HisArc-RDF: prototyping an operating chain, related to the Linked Open Data, on structurally and semantically heterogeneous archaeological data sets

M.-O. Rousset, Francesco Beretta, Emmanuelle Perrin, Vincent Alamercery, Sébastien Durost, Jean-Pierre Girard, François Mistral, Miled Rousset

Linked Pasts 5 Bordeaux, 11-13 December 2019

HisArc-RDF :
prototyping an operating chain, related to the Linked Open Data, on structurally and semantically heterogeneous archaeological data sets

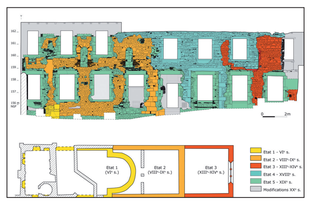
HISARC-RDF project aims to share and reuse archaeological data, by articulating their description on semantic web repositories and standards.

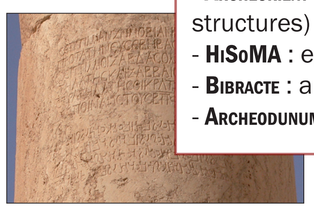
The project will test four data sets, samples of which will be determined to cover different archaeological themes, both in terms of vocabulary and data modelling: hydraulic structure, epigraphic data, ceramic and building structures. To achieve that, we want to reduce veracity concerns on descriptive vocabularies (from artefacts description to period names and limits) so that diferent datasets can be addressed and compared through concepts (instead of words) at a data level. For that prupose, we will create micro-thesauri (at sub-disciplinary level), then vocabulary and data model will be aligned in parallel with the semantic web repositories and CIDOC CRM to build a structured description of database fields as well as database variables. To address this, the project plans to interconnect OpenTheso (thesaurus management tool), OntoME (ontology management environment) and IdRef (database for authority records).

Finally, the project plans to bring together, through workshops, a large network of actors in order to disseminate good practices experienced by LOD implementation.

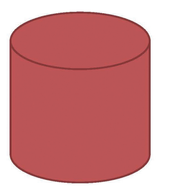
The consortium gets together Archéorient (UMR 5133), Larhra (UMR 5190), Maison de l'Orient et de la Méditerranée (FR 3747), HiSoMA, (UMR 5189), Bibracte EPCC, archaeology laboratories in Strasbourg (Archimède, UMR 7044), Besançon (Chrono-Environnement UMR 6249) and Paris-Sorbonne-ENS (AOrOc UMR 8546), the plateform Spatio (national network Maisons des sciences de l'Homme), ERIC lab (data intelligence), Bibliographic Agency for Higher Education (Abes) and Archéodunum (private company of preventive archaeology).

Volume and variety of archaeological data, variety and veracity concerns about archaeological vocabularies actually blocks interoperability on data sets.



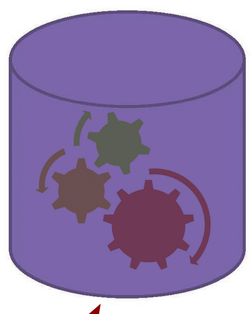


FOUR HETEROGENEOUS DATA SETS
- **ARCHÉORIENT** : data from the survey of the Arid Margins of Northern Syria (hydraulic structures)
- **HISOMA** : epigraphic data from the Greek and Latin Inscriptions of Syria
- **BIBRACTE** : archaeological operations in a programmed context (ceramic)
- **ARCHEODUNUM** : reports of preventive archaeological operations (building structures)

