

---

# HisArc-RDF : une chaîne opératoire pour la production et la réutilisation de données archéologiques et/ou historiques appuyées sur un référentiel terminologique propre et une ontologie inscrits dans le web sémantique

Francesco Beretta<sup>\*1</sup>, Marie-Odile Rousset<sup>\*2</sup>, François Mistral<sup>†3</sup>, Miled Rousset<sup>‡4</sup>, and Vincent Alamertery<sup>§5</sup>

<sup>1</sup>Laboratoire de Recherche Historique Rhône-Alpes - UMR5190 (LARHRA) – Université Pierre Mendès France - Grenoble 2, École Normale Supérieure - Lyon, Université Lumière - Lyon 2, Université Jean Moulin - Lyon III, Centre National de la Recherche Scientifique : UMR5190 – Institut des Sciences de l’Homme - 14, avenue Berthelot - 69363 Lyon cedex 07, France

<sup>2</sup>ARCHEORIENT - Environnements et sociétés de l’Orient ancien – Centre National de la Recherche Scientifique : UMR5133 – France

<sup>3</sup>Agence Bibliographique de l’Enseignement Supérieur – Ministère de l’Enseignement Supérieur et de la Recherche Scientifique – France

<sup>4</sup>Maison de l’Orient et de la Méditerranée - Jean Pouilloux – Centre National de la Recherche Scientifique : FR3747 – France

<sup>5</sup>Laboratoire de Recherche Historique Rhône-Alpes - UMR5190 – École Normale Supérieure - Lyon – France

## Résumé

Construit autour des laboratoires Archéorient et LARHRA, HisArc-RDF réunit depuis 2019 un consortium pluridisciplinaire (archéologie, histoire, terminologie, bibliographie et informatique) qui s’est attaché à développer la preuve de concept d’une chaîne opératoire de partage de données archéologiques et historiques, rendues ” FAIR ” par l’articulation de méthodes développées initialement dans chaque discipline.

Nous présenterons les fondements pratiques de notre démarche, puis les caractéristiques concrètes de trois jeux-tests de données en archéologie préventive (fouilles), archéologie du Proche-Orient (prospection à l’échelle d’une région) et histoire de la littérature (langue régionale), avant d’aborder la description du workflow, sans omettre de souligner au passage les difficultés opérationnelles qu’il a fallu résoudre pour mettre en œuvre un tel modus operandi dans le contexte des recherches en Humanités. Si le ” web des données ” est couramment appelé ” sémantique ”, le rôle joué par le vocabulaire (ce sous-ensemble de la catégorie ” langage ”) y est facilement réduit, de facto, au choix de métadonnées de signalement. Or,

---

\*Intervenant

†Auteur correspondant: [mistral@abes.fr](mailto:mistral@abes.fr)

‡Auteur correspondant: [miled.rousset@mom.fr](mailto:miled.rousset@mom.fr)

§Auteur correspondant: [vincent.alamertery@ens-lyon.fr](mailto:vincent.alamertery@ens-lyon.fr)

en particulier en archéologie, la trace scientifique passe en partie par la description langagière d'une expérience-terrain non reproductible ; exprimée le plus souvent en langage " naturel " ... c'est-à-dire ambigu. A contrario, la spécificité d'HisArc-RDF est de placer la sémantique au cœur du workflow, tant sa composante terminologique que sa composante ontologique. L'interopérabilité de bases hétérogènes repose habituellement sur des concepts de haut niveau qui la réduisent à un plus grand dénominateur sémantique et ainsi excluent de la FAIRisation les spécificités d'un domaine et d'une recherche singulière (le concept " objet physique " englobe, par exemple, aussi bien des pièces de monnaie que des arquebuses). Dans HisArc-RDF, nous minimisons cet effet de dilution sémantique ; l'interopérabilité repose sur le modèle conceptuel (ou ontologie) mais aussi sur deux autres piliers : les référentiels d'autorité et les thésaurus terminologiques, ces derniers à un niveau de granularité plus proche du " terrain " que celui des référentiels bibliographiques.

Articuler l'application OntoME, développée par le Pôle histoire numérique du LARHRA, et l'application de gestion de thésaurus Opentheso, développée à la MOM, crée le lien entre le modèle conceptuel et la terminologie. OntoME permet la création et l'utilisation d'ontologies adaptées au domaine étudié ; Opentheso permet de structurer formellement des terminologies de recherche inscrites dans le web sémantique ; enfin, Opentheso a été interconnecté à OntoME et à la plateforme IdRef qui gère les autorités de l'Agence bibliographique de l'enseignement supérieur. En fin de chaîne, les données hétérogènes ainsi traitées sont devenues interopérables et sont déposées dans un entrepôt permettant leur publication, leur interrogation et leur réutilisation pour de nouvelles recherches grâce à un point d'accès SPARQL.