

Partage et analyse de données multi-points de vues pour la préservation du patrimoine culturel

Une approche guidée par les ontologies

Financement : ANR DIGITALIS¹

Partenaires académiques : CESCМ, CRESTIC, IFAO, LIAS and XLIM

Laboratoire : LIAS (Laboratoire d'Informatique et d'Automatique pour les Systèmes)²

Localisation : Poitiers, Futuroscope, France

Date de début : Octobre 2023 (flexible)

Durée : 3 ans

Date limite d'envoi des candidatures : 26 mai 2023

Salaire mensuel net : 1825 EUR

Encadrants : Stéphane JEAN (directeur de thèse) et Mickaël BARON

(L'encadrant dont le nom est souligné est l'encadrant référent : baron@ensma.fr)

Mots clés : **ontologie, raisonnement, apprentissage automatique, données massives, patrimoine culturel**

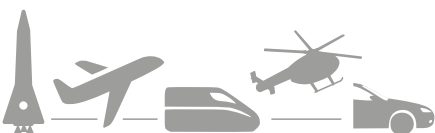
Contexte de la thèse

Analyser, comprendre et préserver l'histoire des sites, des monuments et des artefacts du patrimoine culturel nécessitent de recueillir et d'étudier des informations hétérogènes obtenues par de nombreuses techniques. Ces objets du patrimoine ont évolué au fil du temps de nombreuses manières, à travers l'érosion, la reconstruction ou la rénovation. Les connaissances acquises à partir de l'histoire des objets sont représentées par des documents, des photographies, des dessins, des mesures physiques ou des phases de construction, parfois stockées numériquement. Produire ces connaissances nécessite de nombreuses techniques qui ont également évolué avec le temps. Elles doivent être non invasives, automatisées autant que possible et répondre aux principes FAIR (Facile à trouver, Accessibles, Interopérables et Réutilisables). Dans le contexte de la recherche sur le patrimoine culturel, intégrer de telles données complexes associées aux formes et aux apparences des objets (y compris leur incertitude et leurs propriétés temporelles) pour raisonner sur le passé des objets est une tâche difficile. Lorsqu'ils existent, les outils numériques ne fournissent que des représentations partielles, déconnectées et spécifiques.

La thèse proposée se déroulera dans le contexte du projet ANR DIGITALIS.

1. <https://digitalis.humanities.science>

2. <https://www.lias-lab.fr>



Objectifs scientifiques

L'objectif du projet ANR DIGITALIS est de proposer une représentation numérique intégrée et des outils qui accompagneront la recherche archéologique. Ils faciliteront l'efficacité de la collecte et de la numérisation de données. Il devra permettre des raisonnements sur des objets du patrimoine en s'appuyant à la fois sur la collecte de données et sur la visualisation 3D de ces objets. Des vues photoréalistes et interactives seront ainsi fournies, où les données collectées sont référencées à la fois en 3D et sur des photographies enregistrées.

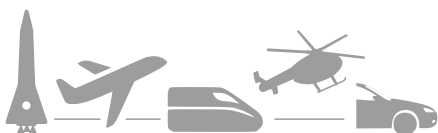
La connaissance sera structurée via un modèle de connaissances qui facilitera le partage en s'appuyant sur des standards existants. Des techniques liées à l'apprentissage automatique pourront être utilisées pour inférer de nouveaux faits historiques.

Plus précisément, les principaux objectifs de la thèse sont décrits ci-dessous.

Définition d'un modèle de données standardisé dédié à la représentation numérique des connaissances historiques. Ce modèle devra permettre de définir des prises de mesures, des annotations dédiées, la localisation des informations sur les photographies et la représentation 3D. Ce modèle sera le cœur numérique du projet. La définition de ce modèle soulève trois défis. (i) Il devra être basé sur des standards existants (en particulier, CIDOC-CRM [1] et Dublin Core <https://dublincore.org>) et les modèles proposés dans des projets connexes (par exemple, OpenArcheo ([2]), ArcheoGrid (<https://www.archeogrid.fr>), Opentheso (<https://opentheso.hypotheses.org>)) ou l'ontologie proposée par Messaoudi et al. [3] qui est dédiée à l'annotation d'images spatialisées pour le suivi de la conservation du patrimoine culturel bâti. Cela permettra d'assurer l'interopérabilité, la traçabilité et la préservation au fil des années. (ii) Il devra prendre en charge le niveau de fiabilité d'une annotation (observation, mesure, interprétation, hypothèse, etc.), ainsi que son statut (existant ou en cours d'élaboration, accès public ou restreint) pour assurer la diffusion des connaissances, l'incertitude et l'évolution. Ces caractéristiques seront utilisées dans les processus de requête et d'inférence pour caractériser les annotations qui peuvent être déduites ou interrogées à partir des annotations produites. (iii) Il devra enfin référencer l'information sur les modèles géométriques afin que les annotations puissent être utilisées sur les photographies et les supports 2D [4].

