



This PDF is a simplified version of the original article published in Internet Archaeology. Enlarged images which support this publication can be found in the original version online. All links also go to the online version.

Please cite this as: Hiebel, G., Danthine, B., Peralta Friedburg, M. and Scherer-Windisch, M. 2023 Prehistoric Mining Data: How to create Open Data from archaeological research for the ARIADNE community and beyond, Internet Archaeology 64. <https://doi.org/10.11141/ia.64.8>

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

Gerald Hiebel, Brigit Danthine, Milena Peralta Friedburg and Manuel Scherer-Windisch

Making archaeological data available for the scientific community received a major boost through the ARIADNE Infrastructure projects that began in 2013. The purpose of this article is to present a methodology for converting excavation and field survey documentation from its original source into a format that corresponds to the FAIR principles and commits to the ARIADNE guidelines, with the ultimate goal to make the data publicly available and incorporating them into the ARIADNE portal.

The methodology is illustrated by excavation and field survey reports that have been created according to the requirements of the Austrian Federal Monuments Office and document the archaeological investigations undertaken in the project 'Prehistoric copper production in the eastern and central Alps'. These were further processed in an Open Research Data pilot project, which initiated the development of the methodology. This consists of a pipeline to convert excavation and field survey reports to lasting file formats and extract information on sites, archaeological structures, stratigraphic units and finds from the reports to create CIDOC CRM encoded RDF data. The goal of the pipeline is to lower the entry threshold for creating such data by making data entry easy, using spreadsheets and applying easy to install and open-source (at least freely available) tools. All components of the pipeline are freely available, and detailed documentation with installation instructions and sample data can be downloaded for those who wish to test the methodology and try it out using their own data.



1. Introduction

The interdisciplinary Research Center [HiMAT](#) (History of mining activities in the Tyrol and adjacent areas, University of Innsbruck) investigates mining history from prehistoric to modern times using an interdisciplinary approach. One of the projects carried out at the research centre was the multinational [DACH](#) project 'Prehistoric copper production in the eastern and central Alps' (FWF I 1670-G19), which ran from 2015 to 2018 (Goldenberg [2021](#)). For the specific geographical region of the project that was covered by the University of Innsbruck (Figure 1), the data transformation to open and reusable data was investigated in a separate [Open Research Data](#) pilot project from 2019 to 2021 (Hiebel *et al.* [2019](#)). The methodological approach used the FAIR principles (Wilkinson *et al.* [2016](#); FORCE11 [2019](#)) to make data Findable, Accessible, Interoperable and Reusable. In the following sections we want to show how the archaeological reports for the [Austrian Federal Monuments Office](#) were transformed. The first step was to change the file formats where necessary and store the documentation on Zenodo and Google Drive. In the second step the information was extracted and conceptually represented with [CIDOC CRM](#) and its extensions (Doerr [2003](#); Bekiari *et al.* [2022](#)) and implemented in [RDF](#), a semantic web standard for data interchange on the Web. RDF is also used by [ARIADNE](#), the European Union Research Infrastructure for archaeological resources (Niccolucci *et al.* [2013](#)) to bring archaeological resources together in the [ARIADNE Portal](#). The methodology of this information extraction and RDF creation with its various steps are the focus of this article. The methodology presented is exemplified here with concrete archaeological data from the Federal Monuments Office documentation but it can be applied to a variety of domains and data sources and was also used to create the resources for the ARIADNE Portal. The single steps to follow together with sample data are elaborated in a separate 'RDF Creation Pipeline', with examples from other domains. A short overview of the pipeline is given in Section 5.

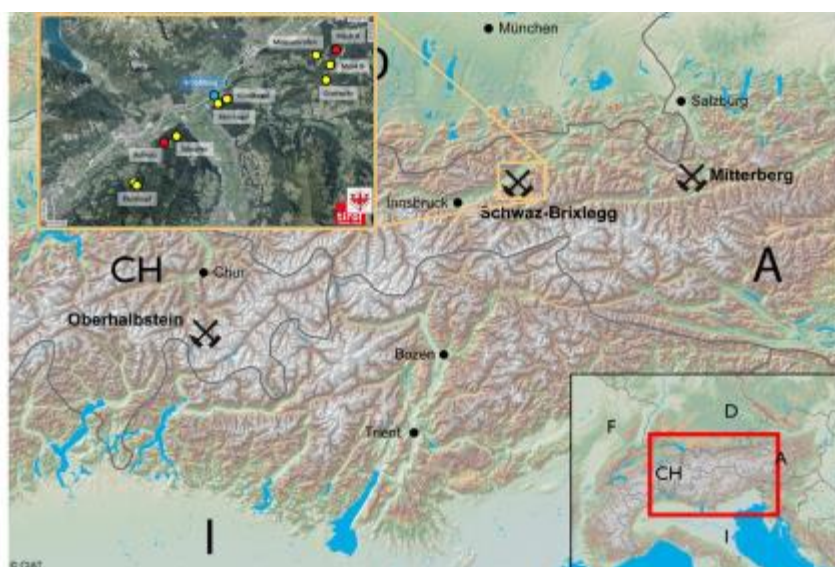


Figure 1: The Schwaz-Brixlegg area in western Austria covered by the University of Innsbruck in the multinational DACH project 'Prehistoric copper production in the eastern and central Alps'



2. Federal Monuments Office Documentation

Every archaeological investigation in Austria has to be documented in accordance with the [requirements of the Austrian Federal Monuments Office](#). In the case of the DACH project 'Prehistoric copper production in the eastern and central Alps', six reports on field surveys and excavations were documented according to these guidelines. They define in detail which reports, lists, photos and plans have to be created for prospections, excavations, stratigraphic units, finds, archaeological objects and groups. Figure 2 shows the categories and directory structure of the documentation as defined by the Federal Monuments Office. The left side illustrates an example for the documentation of an excavation activity at the smelting site '87002.15.01 Verhüttungsplatz südlich der Ruine Rottenburg'. The site was documented in 336 files with a total size of 5.34 Gb. For example, the documentation of a survey on mining sites (underground mines and pits) in the fahlore district of Schwaz-Brixlegg comprised 244 files with a total size of 1.33 Gb, documented under the folder '87007.15.01 Bergbaurevier Schwaz-Brixlegg'. For the excavation site and the fahlore district there are reports for three consecutive years, creating six reports that were further processed. The documentation contains various file formats such as MS Word documents for reports and lists, PDF documents containing scans of handwritten protocols, Autodesk .dwg or .dxf files for plans and .jpg images for photos and similar material.

