :**

- Bon relationnel et sens du travail en équipe
- Disponibilité et réactivité
- Sens de l'analyse et force de proposition
- Adaptabilité et autonomie
- Organisation et rigueur
- Sens du service public
- Confidentialité

**Exposition aux risques professionnels :**

Non

Oui : *si oui, indiquer les informations relatives aux risques physiques (port de charge, machines dangereuses, vibrations...), biologiques, chimiques, rayonnements ionisants ou non ionisants. Si l'agent est exposé aux produits dangereux dont les CMR, il doit impérativement disposer d'une Fiche Individuelle d'Exposition téléchargeable sur intranet dans la rubrique « Prevention-des-risques-professionnels/fiche-individuelle-d-exposition-aux-agents-chimiques-dangereux ».*

**Conditions particulières d'exercice :** néant.

**Rémunération :** Rémunération selon le barème de rémunération au recrutement des contractuels BIATSS de Sorbonne Université.

**Prise de fonction :** septembre 2023. **Contrat de 12 mois renouvelable.**

**Documents requis :**

1. Une lettre de motivation (en français et/ou en anglais) à l'attention de la responsable du projet (Glenn Roe) expliquant l'intérêt du/de la candidat(e) pour cette recherche et les compétences qu'il/elle envisage de mobiliser pour la mener à bien.
2. Un CV détaillé.
3. Une copie du document certifiant le niveau d'étude requis.
4. Les noms et adresses email de deux référents.

**Critères de sélection :**

1. Formation supérieure (Ingénieur ou Master) ou équivalent en informatique ou humanités numériques.
2. Compétences démontrées en modélisation et construction des bases de données (relationnelles et/ou sémantiques).
3. Connaissance des humanités numériques.
4. Un intérêt pour la littérature et langue française.
5. D'excellentes capacités de communication.
6. Une bonne maîtrise du français et de l'anglais.
7. Un grand sens de l'organisation.

**Procédure de sélection :**

Les documents requis doivent être envoyés au responsable du projet Glenn Roe ([glenn.roe@sorbonne-universite.fr](mailto:glenn.roe@sorbonne-universite.fr)) ainsi que le Coordinateur scientifique de l'ObTIC Motasem Alrahabi ([motasem.alrahabi@sorbonne-universite.fr](mailto:motasem.alrahabi@sorbonne-universite.fr)).

Tous les documents doivent être soumis avant le 15 juin 2023, au format PDF. Les candidatures incomplètes ou soumises après l'échéance ne seront pas prises en compte.

Les candidat(e)s retenu(e)s seront informé(e)s par email au plus tard le 20 juin 2023 et invité(e)s à prendre part à des entretiens organisés à Paris (ou en distanciel) début juillet 2023.

\*Conformément à l'annexe de l'arrêté du 18 mars 2013 (NOR : MENH1305559A)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council (ERC). Neither the European Union nor the granting authority can be held responsible for them.