

## SUJET STAGE MASTER

**Titre :** Lac de données pour l'archéologie

**Mots-clés :** Lacs de données (*data lakes*), Métadonnées

**Noms des encadrants :**

Jérôme Darmont (laboratoire ERIC), Sabine Loudcher (laboratoire ERIC), Miquel Molist (Université autonome de Barcelone), Gabriel de Prado Cordero (musée d'Archéologie de Catalogne)

**Lieu :**

Laboratoire ERIC (Université Lyon 2, Campus Porte des Alpes, Bron) / courts séjours en Catalogne

**Durée :**

5 mois à partir de mars ou avril 2021

**Rémunération :**

Indemnités de stage légales (3,60 euros par heure pour 35 heures de travail par semaine, soit environ 550 € par mois)

**Sujet :**

L'université Lumière Lyon 2 et l'université autonome de Barcelone travaillent ensemble depuis 2018 dans le cadre d'un projet visant à l'élaboration d'un lac de données en archéologie (projet HyperThésau du Labex IMU). Même si les données de l'archéologie sont aujourd'hui numériques ou ont été numérisées (carnets de fouille incluant texte, plans, croquis, photos...), il n'existe pas de norme pour ces données, ni de consensus sur les systèmes d'information archéologiques (SIA) : tableur bureautique, base de données FileMaker, serveur SQL propriétaire ou libre, etc. Pour maîtriser cette hétérogénéité, gérer la qualité et faciliter le partage et l'analyse, les partenaires du projet ont décidé d'utiliser le concept de « lac de données » pour conserver celles-ci, primaires ou enrichies, dans leur forme originelle.

Après avoir démontré l'intérêt et la faisabilité d'utiliser un lac de données dans le contexte de l'archéologie, les partenaires du projet (et en particulier le laboratoire ERIC, le laboratoire Archéorient et le musée d'archéologie de Catalogne) souhaitent poursuivre avec la conception et la mise en place d'un lac pour les données du site d'Ullastret du musée d'archéologie de Catalogne.

**L'objectif du stage en informatique** sera, dans un premier temps, de concevoir et de mettre en œuvre le lac de données pour stocker et interroger des données archéologiques hétérogènes. Dans un second temps, il s'agira d'expérimenter l'utilisation de méthodes d'Intelligence Artificielle pour extraire depuis les données primaires des informations qui permettent de les enrichir et de constituer les métadonnées nécessaires à l'analyse des données hétérogènes réunies dans le lac.

**Les missions confiées au stagiaire seront :**

1. la compréhension des données archéologiques à prendre en compte et une familiarisation avec la littérature consacrée aux lacs de données ;

2. la compréhension et l'appropriation d'un lac de données déjà implémenté dans le cadre d'un premier projet en archéologie ;
3. la conception du lac de données, du système de métadonnées et le choix des technologies à utiliser pour la création du lac ;
4. l'implémentation du lac de données et du modèle de métadonnées, en particulier sémantiques ;
5. l'alimentation du lac avec les données primaires dans leur forme originelle ;
6. l'alimentation du lac avec la documentation ou la transcription des données primaires.

**Si le temps le permet, le stagiaire pourra également travailler sur :**

7. l'utilisation de méthodes d'Intelligence Artificielle pour extraire des données primaires des informations qui permettent de constituer de nouvelles métadonnées ;
8. la création d'une interface pour l'alimentation du lac par les archéologues.

**Compétences requises :**

Le sujet de stage s'adresse à des étudiant·es en 1<sup>re</sup> ou 2<sup>e</sup> année de master (ou équivalent) en informatique décisionnelle ou en sciences des données. Des compétences en bases de données, en traitement des données massives ou en technologies liées aux *big data* seront particulièrement appréciées.

**Contact :**

Merci d'adresser, avant le 15 décembre 2020, votre candidature avec un CV, une lettre de motivation ainsi que vos notes de l'année universitaire en cours et de l'année dernière à [jerome.darmont@univ-lyon2.fr](mailto:jerome.darmont@univ-lyon2.fr), [sabine.loudcher@univ-lyon2.fr](mailto:sabine.loudcher@univ-lyon2.fr) et [jean-pierre.girard@mom.fr](mailto:jean-pierre.girard@mom.fr)

Les candidat·es retenu·es seront convoqué·es pour un entretien en visioconférence en janvier.