



Offre n°2024-08496

Data Librarian pour les corpus de Grec Ancien

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Ingénieur scientifique contractuel

Niveau d'expérience souhaité : Jeune diplômé

Contexte et atouts du poste

Ce poste de data librarian est ouvert au sein de l'équipe-projet ALMA_{na}CH du Centre Inria de Paris. ALMA_{na}CH est une équipe de recherche d'une cinquantaine de membres, dont 7 membres permanents, spécialisée dans le traitement automatique des langues (TAL) et les humanités numériques (traitement des sources historiques et littéraires par l'informatique). Elle fait partie des 230 équipes-projets d'Inria, l'institut national de recherche en informatique et en automatique, établissement public de recherche regroupant 9 centres de recherche, dont le Centre de Paris auquel appartient ALMA_{na}CH.

Le poste prend place dans le contexte du projet Corpus Liberatum Linguae Graecae, dont l'un des objectifs est de numériser et mettre à disposition librement des corpus de textes en grec ancien sous la forme de fichiers XML TEI. Spécifiquement, le projet vise à combler les manques dans les corpus ouverts existants tout en fournissant une alternative sérieuse aux projets uniquement commerciaux et fermés.

Ce projet implique, au-delà d'Inria, une vingtaine de partenaires français issus d'une dizaine d'institutions, rassemblés par la volonté d'améliorer la situation de la science ouverte en études classiques et en particulier pour le grec ancien. Parmi ces partenaires, on trouvera des institutions partenaires pour la numérisation (Persée, INST), des partenaires hellénistes (afin d'aider à guider les priorités du projet) et des partenaires du côté apprentissage machine et linguistique computationnelle. Des partenariats avec les projets étrangers sont prévus.

Le projet implique le porteur, Thibault Clérice (chercheur permanent, spécialiste d'humanités computationnelles), ainsi que Benoît Sagot (directeur de recherche, spécialiste de traitement automatique des langues) dans ALMA_{na}CH. Deux postes sont ouverts sur la première année: un poste de data librarian et un poste orienté machine learning (image vers XML). Thibault Clérice accompagnera au quotidien les deux personnes recrutées, tant du point de vue scientifique qu'établissement des partenariats.

La phase exploratoire du projet est financée pour un an.

Mission confiée

Missions :

Les missions principales comprennent l'organisation des partenariats, notamment les réunions du comité d'usagers et les relations avec les institutions de numérisation, ainsi que la mise en œuvre des priorités du comité. Il s'agit aussi d'évaluer l'état des corpus disponibles, de repérer les manques et de produire des fichiers XML TEI simples pour les systèmes de citation. Les missions secondaires, dans la limite du temps disponible, incluent l'annotation linguistique de textes en grec ancien, la préparation à la pérennisation du projet, et l'évaluation qualitative des données produites par la chaîne de traitement. Enfin, la personne participe à la vie de l'équipe et aux projets, avec des déplacements possibles pour des conférences internationales.

Pour une meilleure connaissance du sujet de recherche proposé :

- [The First Thousand Years of Greek](#), L. Muellner
- [Les outils CapiTainS, l'édition numérique et l'exploitation des textes?](#), T. Clérice
- [Continuous Integration and Unit Testing of Digital Editions](#), B. Almas, T. Clérice

Collaboration

La personne recrutée sera en contact avec les membres du projet COLaF autour des questions de schémas XML.

Principales activités

- Missions principales:
 - Organisation des partenariats, en particulier organisation des réunions du comité d'usagers, des relations avec les institutions de numérisation et traduction des priorités établies par le comité d'usagers en actions sur le terrain.
 - État de l'art des corpus ouverts et identifications des manques et documents problématiques.
 - Édition/création de XML TEI "simple" compatible avec des systèmes de citation canonique.
- Missions secondaires:
 - Annotations linguistiques de document en grec ancien
 - Préparation de la pérennisation du projet
 - Correction et évaluation des documents produits par la chaîne de production.
- Missions collectives : participation à la vie de l'équipe et des projets, y compris pour des conférences ou rencontres internationales à l'étranger.

Compétences

Compétences techniques et niveau requis :

- Connaissances en littérature grecque, en particulier du 1er millénaire de notre ère.
 - Un diplôme de master ou une thèse en lettres classiques centrée sur cette période sont un atout.
- Connaissances en langue grecque, pour l'annotation de corpus .
- Compétence en XML TEI (encodage basique)

Langues :

- Niveau de français B2 ou C1 minimum
- Niveau d'anglais B2
- Grec Ancien (Niveau master ou supérieur)

Compétences relationnelles :

- Bonnes capacités d'organisation.
- Bon relationnel.

Compétences additionnelles appréciées :

- La connaissance des corpus ouverts et fermés en grec ancien est un plus.
- La connaissance de git est un plus.
- Un intérêt marqué pour la science ouverte

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 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

Informations générales

- **Thème/Domaine :** Langue, parole et audio
Production, traitement et analyse des données (BAP D)
- **Ville :** Paris
- **Centre Inria :** [Centre Inria de Paris](#)
- **Date de prise de fonction souhaitée :** 2025-09-01
- **Durée de contrat :** 12 mois
- **Date limite pour postuler :** 2025-08-31

Contacts

- **Équipe Inria :** [ALMANACH](#)
- **Recruteur :**
Clerice Thibault / thibault.clerice@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

- Se sentir à l'aise dans un environnement interdisciplinaire, aimer apprendre et écouter sont des qualités essentielles pour réussir cette mission.
- Intéressé par les problématiques de sciences ouvertes
- Une thèse ou un mémoire portant sur un corpus du 1er millénaire de notre ère est un plus.

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.