Der Anschnitt, Beih. 42, S. 57 – 64, 2019

Gerald Hiebel, Gert Goldenberg, Caroline Grutsch, Klaus Hanke, Markus Staudt

A methodology to integrate information in prehistoric mining archaeology research

ABSTRACT: *In prehistoric mining archaeology research, different scientific areas are working together to solve fundamental research questions related to the production and use of metals in prehistoric times. They produce information according to the methodologies of their discipline and in the specific formats that best document their observations and research. These formats may be very different as the documentation of archaeological or geological prospection, archaeological excavations, surveying, mineralogical, geochemical and metallurgical analysis, experimental archaeology and ethnoarchaeology vary significantly even within one discipline. In this paper we want to present a methodology how to integrate information of the above mentioned disciplines in order to answer specific questions in prehistoric mining archaeology research. The methodology is based on building a conceptual model that is able to represent the information of the participating disciplines and to inform the archaeologist on observations, physical structures, geochemical ore/artefact signatures and interpretations related to prehistoric mining, ore processing and extractive metallurgy with all the employed technologies. The CIDOC CRM ontology is used as conceptual background to structure the data. After the main categories of the conceptual model have been defined they need to be specialised in a hierarchical thesaurus. The transformation and mapping of existing data to the conceptual model is realized with semantic web technologies. The integration of the information and the subsequent provision in adequate tools can help researchers of prehistoric mining archaeology significantly to answer their research questions.*

KEYWORDS: INFORMATION INTEGRATION, MINING ARCHAEOLOGY, SEMANTIC TECHNOLOGIES, CIDOC CRM, GEOINFORMATION

Introduction – Why is information integration necessary?

The need for information integration arises from the quest to solve the fundamental research questions related to prehistoric mining, ore processing and extractive metallurgy with all the employed technologies. As different scientific disciplines work on these questions the results they produce correspond to their practices and standards, which may be very different from each other. Looking at prospections, archaeological excavations, surveying, mineralogical, geochemical and metallurgical analysis, experimental archaeology and ethnoarchaeology it is obvious that they use different documentation methodologies. But they still want to answer specific research questions:

- Where are physical structures relevant for prehistoric mining/metallurgy?
- Which information do we have about physical things (structures/objects/samples) that are relevant for prehistoric mining/metallurgy research?
- To which technologies do these physical things relate?
- How are physical structures/objects/samples dated?
- What are the geochemical signatures of samples and from what objects or structures have they been taken?
- Which observations/archaeometric analysis/interpretations have been conducted on physical things?
- What can we learn from experimental Archaeology and Ethnoarchaeology about technology and the production process of objects/structures related to metallurgy?

The integration of information and its subsequent provision in adequate tools can certainly help researchers of prehistoric mining archaeology significantly to answer these questions.