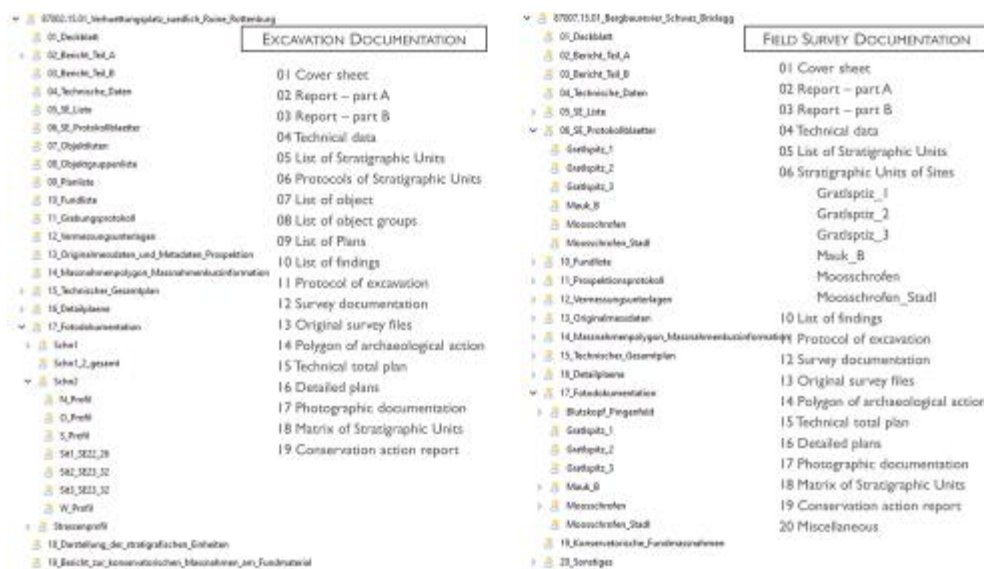


Figure 2: Categories and directory structure of the documentation as defined by the Federal Monuments Office. English translations are on the right of the folder illustrations



2.1. Convert Federal Monuments Office documentation

In order to convert these formats into long-lasting file formats using the Archaeology Data Service (Archaeology Data Service [2019](#)) and IANUS Guidelines (IANUS [2017](#)), a Python program was written that also creates a metadata file with unique identifiers and the links to the location on Google Drive where these files are stored. The long-lasting file formats were converted to a zip file, using the original file structure, and put on [Zenodo](#), a registered repository for long-term preservation of data. The process is illustrated in Figure 3. The reason for this double storage is that in Zenodo the file system has to be zipped and thus the single files are not directly accessible with a specific URI. As we then extracted and identified specific sites, structures, finds and stratigraphic units that were documented in specific files, we wanted to have access in a single documentation to the various individual reports that were created.

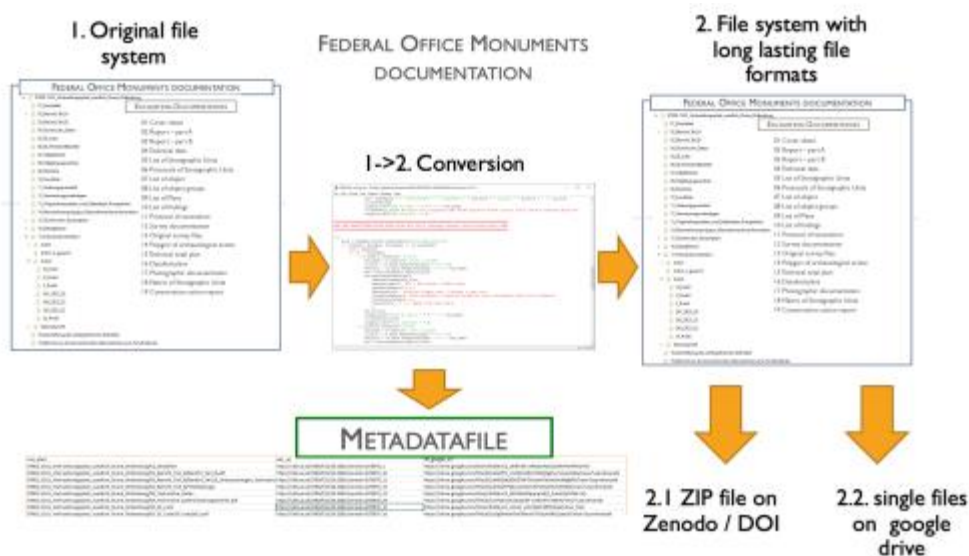


Figure 3: Conversion of the Federal Monuments Office

3. Information Extraction and RDF Creation Methodology

In our methodology there are six steps to create a knowledge graph in RDF from an existing source. First it is necessary to understand what entities are present in the source and which of them should be represented in the knowledge graph. Second, these entities should be represented in an ontology and the relevant ones selected at the level of detail to be available in the graph. As this representation determines the capabilities of querying the graph it makes sense to formulate research questions that can be answered by the graph. In particular, for archaeology the creation of a



controlled vocabulary in a hierarchical structure is essential for the query capabilities. Third, the data has to be made available in a structured format such as spreadsheets or xml. After that, data transformations and linking are performed and unique identifiers have to be created before the final step of creating the RDF network. The next sections detail and illustrate these steps for the Federal Monuments Office documentation.

