

## APPEL À PROJETS 2025 – DONNÉES & OPEN DATA

FÉDÉRATION DE RECHERCHE AGORANTIC  
«CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES»

Titre	Documentarisation et partage de données ouvertes pour l'archéologie islamique en Espagne
Acronyme	METAlbalat
Nom du/des porteur(s)	GILLOTTE, SOPHIE
Coordonnées du/de la gestionnaire de laboratoire	DIMOYAT, PIERRE (UMR 5648-CIHAM) Pierre.Dimoyat@univ-lyon2.fr
Laboratoires associés	UMR 5648-CIHAM UR ERIC (Lyon 2 et Lyon 1)
Budget demandé	8000 €
Résumé  <i>Max. 1 000 caractères espaces compris</i>	METAlbalat est centré sur la valorisation et la diffusion de jeux de données archéologiques issus du site médiéval d'Albalat en Espagne, objet de fouilles programmées depuis 2009. Ce projet vise à améliorer l'accès et la réutilisation des données existantes, notamment par la création de métadonnées enrichies et l'automatisation du traitement des données grâce à la collaboration de membres de l'UR ERIC. Les données seront hébergées en Open Access sur Nakala et enrichies via un thésaurus trilingue développé sur Opentheso, tandis que des modèles 3D seront versés au CND3D, dans le but de les rendre interrogeables et interopérables. Le projet s'inscrit dans une démarche interdisciplinaire alliant archéologie, technologie de l'information et conservation des données, contribuant à la valorisation du patrimoine islamique médiéval et au développement de la recherche future dans ce domaine.

### 1. Contexte, positionnement, objectif(s)

Implantée sur l'un des points de passage permettant de franchir le cours de la moyenne vallée du Tage (région centre-ouest de l'Espagne, province de Cáceres) à l'époque médiévale, l'agglomération fortifiée musulmane d'Albalat fait l'objet depuis 2009 de fouilles annuelles, dirigées par Sophie Gilotte, qui fonctionnent à la fois comme un chantier-école international et un programme d'archéologie pluridisciplinaire. Son enceinte, qui en constitue l'élément le plus monumental et le plus tangible, enferme une superficie d'environ deux hectares qui surplombe le fleuve.

En dépit de son extension relativement réduite, Albalat peut être considéré comme un ensemble d'une importance singulière pour l'histoire d'al-Andalus en raison de la grande qualité et variété des vestiges et du mobilier qui sont progressivement mis au jour. En effet, le lieu fut abandonné brusquement au milieu du XII<sup>e</sup> siècle, à la suite d'une attaque menée par les milices urbaines du roi Alphonse VII de León et Castille. Les destructions par incendie qui mirent définitivement fin à son occupation ont favorisé la conservation de ses constructions, qui se caractérisent par une trame urbanistique complexe dans laquelle de grandes demeures côtoient des espaces dédiés à la production artisanale (métallurgie, tabletterie, four commun, etc.). Le mobilier préservé comprend de riches assemblages céramiques et inclut également de nombreux restes organiques, fauniques, lithiques et métalliques qui contribuent à préciser l'image d'une communauté frontalière, islamisée et militarisée en pleine époque almoravide. L'analyse de ces artefacts, pour certains produits sur place et pour d'autres importés à la faveur d'échanges commerciaux à différentes échelles, nécessite un travail de fond et de longue haleine. Les enjeux scientifiques sont d'autant plus importants que

le registre matériel de cette période était jusqu'à présent très mal connu, bien moins que celui de l'époque qui la précède (taifa, XI<sup>e</sup> siècle) et de celle qui la suit (almohade, 2<sup>de</sup> moitié XII<sup>e</sup>-1<sup>re</sup> moitié XIII<sup>e</sup> siècles).