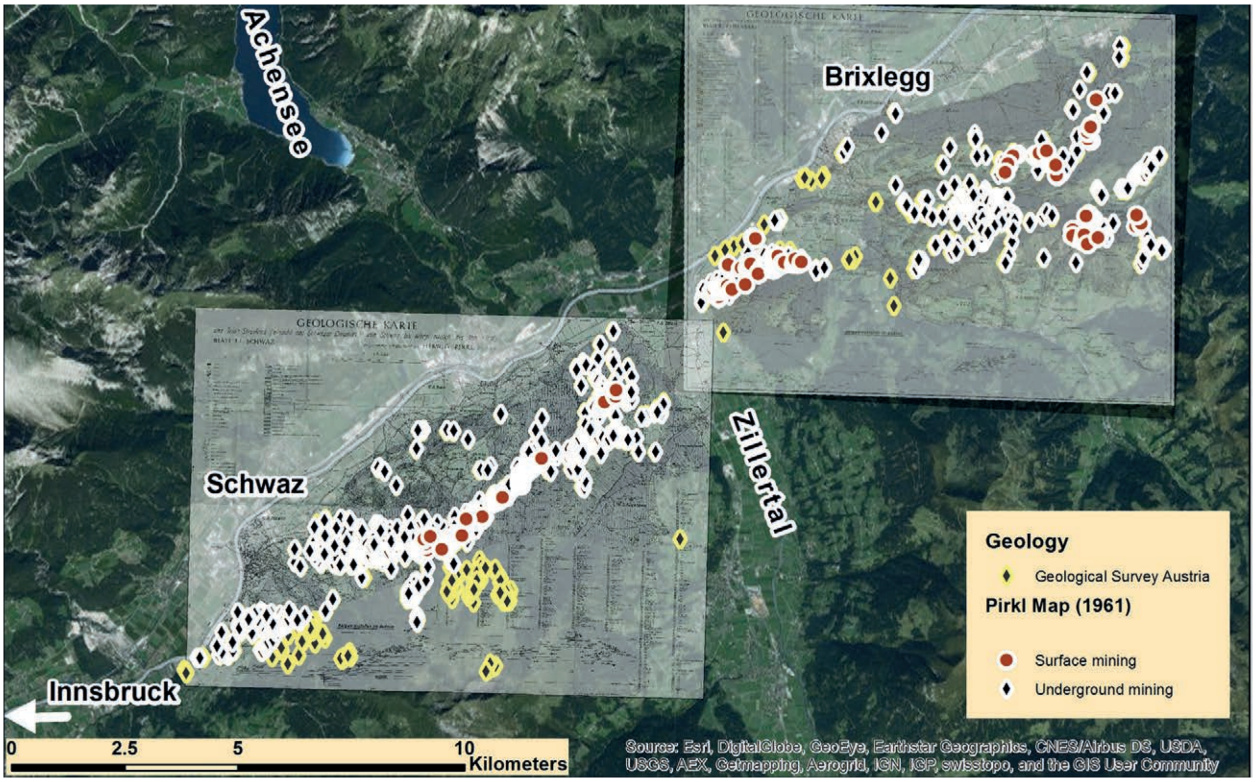

Fig. 1: Mining structures identified by Pirkl and the Geological Survey Austria (source: GBA, 2014).

The first step is to identify the information sources that hold the knowledge which is relevant to these questions.

What information sources hold relevant knowledge?

Scientific research is often organized in projects or specific structures. One of them related to prehistoric mining archaeology is the research centre RC HiMAT of the University of Innsbruck (The History of Mining Activities in the Tyrol and adjacent areas – impact on environment and human societies - www.uibk.ac.at/himat/) that investigates the mining history of the Eastern Alps from Prehistory to Modern Times. As examples for information sources available within the interdisciplinary work of the research centre we choose the area of Schwaz/Brixlegg in North Tyrol, Austria, where the localization, identification and interpretation of mining structures are targeted. Geological prospections are a fundamental information source about structures originating from mining activities. Herwig Pirkl (1961) investigated the Schwaz/Brixlegg mining area in a detail that has not been repeated since. The result was a publication describing the geologic and surface structures of the area. It contains three geological maps in the scale 1:10,000. Two of these maps have been digitized in the course of the works realized in the RC HiMAT. Structures identified by Pirkl as underground mining and surface mining have been registered together with their names and coordinates. In addition, information on mining structures provided by the Geological Survey Austria (GBA, 2014) has been integrated (Fig. 1).

Another source of information for the localization of mining structures is the digital high resolution elevation model (DEM) of the province of Tyrol. The Unit for Surveying and Geoinformation of the research centre examined the DEM for concave and convex surface structures that are in proximity of the structures identified by Pirkl (Fig. 2).

HiMAT's participation in the international research project "Prehistoric copper production in the eastern and central Alps – technical, social and economic dynamics in space and time" (financed by the Austrian Science Fund FWF, the German Research Foundation DFG and the Swiss National Science Foundation SNSF, 2015 - 2018) contributed significantly to the knowledge of archaeological sites related to prehistoric mining activities through archaeological prospections and excavations. In figure 3 these activities are illustrated together with find spots that have been extracted from archaeological literature. A detailed map of the area around a prehistoric mining structure called "Bauernzeche" shows the potential of integrating these sources together with a DEM vizualization (Fig.4). In the map we see the landscape relief and structures created through mining activities which were first observed by Pirkl and documented in geological maps. Decades later these structures were prospected