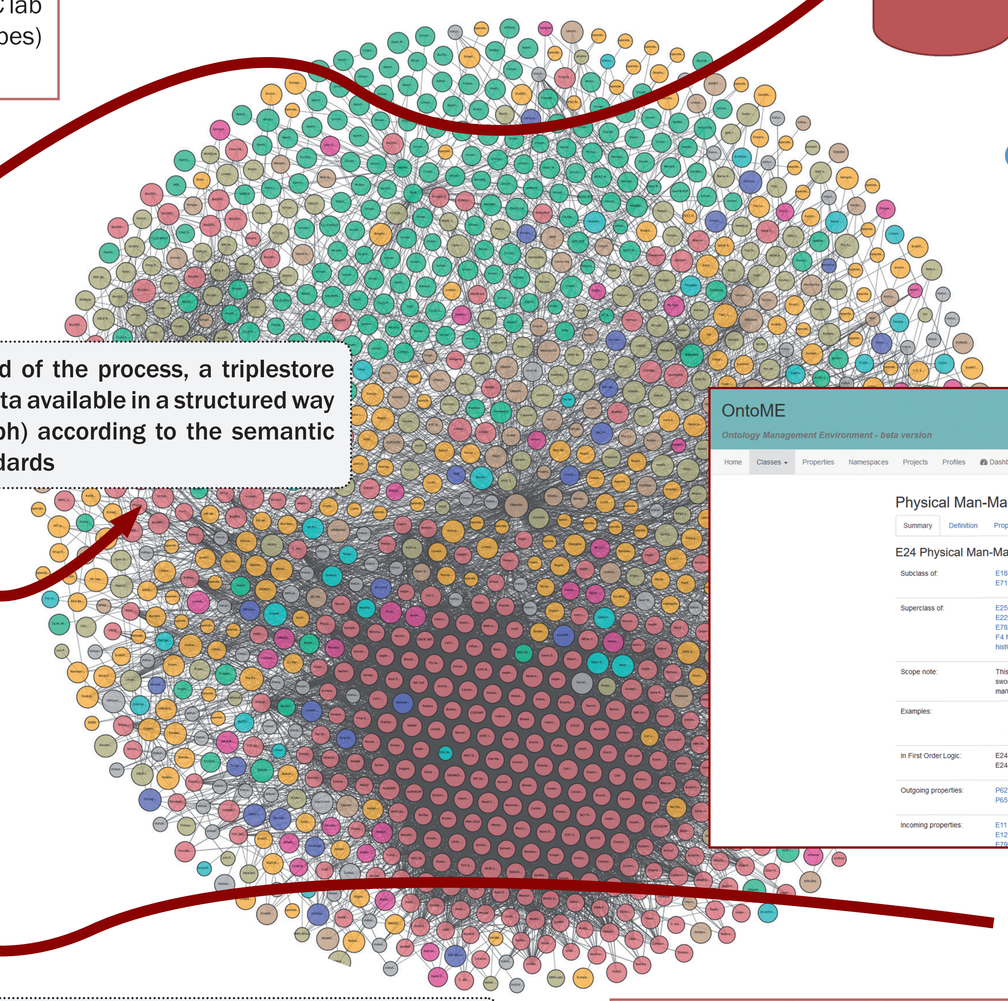


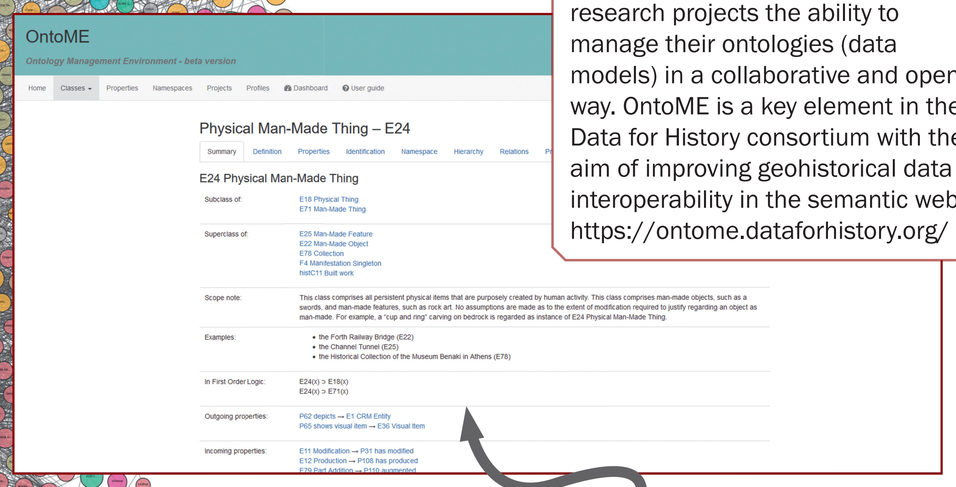
Creating of a data model

Structuring description of database fields as well as database variables, with both data model and controlled vocabularies



