

## Offre d'emploi :

### Poste de chargé de recherche post-doctoral (H / F) au sein du LabEx IMU pour le projet HyperThésau

<b>CDD</b>	<b>Type de poste :</b> Chargé de recherche post-doctoral	<b>Formation requise :</b> Doctorat en informatique
<b>Durée du contrat :</b> 12 mois	<b>Rémunération :</b> Selon expérience, sur la base de la grille de rémunération en vigueur à l'Université de Lyon	<b>Prise de fonction :</b> 01/09/2019

L'Université de Lyon est un site académique d'excellence à vocation mondiale. Labellisée IDEX en 2017, elle se situe au cœur de la Région Auvergne-Rhône-Alpes, sur le bassin Lyon Saint-Étienne.

Structurée autour de 12 établissements membres et d'établissements associés, l'Université de Lyon porte trois ambitions majeures :

- Concevoir une grande université attractive, responsable, bénéficiant d'une réputation d'excellence et d'innovation, et dotée d'un fort rayonnement international ;
- Proposer une offre de formation et des axes de recherche d'excellence, en adéquation avec les attentes et les mutations de la société ;
- Développer et valoriser la dynamique du site Lyon Saint-Étienne, en lien avec tous les acteurs du territoire : citoyens, associations, entreprises, collectivités locales (métropoles de Lyon et Saint-Étienne, Région Auvergne-Rhône-Alpes, autres collectivités territoriales).

## DESCRIPTION DU POSTE

Concevoir et mettre en œuvre un lac de données archéologiques et des méthodes de fouille de données associées

### Positionnement du poste

CDD chercheur au sein du [LabEx IMU](#), sous l'autorité hiérarchique de Gilles Gesquière (responsable scientifique et technique du LabEx IMU) et sous la responsabilité de Marie-Odile Rousset (chargée de recherche, [laboratoire Archéorient](#)) et de Sabine Loudcher (professeur, [Laboratoire ERIC](#)), dans le cadre du projet « HyperThésau – Hyper thesaurus et lacs de données : fouiller la ville et ses archives archéologiques ».

**Mots-clés :** lac de données ; archéologie ; thesaurus ; fouille de données ; science participative

**Organisation du travail :** temps complet

Le projet est fortement interdisciplinaire, au sein d'un consortium très motivé pour faire converger les réflexions et les pratiques scientifiques. La mission s'exercera en étroite collaboration avec une post-doctorante en archéologie, avec un ingénieur d'étude chargé de faire vivre au quotidien l'interdisciplinarité du projet, avec des laboratoires et des praticiens de l'archéologie et avec le Pôle Système d'Information et Réseaux de la Maison de l'Orient et de la Méditerranée. À ERIC, le/la post-doctorant·e sera intégré·e à une groupe de travail « data lake » comprenant les encadrant·es et deux doctorants.

**Lieu de travail**

Université Lumière Lyon 2

Unité de recherche : ERIC

Campus Porte des Alpes, 5 avenue Pierre Mendès France, 69500 Bron

**Missions principales**

Le projet HyperThesau porte, d'une part, sur la constitution d'un thesaurus-métier de l'archéologie prenant en compte l'ensemble de la chaîne opératoire scientifique du terrain à la publication et, d'autre part, sur la conception et la mise en place d'un mini-système d'information pour les données archéologiques.

L'objectif du projet de post-doctorat en informatique sera, dans un premier temps, de concevoir et de mettre en œuvre un lac de données (*data lake*) pour stocker et interroger des données archéologiques hétérogènes. Dans un second temps, il s'agira de déployer et d'adapter des méthodes de fouille de données et de textes, en relation étroite avec les archéologues et leurs besoins, de manière à proposer un enrichissement automatique des données primaires, qui sera soumis à validation.

La mission sera décomposée en six étapes :

1. audit des données archéologiques à prendre en compte ; revue de la littérature consacrée aux lacs de données ;
2. conception, mise en œuvre d'un premier prototype de stockage et de recherche de documents compatible avec les outils mis à disposition par le TGIR HumaNum (preuve de concept) ;
3. conception d'un modèle de métadonnées afin de permettre des interrogations avancées et la fouille de données et de textes ;
4. recueil des besoins d'analyse informatique des archéologues ; revue de la littérature consacrée à la fouille de données en archéologie ;

5. conception de méthodes de fouille adaptées aux données stockées, par exemple, classifieurs ensemblistes capables de travailler sur plusieurs types de documents et données ;
6. déploiement de l'architecture dans l'infrastructure du TGIR HumaNum

## Savoirs

Bases de données, entrepôts de données, lacs de données, gestion de métadonnées, fouille de données, fouille de textes

## Savoir-faire et méthodologie

Bases de données relationnelles et NoSQL, outils d'ingestion de données (Flink, Sqoop...) ou d'ETL, outils d'extraction de métadonnées (Flume, Tika...), plateformes de calcul distribué (Hadoop, Spark...)