3.1. Understand entities present in the documentation and their relations

Archaeological sites associated with prehistoric mining have been documented. They have been encountered during field surveys and, if considered important enough, excavations have been carried out by opening trenches and using a stratigraphic excavation methodology that attempts to dig along stratigraphic units (SU), i.e. layers deposited at specific times by specific events. Finds may have been retrieved from the surface during field surveys or from the stratigraphic units during excavations. Figure 4 illustrates the documented entities and their relations.

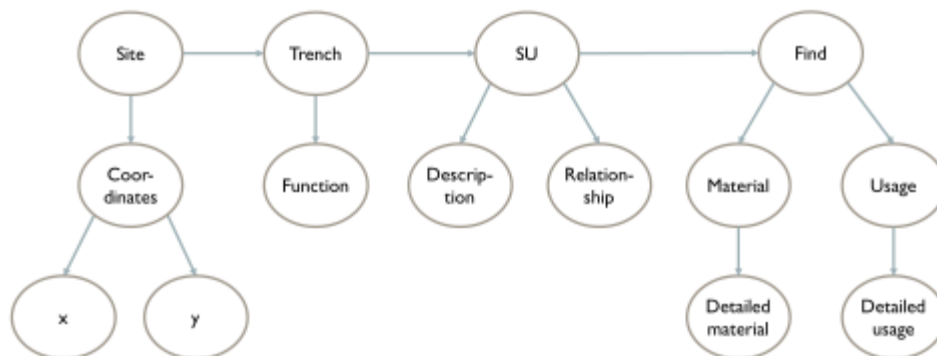


Figure 4: Items present in the Federal Monuments Office documentation

3.2. Conceptual Representation

The CIDOC CRM ontology with its extensions was used for the creation of the metadata for the entities encountered in the previous step. The CIDOC CRM is an ISO standard for cultural heritage Information, which was adopted as the conceptual background by ARIADNE. Various initiatives have arisen to promote the understanding and dissemination of CIDOC CRM like the CRM game (Guillem *et al.* [2018](#)). It was extended in the course of ARIADNE with CRMarchaeo (Doerr *et al.* [2016](#)) to model archaeological excavations. It was built based on the official documentation requirements of different countries, including the Austrian Federal



Monuments Office (Masur *et al.* 2014; Cripps *et al.* 2004). The extension CRMsci (Doerr *et al.* 2012) was used to model scientific observations. CIDOC CRM main classes and their basic relations are illustrated in Figure 5, together with the classes from CRMsci and CRMarchaeo employed to represent excavation-specific entities.

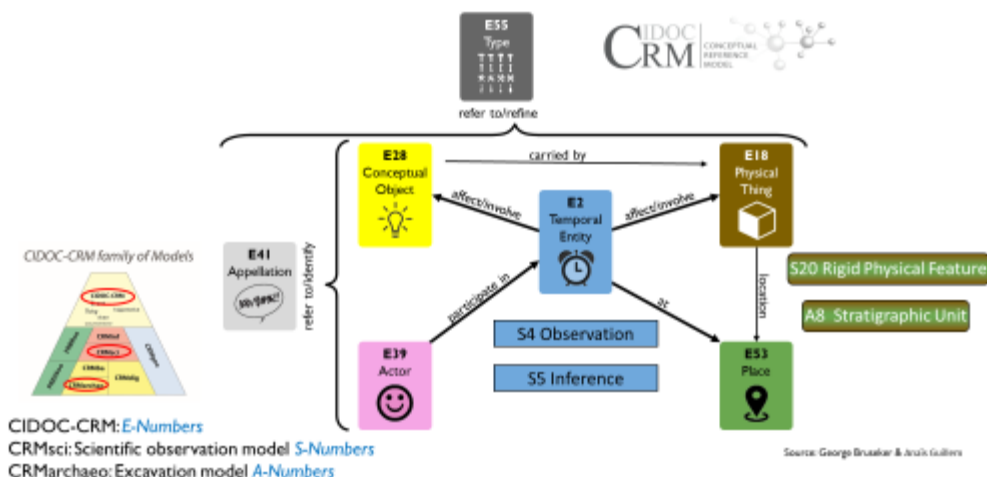


Figure 5: CIDOC CRM main classes and classes from Extensions CRMsci and CRMarchaeo used for the modelling

Archaeological sites and their investigated physical components such as trenches, stratigraphic units and finds are related with the properties illustrated in Figure 6. For all the physical components there are observations or inferences that are explicitly or implicitly stated in the documentation.

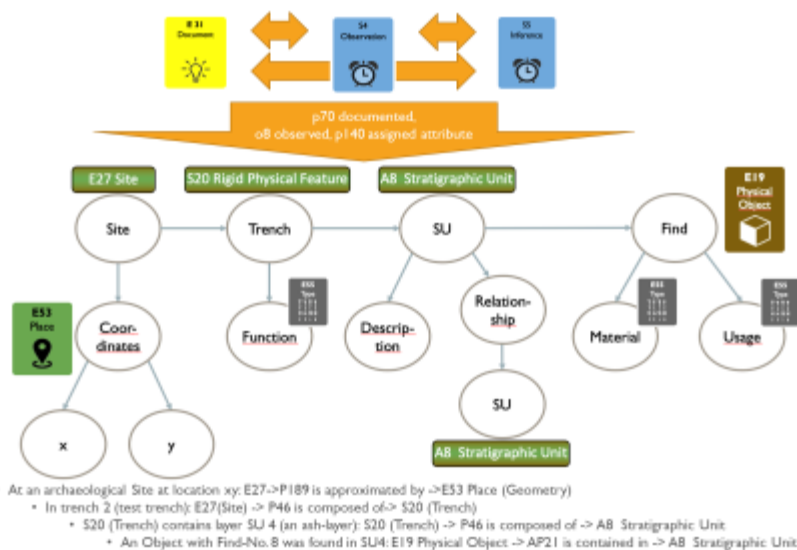


Figure 6: Modelling for entities contained in Federal Monuments Office documentation

Within the ARIADNE community the different approaches to model and implement conceptual representations of archaeological excavations were explored in a workshop (Katsianis *et al.* 2022a), presented at the EAA 2022 in Budapest



(Nenova *et al.* [2022](#)) and are detailed in Katsianis *et al.* [2023](#). The results have been documented in the Archaeological Excavation Modelling Working Group report (Katsianis *et al.* [2022b](#)).