At the end of the process, a triplestore makes data available in a structured way (RDF graph) according to the semantic web standards



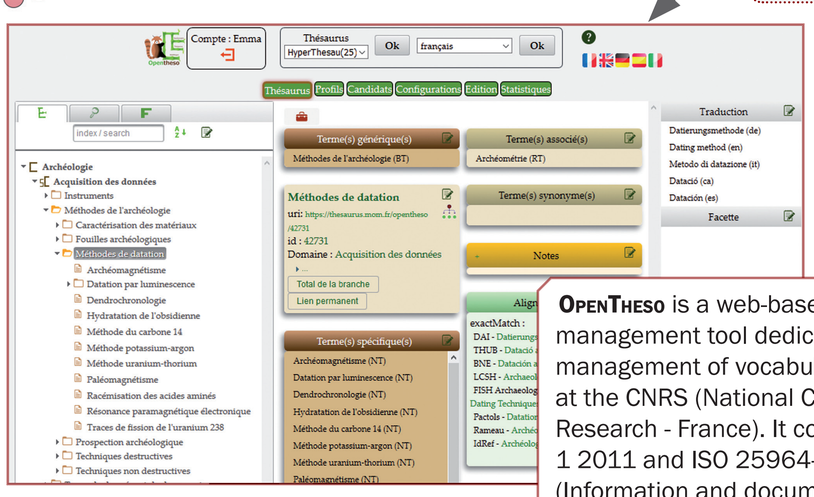


ONTOME, Ontology management environment, is an online application developed by the LARHRA Digital history research team that offers research projects the ability to manage their ontologies (data models) in a collaborative and open way. OntoME is a key element in the Data for History consortium with the aim of improving geohistorical data interoperability in the semantic web. https://ontome.dataforhistory.org/

Matching links with international information systems for people and place authority records

Importing RDF data model and thesaurus

Data model and thesaurus interlinking



OPENTHESO is a web-based thesaurus management tool dedicated to the management of vocabularies. It is developed at the CNRS (National Center for Scientific Research - France). It conforms to ISO 25964-1 2011 and ISO 25964-2:2012 standards (Information and documentation. Thesauri and interoperability with other vocabularies). <https://github.com/miledrousset/opentheso>

IDREF (database for authority records) is a web application developed and maintained by French Bibliographic Agency for Higher Education. IdRef allows users and applications to query, consult, create and enrich authority records. https://www.idref.fr/

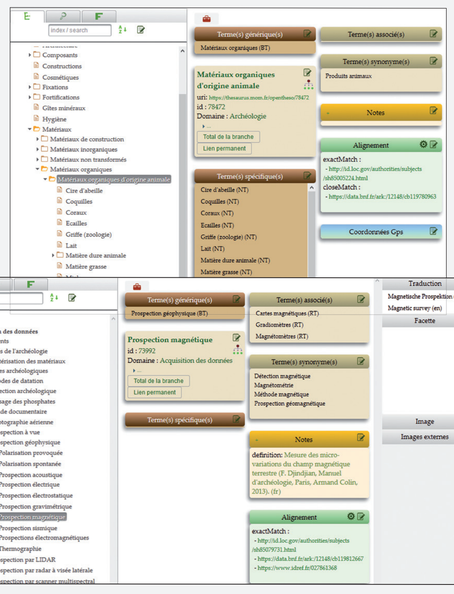
Software module to link to library community's data

ONTOME AND OPENTHESO INTERLINKING

E55 type comprises concepts denoted by terms from thesauri and controlled vocabularies used to characterize and classify instances of CRM classes. Specific subclasses of E55 will be created in OntoME and linked to a top term in OpenTheso.