À ce jour, les différentes publications de l'équipe (catalogue d'exposition, articles de revues, chapitres d'ouvrage et notices variées) rendent compte d'une partie infime de ces découvertes. Même la monographie en cours de préparation ne permettra pas de présenter de manière exhaustive l'ensemble des données archéologiques, qu'il s'agisse des données brutes de fouilles (photos et descriptions raisonnées des unités stratigraphiques), des vestiges construits ou des objets (photos et inventaires versés dans une base de données Access). Leur volume relativement conséquent est en constante croissance en raison de la poursuite des fouilles : à l'heure actuelle ont été inventoriés, par exemple, plus 5 000 objets métalliques ; d'autres corpus sont bien plus larges (vestiges fauniques) mais certains plus restreints, tels ceux des objets en verre, pierre ou encore en matières dures animales, qui comptent seulement entre quelques dizaines et centaines de références. Face aux défis de conservation, archivage et diffusion que soulève une telle richesse qualitative et quantitative, la mise en œuvre des principes FAIR est apparue comme l'une des voies privilégiées à développer.

Le squelette d'un écosystème numérique d'exposition des données a été mis en place à partir de 2022 sur différents supports : gestionnaire de vocabulaire et entrepôts de données (photos et modèles 3D). Un thésaurus a été créé sur la plateforme ouverte Opentheso (<http://thesaurus.mom.fr/opentheso/?idt=th76>) qui comporte actuellement 444 concepts trilingues français-espagnol-anglais, regroupés en 4 ensembles, chacun doté d'une adresse pérenne (format ARK). En parallèle, un jeu d'images-test tiré du corpus des objets en matières dures animales (bois de cervidé, os) a été mis en ligne sur l'entrepôt Nakala d'Humanum (<https://nakala.fr/search/?q=ALBALAT-INDUSTRIE+MATI%C3%88RE+DURE+ANIMALE>). Les images ont été décrites et documentées par des métadonnées également trilingues, selon les standards Dublin Core. Des liens pérennes devraient permettre de passer d'un entrepôt à l'autre. En complément, des modélisations 3D d'un échantillon d'objets (céramiques, os, métaux) ont été hébergées sur la plateforme Sketchfab avec l'aval du musée de Cáceres où les pièces sont déposées après étude (<https://sketchfab.com/albalat>). Il s'agit dans ce cas d'un choix par défaut, car si un compte a bien été ouvert au Conservatoire National des Données 3D afin d'y déposer les modèles 3D d'Albalat, le processus de dépôt (incluant la conversion des fichiers et l'élaboration des métadonnées) s'est révélé trop long et complexe à mettre en œuvre. Une difficulté similaire s'est présentée pour les métadonnées des dépôts effectués sur l'entrepôt Nakala.

Les objectifs de METAlbalat sont franchir certaines étapes-clés pour ce travail dont les bases ont été posées, mais qui ne peut être développé dans l'immédiat faute de moyens humains et financiers, ce qui porte préjudice à la conservation à long terme des données, mais également à leur partage, analyse et valorisation. Au sein du vaste programme annuel qui comprend des campagnes de fouilles et de conservation des vestiges *in situ* ainsi que des missions consacrées aux inventaires et aux études des données récoltées, mobilisant différents spécialistes au fil de l'année, l'opération METAlbalat proposée à la FR Agorantic occupe donc une place à part, dont le périmètre épistémologique est bien cerné.

Cette démarche est inédite, voire avant-gardiste, à plusieurs égards : elle comble un vide dans le domaine de l'archéologie médiévale de l'islam occidental (péninsule Ibérique-Maghreb, VIII<sup>e</sup>-XV<sup>e</sup> s.) où, jusqu'à présent, l'absence d'initiatives comparables freine, par exemple, la simple recherche de parallèles morphologiques, rendant très difficile des études aréales sur la culture matérielle. Cette initiative s'inscrit également dans un contexte où Domus, la seule base de données ouverte espagnole contenant des artefacts archéologiques, toutes époques confondues (<https://ceres.mcu.es/pages/Main>), est spécifiquement conçue pour cataloguer et gérer les collections muséales espagnoles. Or, malgré sa richesse, ce catalogue numérique présente des limitations notables : tous les musées n'y sont pas représentés (dont celui de Cáceres où sont déposées nos pièces) ; il ne fournit aucune métadonnée conformes aux principes FAIR, ce qui restreint l'accès et l'exploitation des données.