3.3. Thesaurus creation

An essential component to represent and query the knowledge contained in the documentation is a controlled vocabulary. Concepts specific to mining archaeology research were organised with the DARIAH Backbone Thesaurus (Doerr *et al.* [2022](#)), a model for sustainable interoperable thesauri maintenance, developed in the European Union Digital Research Infrastructure for the Arts and Humanities (DARIAH). [Simple Knowledge Organization System](#) is a semantic web standard for sharing and linking knowledge organisation systems such as thesauri, taxonomies, classification schemes and subject heading systems and it was used to represent our vocabularies in RDF. One essential step to use the created data for aggregation in ARIADNE was to map the concepts used in the research data to the concepts of the *Getty Art and Architecture Thesaurus* and to [PeriodO](#) a collaborative gazetteer of chronological definitions that allows the association of a period name (e.g. 'Late Bronze Age') with a start and an end date, a corresponding geographic coverage, plus bibliographic reference through unique permalinks, e.g. <http://n2t.net/ark:/99152/p0qhb66cifs> for the late Bronze Age in Austria from 3300 BP to 2800 BP. Getty vocabularies are an evolving hierarchical knowledge base of cultural heritage terms (Harpring [2022](#)) that are available as *Linked Open Data* (LOD).

We used spreadsheets to collect terms with their translations, concept details and to create identifiers. Subsequently we placed them under the DARIAH Backbone Thesaurus hierarchy in mind-mapping software to organise the hierarchical order as shown in Figure 7. Concepts are handled together with their identifiers in a 'concept@identifier' string to facilitate the subsequent transfer to the database, as explained in Section 3.5. In order to unite different thesauri with their identifiers a thesaurus prefix is used, such as bbt for the DARIAH Backbone Thesaurus, aat for the Getty AAT or himat for the mining-specific concepts of the research centre HiMAT.



Central Topic	
▼ activities@bbt:00001 aat:300404112	▼ material things@bbt:00004 aat:300311889
▶ disciplines@bbt:00010 himat:175420 aat:300056134	▶ mobile objects@bbt:000017 himat:175860 aat:300310093
▶ human interactions@bbt:000011 himat:238106	▶ built environment@bbt:000018 aat:300254226
▶ intentional destructions@bbt:00002 aat:300059085	▼ physical features@bbt:000019 himat:175592
▶ functions@bbt:00013	▶ Vegetation@himat:175644 aat:300008932
▶ other activities@bbt:000014	▶ Geologie@himat:175593
▶ natural processes@bbt:00002 himat:175753	▶ Hydrologie@himat:175611 aat:300387652
▶ materials@bbt:00003 himat:175511	▼ Hergestelltes Merkmal (E25) himat:238106 aat:300224592
▼ material things@bbt:00004 aat:300311889	▶ Alm@himat:175691 aat:300008635
▶ mobile objects@bbt:000017 himat:175860 aat:300310093	▶ Deponierung@himat:175727 aat:300195474
▶ built environment@bbt:000018 aat:300254226	▶ Gewässerfund@himat:175728
▶ physical features@bbt:000019 himat:175592	▶ Passfund@himat:175729
▶ structural parts of material objects@bbt:000020	▶ Bestattungsplatz@himat:175730 aat:300387004
▼ types of epochs@bbt:00005 himat:175475 aat:300264088	▶ Religion/Sakrales@himat:175746 aat:300412063
▼ Epochen@himat:175476	▼ Bergbau Merkmal@himat:238107 Mining Features
▶ Epoche Kaiserzeit@himat:175490 aat:300020541 po:p0qb66wmg	▶ Schmiede@himat:175675 Forge
▶ Spätantike@himat:175491	▶ Verhüttungsplatz@himat:240207 hir Smelting site
▶ Mittelalter@himat:175492 aat:300020756 po:p0qb66433w	▶ Aufbereitungsplatz@himat:240241 Ore Processing
▶ Neuzeit@himat:175493 po:p0qb666rmd	▶ Abbau Merkmal@himat:240228 aat: Ore Extraction
▶ Prähistorisch@himat:240000 PRD_period:38 aat:300019257	▶ Ausgrabungsplatz@himat:240062 aat:300266151
▶ Archaisch@himat:240033 aat:300020985	
▶ Geologische Zeiträume@himat:175499 aat:300391250	
▶ conceptual objects@bbt:00006 himat:175419	
▶ groups and collectives@bbt:00007 himat:175981 aat:300025948	
▶ roles@bbt:00008 aat:300435108	
▶ geopolitical units@bbt:00009	

Figure 7: Concepts under the DARIAH Back Bone Thesaurus hierarchy with identifiers from Getty AAT and PeriodO

3.4. Data entry in excel sheets

An essential quality of the methodology is to make it accessible to people without programming skills, as far as possible using available tools that they already know. From experience in the project '[A Puzzle in 4D - Tell el-Daba](#)', where over 50 years of excavation reports have been digitised and metadata created (Aspöck *et al.* 2020), we opted for data entry in a spreadsheet. Five tables were created in an Excel spreadsheet to represent Sites & Structures (E27/S20), Objects (E19), Research Activities (S4), Stratigraphic Units (A8) and Documents (E31) that were explicitly or implicitly referenced in the documentation.

The manual assignment and management of identifiers was critical to link the entities to each other, especially as these appear in various source documents. As mentioned above, the linking from concepts to the Thesaurus created in Section 3.3 was achieved using a concept@identifier string. To create 1:n relationships, several concepts or identifiers had to be linked to one entity (like a site, structure or find). In order to do this, a separator (in our case '|') was used, allowing more than one concept or identifier in the corresponding cell. Therefore, the structure of a relational database can be re-created in the tables. n:m relations can be represented with this structure as well as making the editing and viewing easy. A drawback of this data entry methodology is that the possibilities to control the entry are limited, whereas the advantage is that there is no need to develop a database interface for data entry and that information can easily be copied into many cells, thus enabling quicker data entry. Figure 8 shows the sheets for Sites & Structures (E27/S20), Objects (E19) and Stratigraphic Units (A8), indicating how identifiers relate to other sheets.

