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

To cite this version:
M.-O. Rousset, Francesco Beretta, Emmanuelle Perrin, Vincent Alamercery, Sébastien Durost, et al.. HisArc-RDF: prototyping an operating chain, related to the Linked Open Data, on structurally and semantically heterogeneous archaeological data sets. Linked Pasts 5: Back to the (re)sources, Dec 2019, Bordeaux, France. halshs-02413859

HAL Id: halshs-02413859
https://halshs.archives-ouvertes.fr/halshs-02413859
Submitted on 16 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives| 4.0 International License
**HisArc-RDF** : prototyping an operating chain, related to the Linked Open Data, on structurally and semantically heterogeneous archaeological data sets

**Description**

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-007) managed by the French National Research Agency (ANR).

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 different databases can be addressed and compared through concepts (instead of words) at a data level. For that purpose, 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), Archimède, UMR 7044), Besançon (Chrono-Environnement UMR 5190), Maison de l'Orient et de la Méditerranée (FR 3747), HiSoMA, 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 different datasets can be addressed and compared through concepts (instead of words) at a data level. For that purpose, 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).

**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/](https://ontome.dataforhistory.org/)

**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/](https://www.idref.fr/)

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

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

**OntoME and OpenTheso relationship**

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 "Amphoratype" are automatically recognized as instances of "C22 Amphora type".

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

---

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