METALBALAT ne prétend pas s'y substituer mais vise à créer une collection raisonnée du site d'Albalat, enrichie de métadonnées détaillées et s'appuyant sur des infrastructures publiques ; pour ce faire, placer les jeux de données thématiques d'Albalat (objets métalliques, verres, terres cuites, etc.) au centre d'un graphe sémantique potentiel : Opentheso-Nakala-Conservatoire des données 3D, toutes les dimensions d'un même vestige étant dotés d'identifiants pérennes et tous les vestiges étant eux-mêmes typés par des métadonnées normalisées (DC Terms) ou sémantiquement structurées (concepts dans Opentheso).

## 2. Questionnement scientifique en rapport avec l'intitulé de l'appel

Le récent article d'Anichini, Gattiglia et Gualandi publié dans *Archeologica Data* (2021) rappelle que les données issues de l'archéologie sont particulièrement sensibles en raison de leur caractère unique et des particularités d'obtention intrinsèques à la discipline, qui « détruit » une partie de son registre. Ce constat, bien assumé par les archéologues, a impulsé le développement de méthodes d'enregistrement rigoureuses qui s'adaptent continuellement aux avancées techniques et technologiques de leurs temps. Le but reste fondamentalement identique : préserver le maximum de données, tant quantitatives que qualitatives. Or, une partie des données produites sont désormais nativement numériques. Les clichés numériques (et aussi les vidéos, les ortho-photographies, les planimétries, etc.) remplacent les photos argentiques, les relevés topographiques et les cahiers de fouilles cèdent la place à des fichiers numériques, soulevant certes des défis propres à leur archivage mais ouvrant également la voie à des facilités de diffusion et de partage sans précédent. Dans le contexte de la recherche archéologique, le développement de nouveaux outils d'initiatives publiques conjugué à la double injonction institutionnelle du numérique et de la science ouverte soulèvent d'autres questions clés : la construction de la connaissance repose en partie sur un discours descriptif du terrain, donc sur une sémantique, qui ne fait l'objet d'aucune normalisation quant à ses termes ; comment concilier la réutilisabilité des données partagées avec cette hétérogénéité ? La masse des données concernées se compte en dizaines, voire centaines, de milliers d'enregistrements ; comment automatiser, en tout ou partie, leur documentarisation ? La nécessaire critique des données et des analyses publiées suppose un aller-retour permanent entre toutes les dimensions (terminologie, représentations graphiques, statistiques comparatives, etc.) des corpus ; comment à la fois fluidifier et garantir la rigueur d'un tel dispositif ? Enfin, quel impact concret peut-on attendre de la mise en ligne ouverte de tels jeux de données, tant pour le public amateur que pour les scientifiques eux-mêmes ?

