



Cités, Territoires, Environnement et Sociétés
CNRS-Université de Tours

L A T

Laboratoire Archéologie et Territoires

Offre de contrat post-doctoral

PUBLICATION DE DONNEES ARCHEOLOGIQUES DANS LE WEB DE DONNEES

Unité de recherche de rattachement :

- UMR 7324 CITERES–Laboratoire Archéologie et Territoires, Université de Tours, CNRS.

En collaboration avec :

- EA 6300 Laboratoire d'Informatique, Université de Tours.
- Intelligence des Patrimoines, UMR 7323 CESR, Université de Tours, CNRS.

Lieu : Tours (France)

Durée : 12 mois (2018-2019)

Rémunération : 2200€ net/mois

Objectifs

Ce projet post-doctorat international d'un an a pour objectif la publication des données archéologiques dans le web des données à partir des systèmes d'informations du Laboratoire Archéologie et Territoires de l'UMR CITERES, en collaboration avec le LI et le CESR pour l'ARD Intelligence des patrimoines (IPat). Il s'inscrit au cœur de la plate-forme numérique d'IPat. , l'enjeu de ce post-doctorat est de contribuer à l'émergence du domaine « patrimoine » dans les données interconnectées du web.

Dans la communauté archéologique, on constate qu'une masse importante de données reste difficilement accessible et risque d'être perdue. Ce n'est pas fondamentalement différent des autres disciplines des SHS. En revanche, cela correspond à une originalité propre à la pratique archéologique qui ne permet pas le renouvellement de l'expérience une fois la fouille effectuée et donne, de ce fait, une valeur accrue aux archives de fouilles et une responsabilité supplémentaire aux archéologues. La préservation de cette masse de données scientifiques en la rendant accessible est fondamentale pour la connaissance des patrimoines que ces archives documentent et pour la mémoire de la discipline elle-même dont la publication et le partage des données seront inéluctablement numériques.

Le programme d'archéologie urbaine initié à Tours dès 1970 constitue un cas emblématique en offrant une documentation homogène dont la structure a été conservée lors de l'informatisation du système d'information. Pour l'ouverture des données sur le web sémantique, une notation à cinq étoiles (*Five stars Linked Open Data*) a été mise en place par Berners-Lee en 2010 permettant d'évaluer la qualité d'un jeu de données au regard de son accessibilité sur le web des données. Cette qualité s'évalue par l'accessibilité (en ligne), la structuration, l'utilisation de formats non-propriétaires et répondant à un standard et enfin sur l'interopérabilité. Aujourd'hui, CITERES-LAT dispose de plusieurs bases de données chacune à un stade différent de développement vers l'interopérabilité selon le moment où elles ont été conçues. Notre objectif n'est pas de les homogénéiser mais de les faire dialoguer, entre elles et avec d'autres, de manière à pouvoir les interroger ensemble sans modifier les habitudes des utilisateurs.

La première étape de ce processus consiste à aligner le contenu de ces bases de données avec des vocabulaires partagés et normés (PACTOLS, ensemble de thésaurus multilingues ISO 25964-1:2011). L'étape suivante est le passage à un niveau supérieur d'abstraction afin de permettre l'interrogation croisée de systèmes dont la granularité des entités et des descripteurs sont différents. Pour mettre en correspondance les structures de données et les terminologies utilisées nous faisons appel au CIDOC-CRM (ISO 21127:2014) qui est l'ontologie de domaine pour les données patrimoniales, développée par le CRM *Special Interest Group* coordonné par l'équipe ICS-FORTH à Héraklion avec laquelle une collaboration est engagée. Ensuite, le choix effectué, fondé

sur l'expertise de Béatrice Markhoff au LI en collaboration avec l'Université de Bozen-Bolzano (Italie), est de mettre en place une architecture logicielle, composée de plusieurs briques, pour une interrogation dynamique des systèmes d'information existants via l'ontologie plutôt que de stocker l'ensemble des triplets RDF de manière statique dans des entrepôts de données (triplestore).

Les travaux déjà effectués ont démontré la faisabilité de cette architecture à partir de l'exposition du système ArSol (Archives du Sol <http://arsol.univ-tours.fr/> et <http://citeres.univ-tours.fr/spip.php?article505>) de CITERES-LAT. Cette première preuve de concept constitue la base de ce projet dont l'objectif est bien entendu de rendre interopérable l'ensemble de nos systèmes d'information par ce moyen mais également de faire la démonstration de l'interopérabilité auprès de la communauté archéologique en montrant comment plusieurs systèmes de formats, de structures et de granularités différents peuvent être interrogés de manière croisée depuis un niveau ontologique.

Contribution de la thématique au rayonnement, partenariats et réseaux scientifiques aux échelles nationale et internationale

Ce projet permettra de donner de l'ampleur aux expériences menées en les portant aux niveaux des infrastructures nationales et européennes, en place ou en cours de création, et en créant des sources de données archéologiques dans le graphe du web sémantique.

L'ambition est de démontrer que l'exposition sur le web des données d'informations structurées et interprétées permet d'envisager leur utilisation et leur réutilisation en ouvrant de nouvelles perspectives de recherche. Au-delà, l'enjeu porte sur l'articulation des publications scientifiques avec les données en explorant les nouvelles perspectives qu'offre la mise en œuvre des inférences ontologiques.

Le post-doctorant contribuera à la structuration de la recherche régionale puisqu'il participera à la plate-forme de données hétérogènes en cours de constitution d'Intelligence des Patrimoine. Il bénéficiera de l'insertion dans le réseau national du consortium Mémoire des Archéologues et des Sites Archéologiques (MASA) de la TGIR Huma-Num, coordonné par X. Rodier (CITERES-LAT) pour la MSH Val de Loire de 2017 à 2020. En outre, le projet apportera de la matière à deux projets européens : ARIADNE, *Advanced Research Infrastructure for Archaeological Dataset Networking in Europe* (7e PCRD), en alimentant la plate-forme (<http://portal.ariadne-infrastructure.eu/>) par les données publiées, et PARTHENOS, *Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies* (H2020), auxquels le consortium MASA contribue via la TGIR Huma-Num pour la mise en œuvre du CIDOC-CRM à l'échelle de données de fouilles archéologiques.

Enfin, ce projet s'insère également dans l'action COST *Saving European Archaeology from the Digital Dark Age*, SEADDA, soumise en septembre à l'Union Européenne.

Encadrement

Le post-doctorant sera accueilli à CITERES-LAT dans les locaux de la MSH Val de Loire où il disposera d'un bureau et d'un ordinateur équipé des logiciels nécessaires. Il travaillera en collaboration avec les membres du laboratoire qui développent, gèrent et publient les systèmes d'information : Philippe Husi (IR CNRS, HDR), Olivier Marlet (IE CNRS), Xavier Rodier (IR CNRS, HDR). Il bénéficiera des travaux réalisés à partir du système d'information ArSol en collaboration avec Béatrice Markhoff (MCF HDR) du LI pour la publication d'ArSol dans le web des données. Enfin il veillera à l'adéquation de ces travaux avec la plate-forme numérique d'Intelligence des patrimoines coordonnée par Perrine Thuringer (IR IPat).

Les candidatures sont à envoyer à xavier.rodier@univ-tours.fr avant le 1^{er} janvier 2018 minuit par courriel avec fichiers joints.

Le dossier devra comporter :

- Une lettre de motivation
- Un curriculum vitae avec liste de publications
- Un résumé long de la thèse

Veillez vous adresser à Xavier Rodier pour tout renseignement complémentaire.