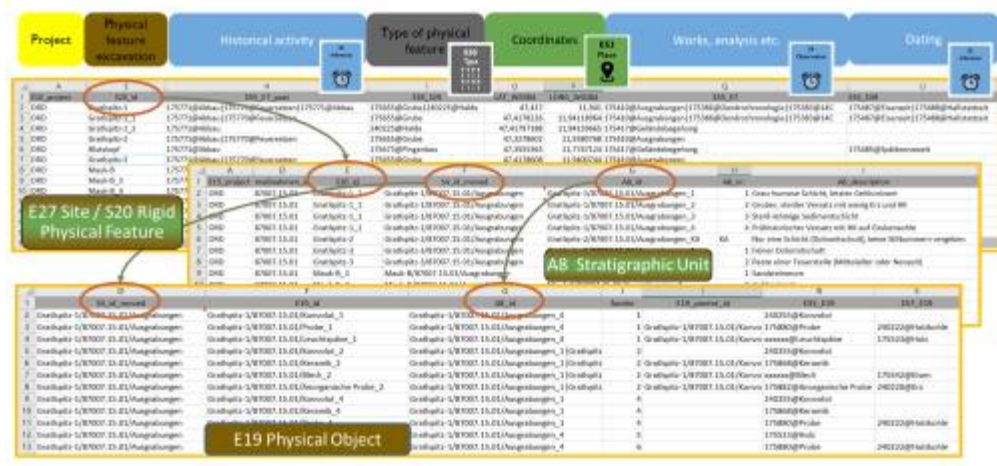


Figure 8: Excel sheets used for data entry

3.5. Data linking, conversion and identifier creation in database

The Excel sheets described in the previous section together with the files of the Thesaurus creation were imported into a Postgres database. The next step was to apply specific transformations to the data, executed with SQL scripts. Where necessary, information from different sheets were linked together through the identifiers. The third task in the database was the creation of URI identifiers which are necessary for the RDF creation. These were generated in specific 'uri' tables.

For the thesaurus file containing the concept hierarchy that was exported from the mind map was transformed to an id/parent_id structure, allowing the RDF creation process to be independent from the hierarchy depth of the thesaurus. In the 'RDF Creation pipeline' handbook there are links to the SQL scripts for the transformations (see Section 5).

3.6. RDF creation

The 'uri' tables created in the Postgres database are the input for the RDF creation. Various tools may be used for this step like [Karma](#), [OntoText-Refine](#) or the X3ML toolkit (Theodoridou and Kritsotakis 2022). As we directly accessed the relational tables in the database we used Ontotext Refine and stored the triples in an [Ontotext GraphDB](#) triple store. The ontologies and the prefixes have to be loaded into a repository of GraphDB, in order to support the process of creating the mapping that was defined in Section 3.2.



4. ARIADNE Data Creation and Portal Presentation

To create ARIADNE aggregation data, additional information specific to the ARIADNE Catalogue had to be entered in an additional 'ADRIANE Metadata' Excel spreadsheet and a mapping to the ARIADNE ontology - the AO-Cat – (Felicetti *et al.* [2023](#)) was performed to create the RDF necessary for the ARIADNE portal. One of the challenges to reusing the network previously described was the fact that the ARIADNE portal uses their own URIs, which are specified and only the last part is optional definable. The ARIADNE Team, who supported us very much in the process, recommended that we create the ARIADNE URIs in the Postgres database through a hash function of the original URIs. The RDF creation pipeline including the 'ARIADNE Metadata' Excel spreadsheet is illustrated in Figure 9. It shows the process from Excel sheets/Thesaurus over the postgres database to the RDF representation in the triple store.

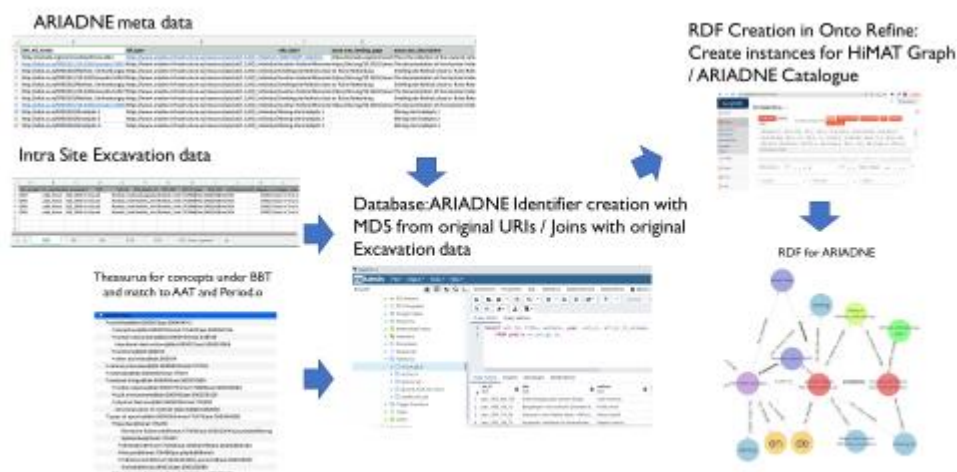


Figure 9: Import of Excel sheets and thesaurus into Postgres database for data transformation with subsequent RDF creation

This strategy allows the re-creation of the same ARIADNE URI from original URIs. In the ARIADNE portal a collection was created for the [prehistoric mining sites in the Lower Inn Valley](#) processed in the Open Research Data Pilot project (ORD) of the University of Innsbruck (UIBK). In the collection the different Federal Monuments Office documentations are represented as 'Fieldwork archive' with a link to the zip file deposited on Zenodo. The sites documented in the Fieldwork archives are represented as 'Site/monument' and link to the [Knowledge graph](#) of the Research Center HiMAT. In the knowledge graph it is possible to navigate to the links on Google Drive of the single folders/documents related to a specific site/trench/stratigraphic unit or find. Both options are illustrated in Figure 10. Of course it is also possible to query the graph with SPARQL to retrieve the links to the documents. In the collection there are only 18 records but through the links to the



knowledge graph many more detailed records of the sites are available. The goal is to apply the pipeline to other archaeological research at the University of Innsbruck.

