

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

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.

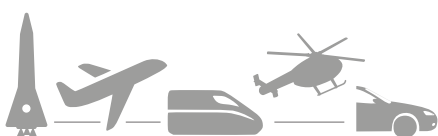
La connaissance sera structurée via un modèle de connaissances qui facilitera le partage en s'appuyant sur des standards existants. Des techniques liées à l'apprentissage automatique pourront être utilisées pour inférer de nouveaux faits historiques.

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 CESC/M/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.

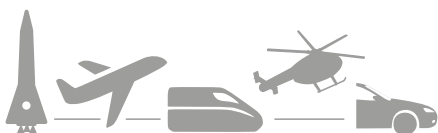
Cas d'application

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.**

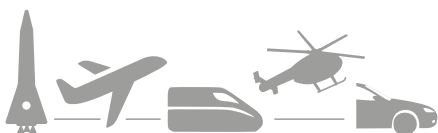


Documents à fournir

- 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.



Data Sharing and Analysis from Multiple Point of Views for Preserving Cultural Heritage

An Ontology-based Approach

Funding : ANR DIGITALIS¹

Academic partners : CESCO, CRESTIC, IFAO, LIAS and XLIM

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

Where : Poitiers, Futuroscope, France

Start : October 2023 (flexible)

Duration : 3 years

Application deadline : May 26, 2023

Monthly net salary : 1825 EUR

Supervisors : Stéphane JEAN (thesis director) and Mickaël BARON

(The supervisor whose name is underlined is the reference person : baron@ensma.fr)

Keywords : ontology, reasoning, machine learning, big data, cultural heritage

Context of the thesis

Analyzing, understanding, and preserving the history of sites, monuments, or artifacts in cultural heritage, requires gathering and studying heterogeneous information, obtained by many types of procedures. These objects have evolved in many ways, from erosion, reconstruction, renovation, painted with many layers, etc. The knowledge gathered from the objects' life corresponds to documents, photographs, drawings, physical measurements, or construction phases, sometimes digitally stored. Producing this knowledge corresponds to many procedures that have also evolved. They need to be noninvasive, automated as often as possible, and fulfill FAIR principles (Findable, Accessible, Interoperable, Reusable). In the context of cultural heritage research, mentally integrating such complex data associated together with object shapes and appearances (including uncertainty and temporal properties) for reasoning on the objects' past is a difficult task. When they exist, digital tools only provide partial, disconnected, and specific representations. The proposed thesis will take place in the context of the French ANR project DIGITALIS (<https://digitalis.humanities.science/>).

Scientific objectives

The goal is to propose an integrated digital representation and tools that will support archaeological research, increasing efficiency in data collection and digitization, providing automatic reasoning, and visualizing objects in 3D with photo realistic and interactive views, where the gathered data is referenced both in 3D and on registered photographs.

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

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

The global knowledge will be structured with a model representation that will be shared within existing standards. Artificial Intelligence procedures will be employed for proposing new inference-based assumptions to historians.

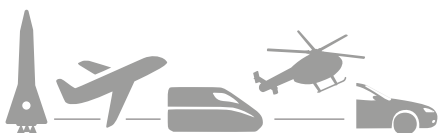
More precisely, the main objectives of the thesis are the following.

Definition of a standardized data model dedicated to the digital representation of knowledge including measurements, dedicated annotations, localization of information on both photographs and 3D representation. This model will be the digital heart of the project. The definition of this model raises three challenges. (i) It should be based on existing standards (in particular, CIDOC-CRM [1] and the Dublin Core <https://dublincore.org>) and models proposed in related projects (e.g, OpenArcheo [2], ArcheoGrid <https://www.archeogrid.fr>, Opentheso <https://opentheso.hypotheses.org>) or the domain ontology, proposed by Messaoudi et al. [3], dedicated to the annotation of images for the characterization and the preservation of built cultural heritage. This would ensure interoperability, traceability and long-term preservation. (ii) It should support the definition of the level of reliability of an annotation (an observation, measurement, interpretation, assumption, etc.), as well as its status (existing or under elaboration, public or restricted access) to ensure knowledge progress, uncertainty and evolution. These characteristics will be used in the querying and deduction processes to characterize annotations that can be inferred or queried from the produced annotations. (iii) It has to reference the information on the geometric models so that the annotations can be distributed on photographs and 2D supports [4].

To fill the digital database with all the objects' properties for several historical objects and sites, to associate information and knowledge, such as photographs, 3D information, reflectance parameters, and any other technical description. The data source will be either automatically gathered from existing databases (selected/verified by experts) or extracted from already existing ones in CESC/IFAO. Filling the database as automatically as possible is an interesting challenge since, for each type of object, the data is today heterogeneous and distributed in various systems. Managing all the data for all selected objects will demonstrate the wide range of possible applications.

To (semi-)automatically inferring conclusions from data and visualization with inconsistency detections, or new assumptions proposal. The main idea is that the visual organization of the gathered data may provide new viewpoints of the object, and a better general observation. Inference from databases based on artificial intelligence techniques can help archaeologists to drive new conclusions [5].

Draw new historical deductions/conclusions for each site and produce new knowledge for each targeted site and object. The goal is to employ the data model to test and validate (or invalidate) new architectural, archaeological, or historical interpretations, thanks to the knowledge structure, the visualization tools and some proposals coming from deductions and search for contradictions in the existing data.



Experimental subjects

Four main experimental subjects of study will be considered to validate the results of this thesis.

- **Saint-Savin's paintings** called "Sixtine of roman art" by André Malraux, listed as UNESCO world heritage site, the church of Saint-Savin's abbey present the most important set of wall paintings in Western Europe from the 11th and 12th centuries (almost 460 m2 in nave, choral and crypt).
- **Hypogée des Dunes** discovered in the 19th century in Poitiers, this small semi-buried Merovingian building (6th-7th century) is known by all art experts of this period (<https://alienor.org>). It combines architecture, funeral remains, decors and inscriptions on a small-scale monument (less than 20 m2), marked by numerous changes over time.
- **Palace of Poitiers** place of residence and exercise of power of the counts of Poitiers, dukes of Aquitaine in the Middle Ages, the building was used from the 11th to the beginning of the 21st century, when the courthouse moved to the new judicial precincts. The site of about 6000 m2 is spread over several levels, partly classified as Historic Monuments since 1862.
- **Cairo ceramics** listed in Cairo library (<https://ifao.egnet.net>) as one of the most representative series of ancient Egypt artifacts, from Neolithic to Medieval periods, coming with a typological and technical analysis, with physical aspects of the used materials, including appearance and paintings.

Profile of applicant

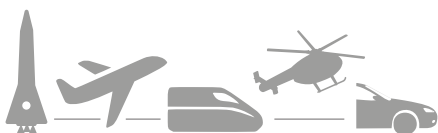
The candidate must hold a master's degree in computer science or an engineering degree. A good knowledge of programming (Java, JavaScript and Python), libraries/frameworks (web frameworks, machine learning library) and storage solutions (RDF storage and Relational DBMS) are required. Fluency in French and in English is also required. **Motivation for the historical field is strongly recommended.**

Documents to provide

- Curriculum Vitae,
- Cover letter,
- Master's degree grades or equivalent,
- Any other document deemed necessary by the candidate which can enrich the application (Master's report, letters of recommendation, etc.).

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.