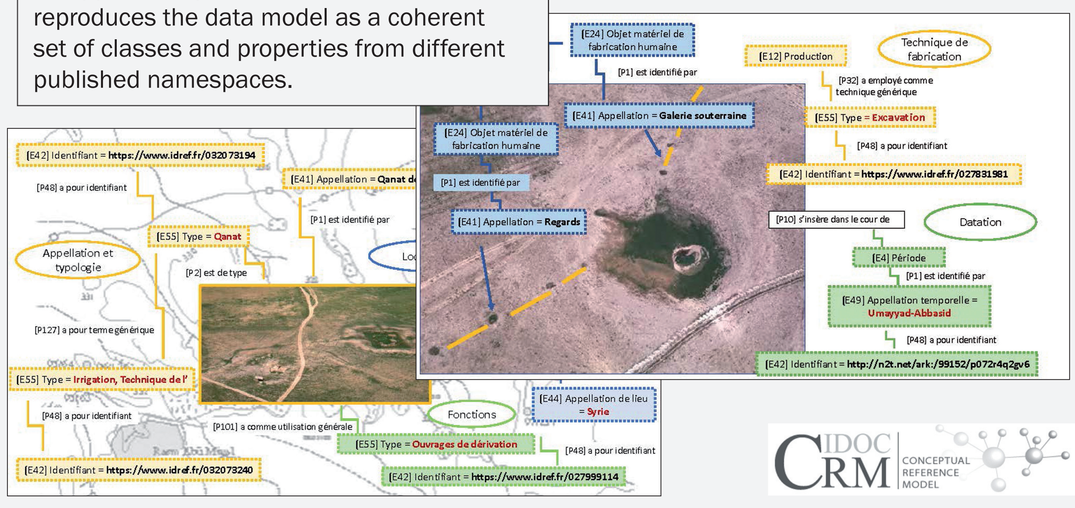
For instance, a "C22 Amphora type" class, subclass of E55 Type, is linked to an "Amphora type" top term in OpenTheso. Then all the different types of amphorae created in the thesaurus as narrower terms of "Amphora type" are automatically recognized as instances of "C22 Amphora type".

Whether on the OntoME or OpenTheso side, the whole process can be community driven.



DATA MODEL

An application profile is created that reproduces the data model as a coherent set of classes and properties from different published namespaces.



Marie-Odile Rousset (CNRS - Université Lumière Lyon 2 – Archéorient)
Francesco Beretta (CNRS – Université de Lyon – LARHRA)
Emmanuelle Perrin (Université de Lyon – Archéorient)
Vincent Alamercery (ENS de Lyon – Université de Lyon – LARHRA)
Sébastien Durost (Bibracte EPCC)
Jean-Pierre Girard (Université de Lyon – Archéorient)
François Mistral (ABES)
Miled Rousset (CNRS - Maison de l'Orient et de la Méditerranée)

PHOTO CREDITS
Marges arides de la Syrie du Nord (Archéorient) : qanat (aqueducs souterrains repérés par leurs regards réguliers). @ Bernard Geyer.
Bibracte : vases à boisson. @ Antoine Maillier. http://www.bibracte.fr/fr/le-musee-de-bibracte
HiSoMA : Palmyre (Syrie), grande colonnade. Dédicace bilingue, grecque et araméenne, à la reine Zénobie (271 apr. J.-C.). @ Julien Aliquot. https://www.hisoma.mom.fr/recherche-et-activites/inscriptions-grecques-et-latines-de-la-syrie
Archéodunum : Relevé phasé du mur gouttereau Sud de l'ancienne église Saint-George. @ Quentin Rochet & Camille Collomb. http://www.archeodunum.com/wp-content/uploads/PDF/ABAD_poster_1.pdf
Linked open data : lod-cloud.net.

This project was designed with the financial support of the LABEX IMU (ANR-10-LABX-00) of the University of Lyon, as part of the "Investissements d'Avenir" program (ANR-11-IDEX-0007) managed by the French National Research Agency (ANR).