



Offer #2024-08496

Data Librarian pour les corpus de Grec Ancien

The offer description below is in French

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : Temporary scientific engineer

Level of experience : Recently graduated

Context

Ce poste de data librarian est ouvert au sein de l'équipe-projet ALMAnaCH du Centre Inria de Paris. ALMAnaCH 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 ALMAnaCH.

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 ALMAnaCH. 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.

Assignment

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.

Main activities

- 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.

Skills

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

Benefits package

- 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

General Information

- **Theme/Domain** : Language, Speech and Audio
Data production, processing, analysis (BAP D)
- **Town/city** : Paris
- **Inria Center** : [Centre Inria de Paris](#)
- **Starting date** : 2025-03-01
- **Duration of contract** : 12 months
- **Deadline to apply** : 2025-02-20

Contacts

- **Inria Team** : [ALMANACH](#)
- **Recruiter** :
Clerice Thibault / thibault.clerice@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

- 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.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.