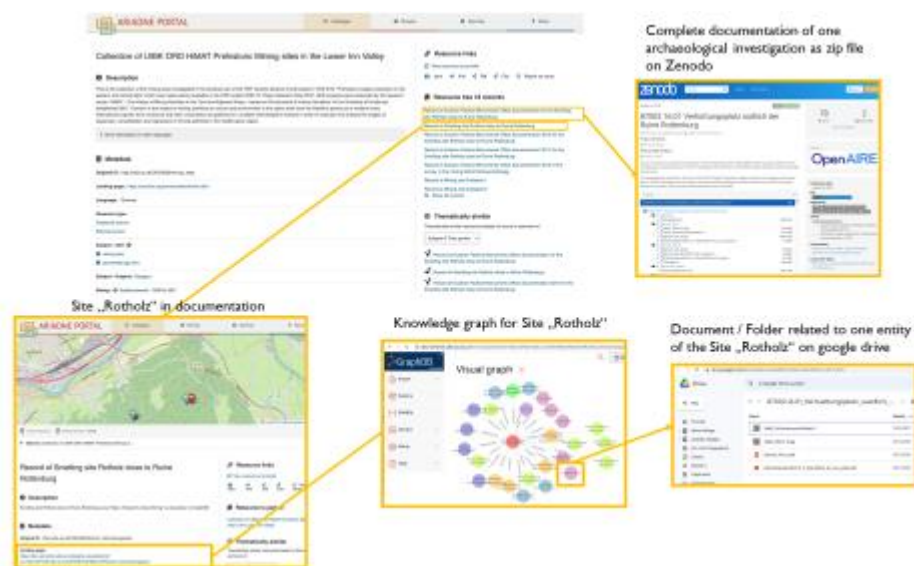


Figure 10: Federal Monuments Office documentation as fieldwork archive in ARIADNE portal linking to Zenodo and sites within archive linking to the HiMAT knowledge graph

5. RDF Creation Pipeline Handbook

In the course of the project a handbook has been produced (Figure 11) detailing the methodology and providing sample data so that anybody interested can walk through the examples and create RDF data using the tools and data (Hiebel and Peralta Friedburg 2023).

RDF Creation Pipeline		RDF Creation Pipeline		RDF Creation Pipeline	
1. Introduction	4	5.2.1. Alpine Magazine	38	5.3.1. Create a repository in Triple Store	58
2. Handbook and Data Structure	4	5.2.2. Excavation	39	5.3.2. Load data in specific graph in repository	58
3. Software Installation	4	5.2.3. Thesaurus	39	5.3.2.1. Alpine Magazine	58
3.1. Topshelf	5	5.3. Load data into Postgres	39	5.3.2.2. Excavation & Thesaurus	62
3.2. Postgres	5	5.3.1. Move data to input directory for postgres	39	5.3.2.3. Triple Store	63
3.2.1. Install PostGIS	7	5.3.1.1. User-scale postgres installation	39	5.3.2.4. Create a repository in Triple Store	64
3.3. GraphDB	18	5.3.1.2. Create database and postgres extension	39	5.3.2.5. Load Data	64
3.4. OntoBake Refine	18	5.3.2.1. Create database	39	5.3.3. Graph DB Settings	66
3.4.1. OntoBake Refine script	13	5.3.2.2. Create PostGIS Extension in database	34	5.3.3.1. Autocomplete	66
4. Ontologies and Modeling	17	5.3.3. SGA Scripts to load data	35	5.4. Data access and visualization	67
4.1. Formal ontologies	17	5.3.3.1. Alpine Magazine SQL Load Files	38	5.4.1. Create Repository and access named graph in repository	67
4.2. The CIDOC Conceptual Reference Model (CIDOC CRM)	18	5.3.3.2. Excavation SGA Load Files	38	5.4.2.1. Alpine Magazine	67
4.2.1. CIDOC CRM - Inities	18	5.3.3.3. Thesaurus SGA Load Files	38	5.4.2.2. Excavation	67
4.2.2. Top Level Entities	20	5.4. Query / transform data & create URI identifiers (and labels) for concepts in postgres	39	5.4.2.3. Access Inference in repository and visualization	68
4.3. Resources	20	5.4.1. Query data	39	5.4.2.4. Alpine Magazine	68
4.3.1. CIDOC CRM Entities	21	5.4.1.1. Transform data	39	5.4.2.5. Excavation	69
4.3.2. CRMClasses extension	21	5.4.2. Transform data	39	5.5. SPARQL Queries	70
4.4. WQS for Thesaurus management	21	5.4.2.1. Excavation	42	5.5.1. SPARQL Query window	71
5. RDF Creation Pipeline	22	5.4.2.2. Thesaurus	42	5.5.1.1. Alpine Magazine	72
5.1. Create all/ sub defined structured data from a spreadsheet or from a database	22	5.4.3. Create URI identifiers	43		
5.1.1. Alpine Magazine	22	5.4.3.1. Alpine Magazine	44		
5.1.1.1. Excel Sheet for Alpine Magazine data	22	5.4.3.2. Excavation	45		
5.1.1.2. Generate data for Alpine Magazine	24	5.4.3.3. Thesaurus	45		
5.1.2. Excavation	24	5.5. Import in RDF creation software	45		
5.1.2.1. Excel Sheet for Excavation sample for Rotholz	24	5.5.1. Load URI data in OntoBake Refine from Database	45		
5.1.3. Thesaurus	25	5.5.1.1. Onto Refine	52		
5.1.3.1. Ideal Map for Thesaurus Hierarchy	26	5.5.1.2. Thesaurus URI files	54		
5.1.3.2. Excel Sheet Thesaurus Hierarchy	28	5.6. Apply existing model or create conceptual model	54		
5.1.3.3. Excel Sheet Thesaurus Terms	28	5.6.1. Apply existing model	54		
5.2. Create conceptual model based CIDOC CRM	29	5.6.1.1. Alpine Magazine	54		
		5.6.1.2. Excavation	56		
		5.6.1.3. Thesaurus	56		
		5.6.2. Create model according to conceptual model designed in step 2	56		
		5.6.2.1. Alpine Magazine	58		
		5.7. Load RDF in triple store (InSPARQL)	58		