## 3. Méthodologie

- *Nettoyage et enrichissement des enregistrements* : Outre le nettoyage préalable des données incluses dans la BD Access de gestion des données de fouille, on procédera à l'enrichissement et à la reprise des enregistrements pour y intégrer ou en extraire des métadonnées réutilisées, via leur structuration dans Opentheso, pour la documentarisation des dépôts dans Nakala et dans le Conservatoire des données 3D.
- *Transfert automatisé* : Les données primaires, actuellement stockées localement dans une base de données développée sous Microsoft Access, seront exportées vers la plateforme Nakala. Ce travail, effectué dans le cadre d'un stage de L3 en informatique, sera l'occasion de tester l'outil Mynkl, développé par le consortium Cahier et actuellement en cours de finalisation par le consortium Ariane, pour aider aux dépôts de données avec leurs métadonnées dans Nakala.
- *Enrichissement des métadonnées* : Un stagiaire en archéologie de niveau Licence (L3) ou Master (encadré par Jérôme Darmont et/ou Sabine Loudcher) aura pour mission d'enrichir, sur les plans linguistique et sémantique, les métadonnées associées aux champs archéologiques. L'une des tâches sera de développer le thésaurus trilingue (ES-FR-GB, sur la plateforme Opentheso) des termes/concepts utilisés pour décrire les vestiges d'Albalat et leurs contextes opérationnels et scientifiques : label, définition, relations croisées et positionnement au sein d'une poly-hiérarchie sémantique, notes éditoriales et scientifiques. Cette mission implique une bonne compréhension des termes techniques archéologiques.
- *Collections d'images en Open Access* : Création de collections thématiques d'images des vestiges documentées dans Nakala, dont les métadonnées renvoient à la fois aux catégories Dublin Core et aux concepts définis dans Opentheso.
- *Acquisition de données en 3D* : Le projet vise non pas à acquérir de nouvelles données primaires, mais à diversifier les formats et supports de celles déjà obtenues pour en garantir la conservation tout en proposant un accès plus interactif. De nouveaux modèles 3D de divers artefacts remarquables d'Albalat, conservés dans les réserves du projet de fouilles et du musée de Cáceres en Espagne, seront générés par photogrammétrie. Les modèles seront eux-mêmes documentés par des métadonnées extraites d'Opentheso ou du référentiel Dublin Core.

#### 4. Résultats attendus et caractère innovant de la recherche

Compte tenu du temps imparti et du volume conséquent des données à traiter, l'objectif principal de ce projet est de rendre interrogeables et réutilisables plusieurs jeux de données extraits du corpus des artefacts d'Albalat, incluant les matières dures animales, le verre, la terre cuite, et une sélection de métaux. METALbalat vise à améliorer, tant quantitativement que qualitativement, les métadonnées associées, afin d'augmenter l'accessibilité et la richesse des rares informations actuellement disponibles en Open Access. La collaboration avec l'équipe de recherche en informatique ERIC permettra d'automatiser une partie du traitement des données, intégrant ainsi une composante innovante au projet. Cette automatisation permettra d'optimiser le travail de dépôt et de tester, par la même occasion, l'efficacité de l'outil Mynkl. Le versant informatique devrait également rendre possible la génération automatique de fiches-types pour préparer le versement de données dans les BD Domus (à proposer au musée de Cáceres) et Artefacts.mom. L'exposition des collections en Open Data sera rendue possible par la création d'une page Nakala-Press à interface trilingue. Des liens assureront, entre autres, la connexion vers le CND3D où seront déposés les modèles 3D, ainsi que vers le thesaurus d'Opentheso pour l'enrichissement sémantique des métadonnées. METALbalat représentera un premier pas vers l'ouverture de collections d'artefacts archéologiques d'al-Andalus ; il espère ainsi enrichir les recherches actuelles en facilitant, entre autres, les comparaisons et réutilisations et, pourquoi pas, stimuler l'émergence de projets futurs dans le domaine de la gestion et de la diffusion des données de l'archéologie de l'islam occidental.

#### 5. Dimension interdisciplinaire

*Indiquez les champs disciplinaires associés, la cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques », les axe(s) de la FR concerné(s).*

Le projet METALABALAT se situe à l'intersection de l'archéologie, de la technologie de l'information et de la conservation/exposition des données, avec un fort accent sur l'optimisation des processus. La dimension informatique, assurée par l'UR ERIC, vise à optimiser la gestion de données archéologiques qui sont la base du projet, en automatisant le passage d'une base de données locale à l'entrepôt Nakala. L'exposition des données (représentations numériques, métadonnées, thesaurus) permet également de renouveler les méthodes traditionnelles de l'archéologie.

