Prehistoric Mining Data for ARIADNE

How to create Open Data from archaeological research for the ARIADNE community and beyond

Gerald HIEBEL, University of Innsbruck – Institute of Archaeologies, Austria
Brigit DANTHINE, Austrian Academy of Science - Austrian Archaeological Institute , Austria
Manuel SCHERER-WINDISCH, , University of Innsbruck – Institute of Archaeologies, Austria

Keywords: *FAIR data — ARIADNE — Archaeology — Semantic Web — Prehistoric Mining*

CHNT Reference: add full reference here: Hiebel, G., Danthine, B ., and Scherer-Windisch, M . (2022). 'Prehistoric Mining Data for ARIADNE', in CHNT Editorial board. *Proceedings of the 27th International Conference on Cultural Heritage and New Technologies, November 2022.* Heidelberg: Propylaeum.
DOI:

Introduction

In accordance with the goals of this round table we want to present an approach that provides in a first step open excavation data in RDF for prehistoric mining archaeology applying the FAIR principles and in a second step creates metadata for the ARIADNE catalogue based on the data created in the first step with little additional input.

Step 1: Creating open excavation data for prehistoric mining archaeology in RDF

We want to present an approach to create RDF data that uses Excel spreadsheets for data entry, Postgres database for data transformations and sematic tools like Karma or OntoRefine for RDF creation. For Thesaurus creation mind mapping software is used in addition to ease the creation of a hierarchical structure of terms. The pipeline for RDF data creation is illustrated in Figure 1.

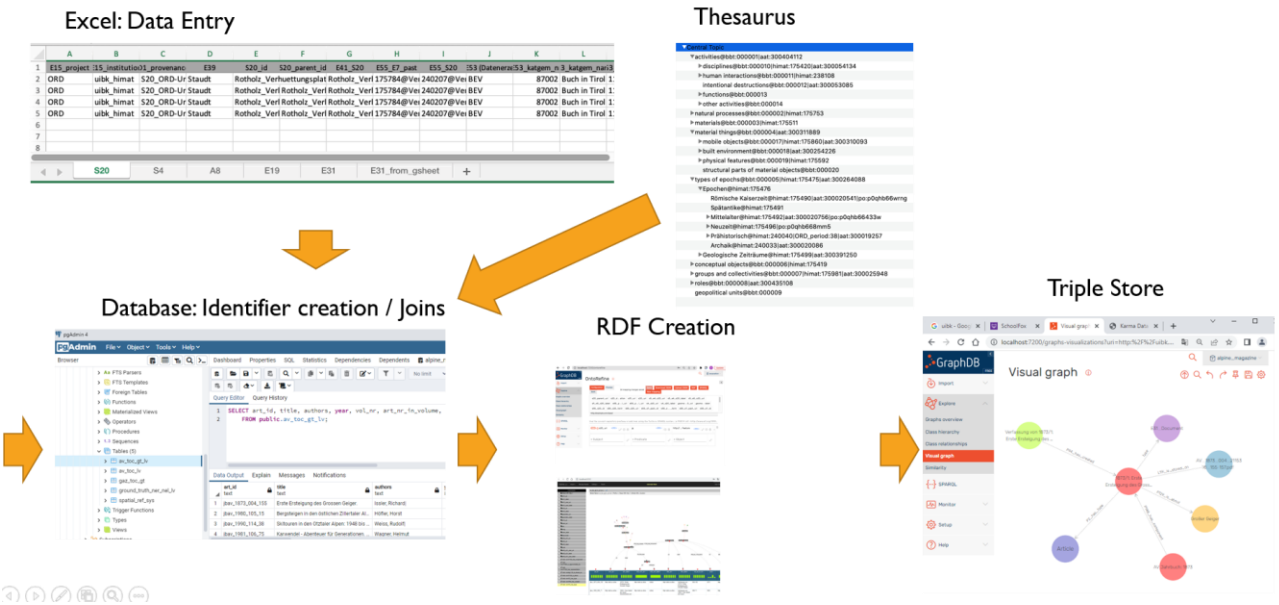

Fig. 1 Pipeline for RDF data creation

The approach was executed with data from a project related to prehistoric mining activities in the eastern Alps of Austria. The data were collected during several scientific research campaigns and as the data are of archaeological nature, the methodologies and guidelines of ARIADNE (Advanced Research Infrastructure for Archaeological Data Networking in Europe) were already used to process the data making them FAIR (FORCE11 2019).

