

Offre d'Ingénieur-e d'études en production, traitement, analyse de données et enquêtes (D2A41)

Type de contrat : CDD de 5 mois

Date de début souhaité : 1^{er} février 2025

Quotité : 100%, possibilité de temps partiel si l'expérience le justifie

Niveau requis : master + 2 ans d'expérience minimum

Rémunération : de 2763 à 3494 euros bruts par mois selon expérience

Laboratoire d'accueil : UMR5190 – Laboratoire de recherche historique Rhône-Alpes

Adresse : MSH Lyon Saint-Etienne - 14, avenue Berthelot, 69363 Lyon Cedex 07

Site web : <https://larhra.fr/>

Mission

L'Université Lumière Lyon 2 recrute un ingénieur ou une ingénieure d'études dans le cadre du projet OBJECTive. En relation étroite avec la porteuse du projet, l'ingénieur-e mettra en œuvre le protocole de collecte et de traitement de données du corpus, participera au choix, à la modélisation, mobilisera et adaptera les méthodes propres aux humanités numériques en relation avec les problématiques du projet.

À partir d'un corpus déjà établi, l'ingénieur-e recruté-e devra organiser et transcrire les catalogues de vente (pour la plupart disponibles en ligne) et créer un flux de travail afin de classer automatiquement les données pour les adapter au modèle de base de données. Des méthodes de traitement du langage naturel seront utilisées et l'indexation sera effectuée en liaison avec l'initiatrice du projet. La base de données sera modélisée par le LARHRA et Kleiolab dans le cadre contextuel du CIDOC CRM, une norme ISO (ISO 21127) utilisée pour le patrimoine culturel, afin de produire et de publier des informations de haute qualité, sémantiquement cohérentes et réutilisables. Elle sera hébergée dans l'environnement de recherche virtuel Geovistory (<https://geovistory.org>).

Le projet OBJECTive

OBJECTive tend à développer une base de données similaire au Getty Provenance Index, mais spécifiquement conçue pour répondre à l'hétérogénéité des informations sur les arts décoratifs et les acteurs impliqués dans leur circulation. Cette base de données contribuera au développement des études sur le marché de l'art et les provenances, un domaine interdisciplinaire à la croisée de l'histoire de l'art, de l'histoire économique et de l'histoire sociale.

Dans cette phase initiale - dans le cadre du financement ACCESS-ERC - nous visons à réunir, trier et intégrer des données provenant de 100 catalogues de ventes aux enchères parisiennes (environ 20 000 objets) datant des années 1820 à 1890, et à démontrer comment des objets d'origines diverses ont été collectés, valorisés et mis en circulation.

Activités principales

- Mettre en œuvre la démarche méthodologique et le protocole définis.
- Assurer la conduite des opérations de production, de recueil et de remédiation des données, assurer le bilan de collecte.
- Organiser le traitement des données et participer à l'analyse des résultats.
- Participer à la modélisation des données, à l'importation et à l'alimentation de la base de données Geovistory.
- Participer à la diffusion des protocoles et des résultats auprès de la communauté scientifique, en collaboration avec l'initiatrice du projet et les équipes impliquées (colloques, rapports de recherches...).
- Établir les collaborations pertinentes d'un point de vue scientifique et technologique permettant de réaliser au mieux les objectifs du projet.
- Rédiger des rapports d'avancement, des rapports méthodologiques réguliers et de la documentation scientifique.
- Rédiger des tutoriels à destination des membres du projet.
- Établir et mettre à jour le Plan de gestion des données (PGD) et définir les protocoles de sauvegarde des données.

Conditions particulières d'exercice

- Obligation de respecter le secret statistique ou professionnel dans le cadre législatif existant
- Possibilité de déplacements en France ou à l'étranger (exceptionnel)

Connaissances

- Connaissance d'un domaine en sciences humaines de préférence en histoire ou histoire de l'art - avec une ouverture vers les collaborations interdisciplinaires – et notions d'économétrie seraient un plus.
- Méthodes et outils en production de données en particulier extraction et transcription des données (Json, XML et formats TEI et ALTO, OCR, structuration des données brutes, alignement de données)
- Méthodes et outils en modélisation de données : standards, technologies et outils du web sémantique et des données ouvertes liées (RDF, SPARQL, ontologies, vocabulaires contrôlés)
- Méthodes et outils en Traitement Automatique des Langues : entités nommées, lemmatisation, CQL, word embedding
- Méthodes et outils en traitement, analyse et visualisation de données
- Systèmes de gestion de base de données relationnelles
- Cadre légal et déontologique
- Techniques de présentation écrite et orale.
- Langue anglaise : B2 minimum (cadre européen commun de référence pour les langues).

Compétences opérationnelles

- Savoir organiser un corpus de données, en vue de son exploitation
- Savoir exploiter les bases de données relevant de différentes technologies (SQL, XML, RDF, graphes).
- Savoir manipuler des données à l'aide du langage Python et des bibliothèques et modules dédiés, notamment NumPy et Pandas
- Savoir transcrire automatiquement des données textuelles (OCR)
- Savoir nettoyer des données (expressions régulières)
- Savoir appliquer des méthodes d'analyses statistiques et spatiales
- Savoir s'insérer dans un collectif de travail tout en sachant travailler en autonomie.
- Savoir travailler en équipe

Contexte de travail

Le LARHRA est une UMR en histoire moderne et contemporaine, multi-tutelles (5) et multi-sites, composée d'une centaine de permanents et d'autant de doctorants. L'ingénieur-e travaillera au sein de l'Axe de recherche en histoire numérique (ARHN) du LARHRA.

Au sein de l'ARHN, l'accent est mis sur les recherches conduites à l'aide d'outils numériques et sur la méthodologie numérique appliquée à la recherche. Avec plus de douze ans d'expérience dans ce domaine, fondée sur une collaboration étroite entre chercheurs et enseignants-chercheurs, ingénieurs et doctorants, l'axe est actuellement composé de 6 ITA en développement web et gestion de projet, d'une douzaine de chercheurs et enseignants chercheurs, ainsi que plusieurs doctorants. L'expertise de l'axe porte particulièrement sur la production de données par les projets au sein d'un environnement de recherche collaboratif (projet *geovistory.org*), leur ouverture et connexion avec d'autres ressources au sens des données ouvertes et du web sémantique (*linked open data*), ainsi que sur la modélisation collaborative des données et les ontologies pour l'histoire.

Modalités de candidature

Envoyer par mail une lettre de motivation et un CV détaillé à Camille Mestdagh camille.mestdagh@univ-lyon2.fr et Vincent Alamertery vincent.alamertery@ens-lyon.fr

Date limite d'envoi : le **3 janvier 2025**

Les entretiens auront lieu à **Lyon le 16 janvier 2025**

Contact : Camille Mestdagh camille.mestdagh@univ-lyon2.fr