## Savoir être

Travail en équipe dans un environnement interdisciplinaire, communication et vulgarisation scientifique, écoute et analyse des besoins, capacité de synthèse

## CANDIDATURES

---

### Renseignements sur le poste (personnes à contacter) :

#### Jérôme Darmont et Sabine Loudcher

[jerome.darmont@univ-lyon2.fr](mailto:jerome.darmont@univ-lyon2.fr), [sabine.loudcher@univ-lyon2.fr](mailto:sabine.loudcher@univ-lyon2.fr)

### Envoi des candidatures

lettre de motivation + CV **exclusivement par e-mail avant le 1<sup>er</sup> avril minuit** à [rh.recrutement@universite-lyon.fr](mailto:rh.recrutement@universite-lyon.fr) , [sabine.loudcher@univ-lyon2.fr](mailto:sabine.loudcher@univ-lyon2.fr) **et** [jerome.darmont@univ-lyon2.fr](mailto:jerome.darmont@univ-lyon2.fr)

### Entretien

Les entretiens se dérouleront à Lyon ou par visioconférence à partir du 01/05/2019 devant un jury pluridisciplinaire composé de chercheurs en informatique, en archéologie et de l'ingénieur d'études chargé de l'animation du projet.

## Job offer:

# Postdoctoral researcher (M/W) within the Excellence Lab named Urban Worlds Intelligence (IMU) and project HyperThesau

<b>Job type:</b> Postdoctoral researcher		<b>Required education:</b> PhD in Computer Science
<b>Duration:</b> 12 months	<b>Salary:</b> Depending on experience, Selon expérience, with respect to Université de Lyon's remuneration policy	<b>Starting date:</b> September 1, 2019

The Université de Lyon is an international, excellence academic site that earned the IDEX label in 2017. It is located at the heart of the Auvergne-Rhône-Alpes region, in the Lyon Saint-Étienne area (France). Structured around 12 member and associate member institutions, the carries three major ambitions:

- build a great, attractive university that is responsible, benefits from a reputation of excellence ad innovation, and wielding a strong international aura;
- propose excellence education and research axes, in line with the society's and mutations de la société;
- develop dynamics within the Lyon Saint-Étienne site, in collaboration with all actors of the territory: citizens, organizations, companies, local authorities (Lyon and Saint-Étienne metropolises, Auvergne-Rhône-Alpes Regions, other local authorities).

## JOB DESCRIPTION

Designing and mining an archaeological data lake

### Job position

Researcher with [LabEx IMU](#), under the hierarchical authority of Mr Gilles Gesquière (technical and scientific head of LabEx IMU) and under the responsibility of Marie-Odile Rousset (researcher, [Archéorient lab](#)) and Sabine Loudcher (professor, [ERIC lab](#)), within the HyperThesau project ("Hyper thesaurus and data lake: mine the city and its archeological archives").

**Keywords:** data lakes; archaeology; thesaurus; data mining; participative science



**Work organization:** Full time

The project is highly interdisciplinary and takes place within a consortium that is much motivated to have thoughts and scientific practices converge. The mission will be performed in close collaboration with another postdoctoral fellow in archaeology, an engineer in charge of promoting interdisciplinarity within the project, archaeological labs and field actors and the Networks and Information Systems service at the Maison de l'Orient et de la Méditerranée. At ERIC, the postdoctoral fellow will be part of a data lake working group including the coordinators and 2 PhD students.

**Workplace**

Université Lumière Lyon 2

Research center: ERIC

Campus Porte des Alpes, 5 avenue Pierre Mendès France, F-69500 Bron

**Main missions**

The HyperThesau project aims, on one hand, at building an archeological thesaurus that takes the whole scientific process into account, from the field to publication and, on the other hand, at designing and implementing an information system for archaeological data.

The objective of this computer science postdoc project will be first to design and implement a data lake for storing and querying heterogeneous archaeological data. Second, the postdoctoral fellow will adapt and deploy data and text mining methods, in close relationship with archeologists and their needs, so as to propose an automatic enrichment of primary data, which will be submitted to validation.

The mission will be decomposed into 6 steps:

7. audit of archeological data; data lake literature review;
8. design and implementation of a first prototype for storing and searching documents that is compatible to the tools available at the HumaNuM research infrastructure (proof of concept);
9. design of a metadata model to allow advanced queries, data and text mining;
10. collection of the archeologists' analysis needs; data mining in archaeology literature review;
11. design of data mining methods that are adapted to the data stored in the lake, e.g., ensemble classifiers able to operate on various types of documents and data;

12. deploy the architecture within HumaNum's infrastructure.

## Required knowledge

Databases, Data warehouses, Data lakes, Metadata management, Data mining, Text mining

## Know-hows and methodology

Relational and NoSQL databases, Data ingestion tools (Flink, Sqoop...) and/or ETL frameworks, Metadata extraction tools (Flume, Tika...), Distributed computing frameworks (Hadoop, Spark...)

## Soft skills

Working in a team within an interdisciplinary environment, Scientific communication and mediation, Collection and analysis of users' needs, Synthesis ability

## APPLICATION

---

### Contact people: Jérôme Darmont and Sabine Loudcher

[jerome.darmont@univ-lyon2.fr](mailto:jerome.darmont@univ-lyon2.fr), [sabine.loudcher@univ-lyon2.fr](mailto:sabine.loudcher@univ-lyon2.fr)

### Application

Motivation letter + resume, **only by e-mail before April 1, 2019 midnight GMT** to [rh.recrutement@universite-lyon.fr](mailto:rh.recrutement@universite-lyon.fr), [sabine.loudcher@univ-lyon2.fr](mailto:sabine.loudcher@univ-lyon2.fr) **and** [jerome.darmont@univ-lyon2.fr](mailto:jerome.darmont@univ-lyon2.fr)

### Interview

Interviews may take place either in Lyon or by videoconference from the 1<sup>st</sup> May 2019, in front of a multidisciplinary jury made up of researchers in computer science, in archeology, and the engineer in charge of managing of project.