Methods and Materials

Research data were deposited at the Zenodo repository located at the CERN Data Center which has experience in long-term archiving and assigns DOIs (Digital Object Identifier).

The modelling of the excavation data was based on the three different main components of (1) the material research objects like sites, trenches, stratigraphical units and finds, (2) the activities like excavating or analyzing a find and (3) the actors like archaeologist or a labor. These were each entered in different excel sheets and related through identifiers as it would have been done in a relational database. Because of such detailed modelled provenances, it is possible to explicitly specify the different research activities with their specific methodologies and lining them to the investigating person or institution.

The excavation data was derived from projects at the research center HIMAT (History of Mining Activities in Tyrol and adjacent areas) conducted by the Archaeological Department of the University of Innsbruck. The documentations were done according to the guidelines of the Federal Monuments Office (BDA – Bundesdenkmalamt).

To ensure long-term preservation all documentation files were converted to file formats specified by the guidelines of the ADS (Archaeology Data Service, York) before uploading them to Zenodo and receiving a citable DOI. For the conversion a python-program was written within the research-project “FAIR Data for Prehistoric Mining Archaeology” (Hiebel et al. 2019). Because Zenodo gives one single DOI to a whole dataset the python-program also writes a metadata-file where each folder and file receives a unique identifier consisting of the Zenodo-DOI and an ongoing number. While this

"step", although important, is pretty simple and not really innovative, it is the combination of the different data and documents from the individual projects which make the data truly Interoperable and Reusable. It needs to fulfil basically two requirements. First the model behind all resources needs to be general enough that all projects can be included, but compact enough to allow the research on specific questions. And secondly a shared vocabulary/thesaurus and ontology which is usable but extended enough to fulfil all needs.

For the creation of the metadata for all generated and deposited files and research documents the CIDOC CRM ontology with its extension was used. CIDOC CRM is an ISO standard for Cultural Heritage Information, which was adopted as the conceptual background by ARIADNE, the European Union Research Infrastructure for archaeological resources. CIDOC CRM was extended in the course of ARIADNE with CRMarchaeo to model archaeological excavations and which was build based on the official documentation requirements of different countries including the Austrian Federal Monuments Office (BDA). CRMsci was used to model scientific observations. Concepts specific to mining archaeology research are organized with the DARIAH Back Bone Thesaurus, a model for sustainable interoperable thesauri maintenance, developed in the European Union Digital Research Infrastructure for the Arts and Humanities (DARIAH). SKOS (Simple Knowledge Organization System) was used to organize our vocabularies, another semantic web standard for sharing and linking knowledge organization systems such as thesauri, taxonomies, classification schemes and subject heading systems. One essential step to use the created data for aggregation in ARIADNE as described in step 2 was to map the concepts used in the research data to the concepts of the Getty AAT and PeriodO.

Step 2: Creating ARIADNE Aggregation Data

To create ARIADNE Aggregation Data additional information specific to the ARIADNE Catalogue was entered in a "ADRIANE Metadata" Excel spreadsheet and a modelling to the ARIADNE catalogue had been performed to create the RDF necessary for the ARIADNE Portal. Through using URI identifiers from the RDF network created in the previous step one within the "ADRIANE Metadata" Excel lots of the information provided in the original network can be accessed and used for the ARIADNE catalogue. SPARQL Queries have to be written to provide these additional data for the ARIADNE METADATA catalogue and we are now in the course of doing that. Meanwhile the information included in the "ADRIANE Metadata" Excel spreadsheet will be available in the catalogue. Figure 2 illustrates the desired workflow.