Pour alimenter la base de données numérique avec toutes les propriétés des objets pour plusieurs sites et monuments historiques. L'objectif est d'associer aux objets des informations en 3D et toute autre description technique. La source de données sera soit collectée automatiquement à partir de bases de données existantes (sélectionnées et vérifiées par des experts), soit extraite de celles déjà existantes au CESCO/IFAO. L'alimentation de la base de données de manière aussi automatique que possible constitue un défi intéressant, car les données pour chaque type d'objet sont aujourd'hui hétérogènes et distribuées dans divers systèmes. La gestion de toutes les données pour tous les objets permettra de démontrer la capacité de traiter différents cas possibles.

Inférer (semi-)automatiquement des nouveaux faits à partir des données et de la visualisation en détectant les incohérences ou en proposant de nouvelles hypothèses. L'idée principale est que l'organisation visuelle des données collectées peut fournir de nouveaux points de vue sur



l'objet et une meilleure observation générale. L'inférence à partir de bases de données basées sur des techniques d'intelligence artificielle pourra aider les archéologues à tirer de nouvelles conclusions [5].

Formuler de nouvelles déductions/conclusions historiques et produire de nouvelles connaissances pour chaque site et monument du patrimoine. L'objectif est d'utiliser le modèle de données pour tester et valider (ou invalider) de nouvelles interprétations architecturales, archéologiques ou historiques, grâce à la structure des connaissances, aux outils de visualisation et à certaines propositions découlant des déductions et de la recherche de contradictions dans les données existantes.

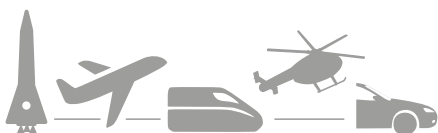
Cas d'application

Quatre sujets expérimentaux d'étude seront considérés pour valider les résultats de cette thèse.

- **Les peintures de Saint-Savin** surnommée « la Sixtine de l'art roman » par André Malraux, inscrite au patrimoine mondial de l'UNESCO, l'église de l'abbaye de Saint-Savin présente le plus important ensemble de peintures murales d'Europe occidentale des XI^e et XII^e siècles (près de 460 m² dans la nef, le chœur et la crypte).
- **L'Hypogée des Dunes** découvert au 19^e siècle à Poitiers, ce petit édifice semi-enterré mérovingien (VI^e VII^e siècle) est connu de tous les experts en art de cette période (<https://alienor.org>). Il combine architecture, restes funéraires, décors et inscriptions sur un monument de petite échelle (moins de 20 m²), marqué par de nombreux changements au fil du temps.
- **Palais des comtes de Poitiers-ducs d'Aquitaine**, lieu de résidence et d'exercice du pouvoir des comtes de Poitiers, ducs d'Aquitaine au Moyen Âge, le bâtiment a été utilisé du XI^e siècle jusqu'au début du XXI^e siècle, lorsque le palais de justice a déménagé vers de nouveaux locaux judiciaires. Le site d'environ 6000 m² est réparti sur plusieurs niveaux, partiellement classé comme Monument Historique depuis 1862.
- **Céramiques du Caire** répertorié dans la bibliothèque du Caire (<https://ifao.egnet.net>) comme l'une des séries les plus représentatives d'artefacts de l'Égypte ancienne, allant de la période néolithique à la période médiévale, avec une analyse typologique et technique, incluant les aspects physiques des matériaux utilisés, y compris l'apparence et les peintures.

Profil du candidat

Le candidat doit détenir un diplôme de Master en Informatique ou un diplôme d'ingénieur spécialité Informatique. Une bonne connaissance de la programmation (Java, JavaScript et Python), des bibliothèques/frameworks (frameworks web, bibliothèque d'apprentissage automatique) et des solutions de stockage (stockage RDF et SGBDR) est requise. Un bon niveau en français et en anglais est fondamental. **Une motivation pour le domaine historique est fortement recommandée.**



Documents à fournir

- Curriculum Vitae;
- Lettre de motivation;
- Notes de Master ou équivalent;
- Tout autre document jugé utile pour enrichir le dossier de candidature (e.g. lettres de recommandation, articles scientifiques, rapport de master).

Références

- [1] M. Doerr, "The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata", *AI Mag.*, vol. 24, no. 3, pp. 75–92, 2003. DOI: 10.1609/aimag.v24i3.1720. [Online]. Available: <https://doi.org/10.1609/aimag.v24i3.1720>.
- [2] O. Marlet, T. Francart, B. Markhoff, and X. Rodier, "OpenArchaeo for Usable Semantic Interoperability", in *ODOCH 2019 @CAiSE 2019*, Jun. 2019. [Online]. Available: <https://hal.science/hal-02389929>.
- [3] T. Messaoudi, P. Véron, G. Halin, and L. de Luca, "An ontological model for the reality-based 3D annotation of heritage building conservation state", in *Journal of Cultural Heritage*, 2018.
- [4] L. D. Luca, "Towards the Semantic-aware 3D Digitisation of Architectural Heritage: The "Notre-Dame de Paris" Digital Twin Project", in *SUMAC, ACM Multimedia*, 2020.
- [5] F. G. A.E.A. Djebri A. Tettamanzi, "Linking and Negotiating Uncertainty Theories Over Linked Data", in *Workshop of the World Wide Web Conference*, 2019.