En ce sens, METALABALAT s'inscrit dans l'axe de recherche 2 *Culture and Digital* de la FR Agorantic par sa capacité à intégrer les technologies numériques pour la gestion et la valorisation d'un patrimoine historique et culturel spécifique (archéologie islamique médiévale et plus concrètement al-Andalus), quasiment absent des espaces numériques publics et ce, malgré « une offre éditoriale de data journals » en archéologie (V. Rebouillat, Lyon 1, 2021) et, plus largement, des répertoires ouverts de données archéologiques (ADS au Royaume-Uni, tDAR aux États-Unis et, à vocation plus internationale, Open Context).

#### 6. Partenariats extérieurs envisagés

- Musée provincial de Cáceres
- Sabine Sorin-Mazouni, consortium 3D & Miguel Angel García Pérez, projet Albalat : modélisations 3D et soutien aux dépôts des modèles 3D.
- UMR 5138 ArAr, porteuse de la base d'objets mobiliers *Artefacts*.

#### 7. Valorisation (si prévue)

*Déclaration d'invention permettant de valoriser un savoir-faire, une base de données ou un logiciel*

Malgré l'absence d'innovation dans le domaine du savoir-faire, le projet METALbalat permettra de tester l'efficacité de l'outil Mynkl. Il est donc prévu, dans une seconde étape (2025-2026), de travailler à la publication d'un *data paper* collectif pour présenter la publication argumentée du jeu de données ainsi que le programme informatique développé.

#### 8. Expression des besoins en assistance informatique

*Indiquez ici les tâches que vous voudriez faire réaliser par les ingénieur-es de la FR. Leur temps étant limité, il est possible que cela ne soit pas possible : pour cette raison, il est également nécessaire d'identifier une solution de repli pour mener ces tâches à bien si elles sont critiques pour le projet.*

#### 9. Évaluation du projet

Indiquez ici les noms et les coordonnées mail de 4 experts externes que l'on pourrait solliciter en vue de l'évaluation de votre projet.

- 1- DUROST Sébastien (Bibracte EPCC) : systèmes numériques d'information archéologique et publications, [s.durost@bibracte.fr](mailto:s.durost@bibracte.fr)
- 2- ROUSSET Marie-Odile (DR, Archéorient, MOM) : données ouvertes de la recherche archéologique, [marie-odile.rousset@mom.fr](mailto:marie-odile.rousset@mom.fr)
- 3- GUTIÉRREZ LLORET Sonia (PR. Univ. Alicante) : archéologie islamique (al-Andalus), [sonia.gutierrez@gcloud.ua.es](mailto:sonia.gutierrez@gcloud.ua.es)
- 4- Marianne REBOUL (MCF ENS Lyon, UMR Ihrim) : Humanités numériques, [marianne.reboul@ens-lyon.fr](mailto:marianne.reboul@ens-lyon.fr)

10. Budget (€) prévisionnel *		
	Brève description	Montant
Missions	Missions en Espagne pour l'acquisition des captures photogrammétriques d'artefacts, déposés en Espagne (dépôt temporaire de fouille, réserves du musée de Cáceres) : logement et manutention	1400 €
Consommables, petits matériels**	Appareil photo (appui photogrammétrie), 2 Disques durs 4-5 TB (stockage temporaire des données primaires)	600 €
Organisation de réunions	Réunions de travail avec les stagiaires et les collaborateurs (Lyon)	120 €
Stages***, vacations	2 x stage 4 mois (2 étudiants Licence-Master)	4800 €
Prestations de service	Génération de modèles 3d d'artefacts archéologiques	1043 €
	Achat url (OVHcloud)	37 €
Budget total		8000 €
Co financements le cas échéant		0 €
Budget demandé à Agorantic		8000 €
Recettes extérieures		0 €

Mon directeur d'unité est informé du dépôt de ce projet **X** (Marilyn Nicoud, [marilyn.nicoud@univ-avignon.fr](mailto:marilyn.nicoud@univ-avignon.fr))