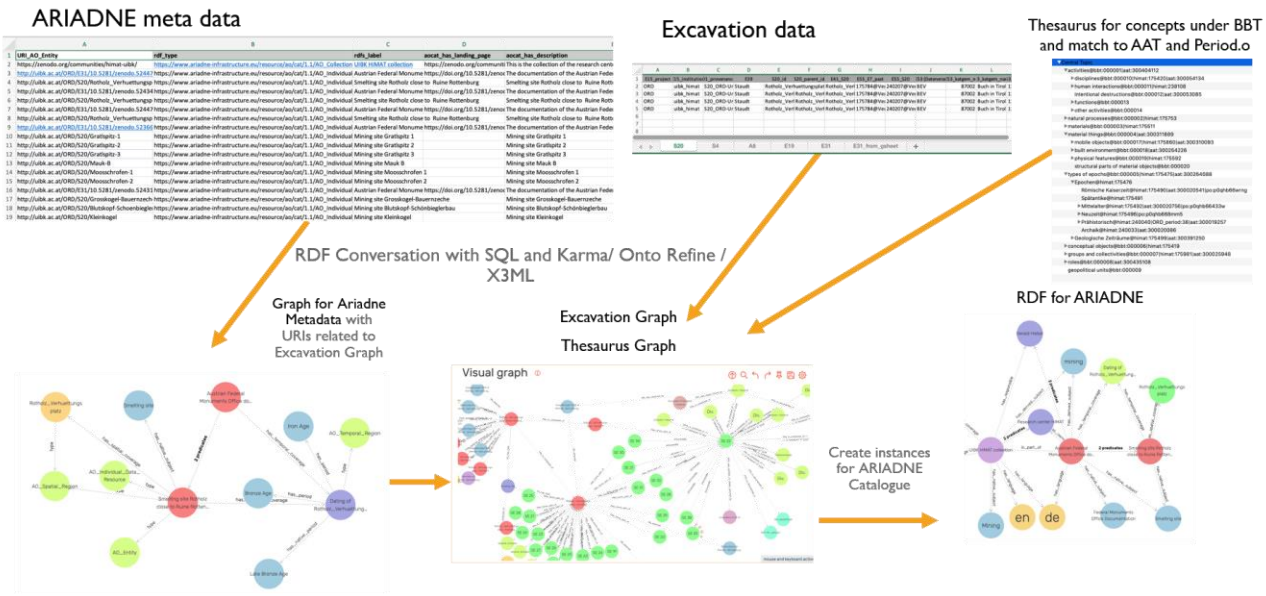

Fig. 2. Adding the ARIADNE Metadata Spreadsheet to the Pipeline

Funding

The research presented in this paper was financed by the Austrian Science Fund (FWF) in the course of the Open Research Data Pilot funding schema with the project ”Mining, Technology and Trade in Prehistory - Open Data from Prehistoric Mining Archaeology“ (FWF, ORD 74-VO) and the Project “ Information Integration for Prehistoric Mining Archaeology” (FWF, P31814-N38).

References

Chryssoula Bekiari, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, Athanasios Velios, Erin Canning, Philippe Michon. 2022. “Volume A: Definition of the CIDOCConceptual Reference Model.” Accessed June 27, 2022. https://www.cidoc-crm.org/Version/version-7.2.1

FORCE11. 2019. “Guiding Principles for Findable, Accessible, Interoperable and Re-Usable Data Publishing Version B1.0.” Accessed November 14, 2019. https://www.force11.org/fairprinciples.

Hiebel, Gerald, Gert Goldenberg, Caroline Grutsch, Klaus Hanke, and Markus Staudt. 2019. „FAIR Data for Prehistoric Mining Archaeology.“ *International Journal on Digital Libraries* **22,** 267–277 (2021). https://doi.org/10.1007/s00799-020-00282-8