Figure 11: Table of contents for the RDF Creation pipeline



Currently there are sample data for three different datasets:

- Alpine Magazine
- Excavation
- Thesaurus

The structure is the same for all three, corresponding with the numbering in the handbook:

5.1 Create utf8 tab-delimited structured data from a spreadsheet or from a database

5.2 Understand your data and create a conceptual model based on CIDOC CRM ontology

5.3 Load data into Postgres

5.4 Transform data & create URI identifiers (and labels) for concepts in Postgres

5.4.1 Transform data

5.4.2 Create URI identifiers

5.5 Import in RDF creation software (Ontotext Refine)

5.6 Create conceptual model or apply existing model

5.7 Publish RDF to triple store (GraphDB)

For all three datasets there are folders and files as illustrated in Figure 12.

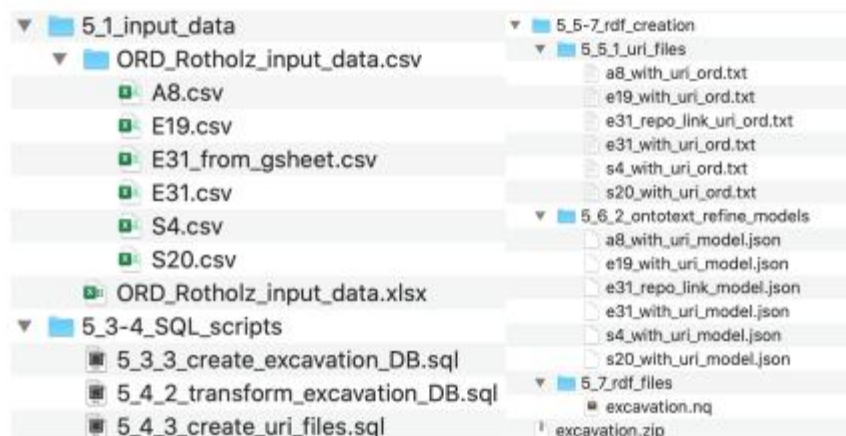


Figure 12: Structuring of the sample data for excavations



6. Conclusion and Outlook

In this article we have shown how the Austrian Federal Monuments Office documentation for the project 'Prehistoric copper production in the eastern and central Alps' was transformed to create FAIR data and a knowledge graph in RDF using the CIDOC CRM ontology with extensions. The same methodology was applied to create RDF data using the ARIADNE ontology for the ARIADNE portal. The methodology section outlined the process of RDF creation and pointed to a detailed 'RDF Creation Pipeline' handbook that should make the steps repeatable. The aim of the handbook is to give others the ability to apply the methodology to their own data. Understanding of data structures and skills in data modelling with CIDOC CRM and SQL will be necessary but no programming knowledge is needed to perform the tasks. An understanding of the ARIADNE ontology is necessary to apply the pipeline in order to create the RDF for the ARIADNE portal. Sample data and modelling patterns for ARIADNE data creation are not yet included in the 'RDF Creation Pipeline' but are planned to be included in a subsequent version of the handbook. Creating ARIADNE data helped us to see what was still missing in the original data creation process. As it was the primary provenance that was not sufficiently captured and modelled, a desirable goal for future RDF generation is to include the necessary data in the modelling and create ARIADNE data simultaneously. We are currently exploring the use of named graphs and reification constructs like [RDF-star](#) to model provenance and also represent proposition sets like inferences. The Austrian Federal Monuments Office documentation of the archaeological investigations that have been carried out in the project 'Information Integration for Prehistoric Mining Archaeology' are currently being processed and will be included in the HiMAT knowledge graph as well as the ARIADNE portal.

Acknowledgements

The research presented here was financed by the Austrian Science Fund (FWF) in the course of the Stand Alone Project 'Information Integration for Prehistoric Mining Archaeology' (FWF, P 31814-N38) and 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). Federal Monuments Office documentation was created in the trinational joint project (DACH) 'Prehistoric copper production in the eastern and central Alps' (FWF, I 1670-G19).

Bibliography

Archaeology Data Service 2019 *Archaeology Data Service/Digital Antiquity Guides to Good Practice*. <https://guides.archaeologydataservice.ac.uk/g2gpwiki/> [Last accessed: 2 July 2019]

Aspöck, E., Hiebel, G., Kopetzky, K. and Durco, M. 2020 'A puzzle in 4D: archiving digital and analogue resources of the Austrian excavations at Tell el Daba, Egypt' in



E. Aspöck, S. Stuhec, K. Kopetzky and M. Kucera 'Old excavation data. What can we do?', *Proceedings of the Workshop held at the 10th ICAANE in Vienna, April 2016*. Wien: Verlag der Österreichischen Akademie der Wissenschaften (= OREA - Oriental and European Archaeology, 16). 79-100. <http://dx.doi.org/10.1553/0x003bca0e>

Bekiari, C., Christaki, E., Doerr, M., Bruseker, G., Ore, Ch.E., Stead, S., Velios, A., Erin Canning, E. and Michon, P. 2022 *Volume A: Definition of the CIDOC Conceptual Reference Model*. <https://www.cidoc-crm.org/Version/version-7.2.1> [Last accessed: 27 March 2023]

Cripps, P., Greenhalgh, A., Fellows, D., May, K. and Robinson, D., 2004 *Ontological Modelling of the Work of the Centre for Archaeology*, CIDOC CRM Technical Paper. https://CIDOC-CRM.org/sites/default/files/cidoc_crm_version_4.3_Nov08.pdf [Last accessed: 20 March 2022]

Doerr, M. 2003 'The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata', *AI Magazine* **24**, 75-92.

Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., Theodoridou, M. *et al.* 2012 *CRMsci: the Scientific Observation Model An Extension of CIDOC-CRM to support scientific observation*, Version 1.2. <https://cidoc-crm.org/crmsci/sites/default/files/CRMsci1.2.pdf> [Last accessed: 20 March 2022]

Doerr, M., Achille F., Sorin H., Hiebel, G., Kritsotaki, A., Masur, A., May, K., Ronzino, P., Schmidle, W., Theodoridou, M., Tsiafaki, D. *et al.* 2016 *Definition of the CRM Archaeo: An Extension of CIDOC CRM to support the archaeological excavation process*, Technical Report 1.4. Prato, Italy: PIN S.c.R.L. https://cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.1.pdf [Last accessed: 20 March 2022]

Doerr, M., Daskalaki, M., Goulis, H., Katsiadakis, H., Chrysovitsanos, G., Kalafata, P., Tzedopoulos, Y., Nouvel, B., Sinigaglia, E., Colombi, C., Bekiari, C., Tsoulouha, E., Harami, L. and Almeida, B. 2022 *DARIAH Backbone Thesaurus (BBT) - Definition of a model for sustainable interoperable thesauri maintenance*, Thesaurus Maintenance Working Group, VCC3, DARIAH EU, <https://www.backbonethesaurus.eu/>. [Last accessed: 29 January 2023]

Felicetti, A., Meghini, C., Richards, J.D. and Theodoridou, M. 2023 *The AO-Cat Ontology*. <https://doi.org/10.5281/zenodo.7818375>

FORCE11 2019 *Guiding Principles for Findable, Accessible, Interoperable and Re-Usable Data Publishing*, Version B1.0. <https://www.force11.org/fairprinciples> [Last accessed: 14 November 2019]

Goldenberg, G. 2021 'Prehistoric mining, beneficiation and smelting of copper ores in the Lower Inn Valley, North Tyrol, Austria' in P. Bellintani and E. Silvestri, E. *Fare Rame. La metallurgia primaria della tarda età del Bronzo in Trentino: nuovi scavi e stato dell'arte della ricerca sul campo*, Trento: Provincia autonoma di Trento. 223-36.



Guillem, A., Bruseker G. T. and Lercari, N. 2018 'Learning through doing: an online game for appropriating ontological modelling methodology', Paper presented at 46th Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2018), Tübingen, Mar. 19-23.

Harpring, P. 2022. Getty Vocabularies and Linked Open Data (LOD). https://www.getty.edu/research/tools/vocabularies/Linked_Data_Getty_Vocabularies.pdf [Last accessed: 20 October 2023]

Hiebel, G. and Friedburg, P. 2023 *RDF Creation Pipeline*. <https://doi.org/10.5281/zenodo.7808036>

Hiebel, G., Goldenberg, G., Grutsch, C., Hanke, K. and Staudt, M. 2019 'FAIR data for prehistoric mining archaeology', *International Journal on Digital Libraries* **22**, 267–77. <https://doi.org/10.1007/s00799-020-00282-8>

IANUS 2017 IANUS - IT-Empfehlungen. <https://ianus-fdz.de/it-empfehlungen/>

Katsianis, M., Styliaras, G., Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Kritsotakis, V., Martlet, O., Nenova, D., Nurra F., Ore, C-E. and Theodoridou, M. 2022a *ARIADNEplus Virtual Workshop on Semantic mapping of archaeological excavation data*, ARIADNE Project. <https://doi.org/10.5281/zenodo.7112918>

Katsianis, M., Nenova, D. Bruseker, G., Derudas, P., Felicetti, A., Hiebel, G., Hivert, F., Martlet, O., Opitz, R., Richards, J., Roulet, T., Styliaras, G., Ore, C-E. and Uleberg, E. 2022b *Archaeological Excavation Modelling Working Group: WP 4.4.12 excavation data*, ARIADNE Project. <https://doi.org/10.5281/zenodo.7377910>

Katsianis, M., Bruseker, G., Nenova, D., Marlet, O., Hivert, F., Hiebel, G., Ore, C.-E., Derudas, P., Opitz, R. and Uleberg, E. 2023 'Semantic Modelling of Archaeological Excavation Data. A review of the current state of the art and a roadmap of activities', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.12>

Masur, A., May, K., Hiebel, G. and Aspöck, E. 2014 'Comparing and mapping archaeological excavation data from different recording systems for integration using ontologies', *Proceedings of the 18th International Conference on Cultural Heritage and New Technologies 2013 (CHNT 18, 2013)*. https://archiv.chnt.at/wp-content/uploads/Masur_etal_2014.pdf

Nenova, D., Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Katsianis, M., Marlet, O., Opitz, R., Ore, C.-E. and Uleberg, E. 2022 'Bringing excavation data together. Are we there yet and where is that?', Paper presented at *28th EAA Annual Meeting (EAA 2022)*, Budapest. <https://doi.org/10.5281/zenodo.7117049>.

Niccolucci, F. and Richards, J.D. 2013 'ARIADNE: Advanced research infrastructures for archaeological dataset networking in Europe. A new project to foster and support archaeological data sharing', *The European Archaeologist* **39**, Summer 2013.



Theodoridou, M. and Kritsotakis, V. 2022 'The X3ML toolkit: how to map excavation data to CIDOC CRM', ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological Excavation Data, 15 June 2022, Zenodo. <https://doi.org/10.5281/zenodo.7108577> (Presentation video accessible at <https://vimeo.com/753116253>) [Last accessed: 20 March 2022].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva, Bonino, Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenberg, P., Wolstencroft, K., Zhao, J. and Mons, B. 2016 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data* **3**, 160018. <https://doi.org/10.1038/sdata.2016.18>