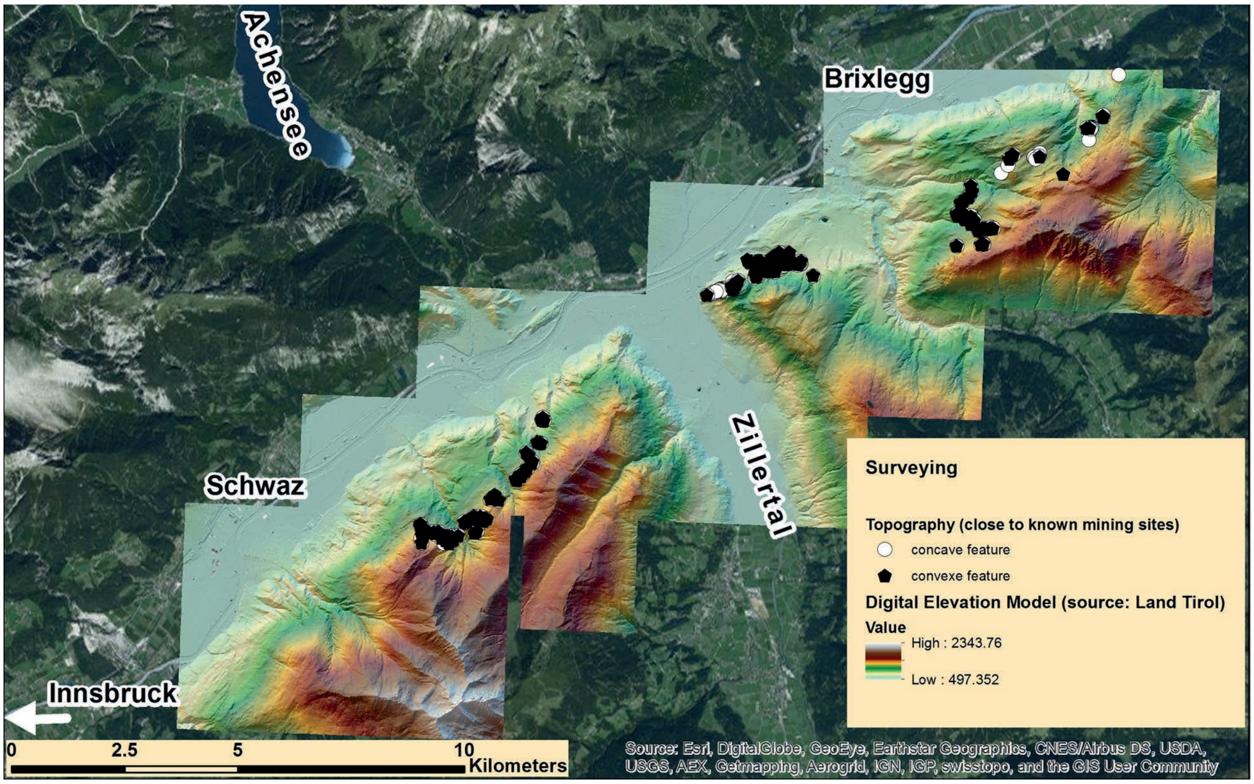

Fig 2: Surface structures identified in the high resolution elevation model of the province of Tyrol (source: Land Tirol 2009).

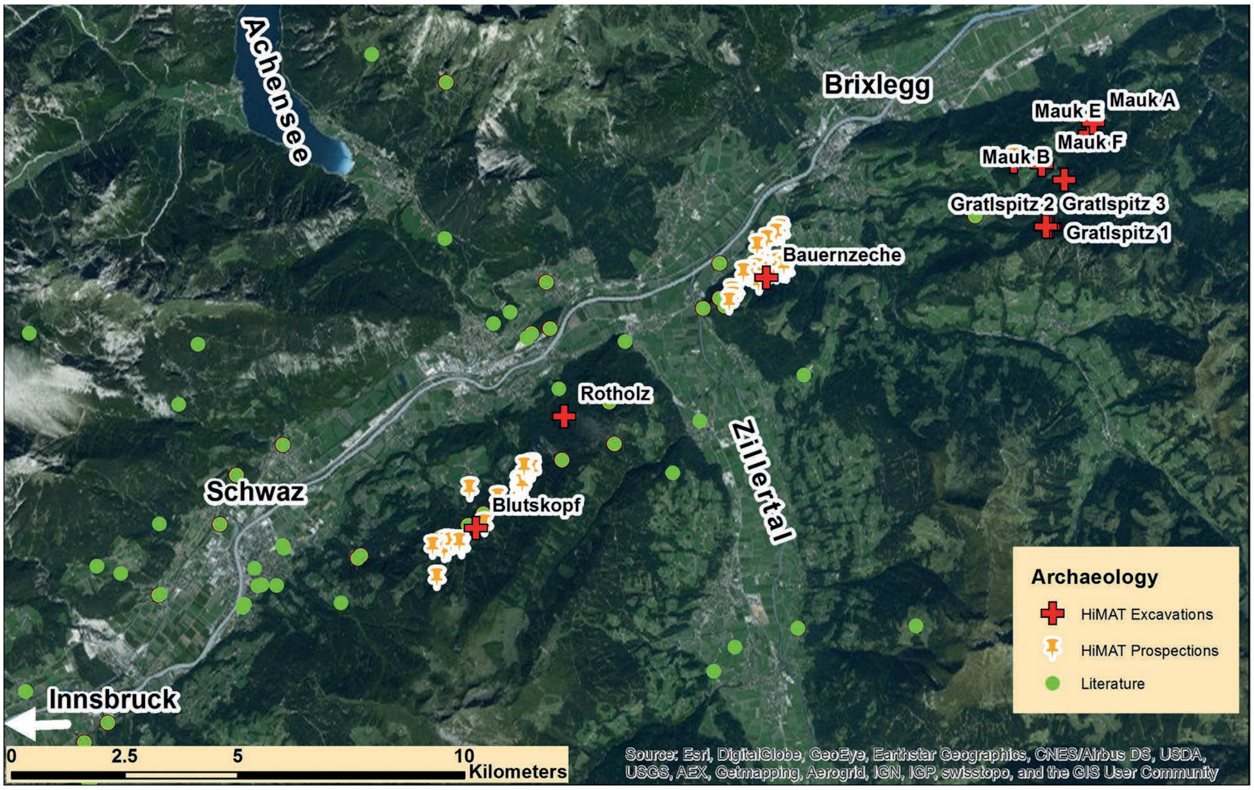

Fig. 3: RC HiMAT excavation and survey activities and archaeological sites extracted from archaeological literature (source: RC HiMAT, ESRI).

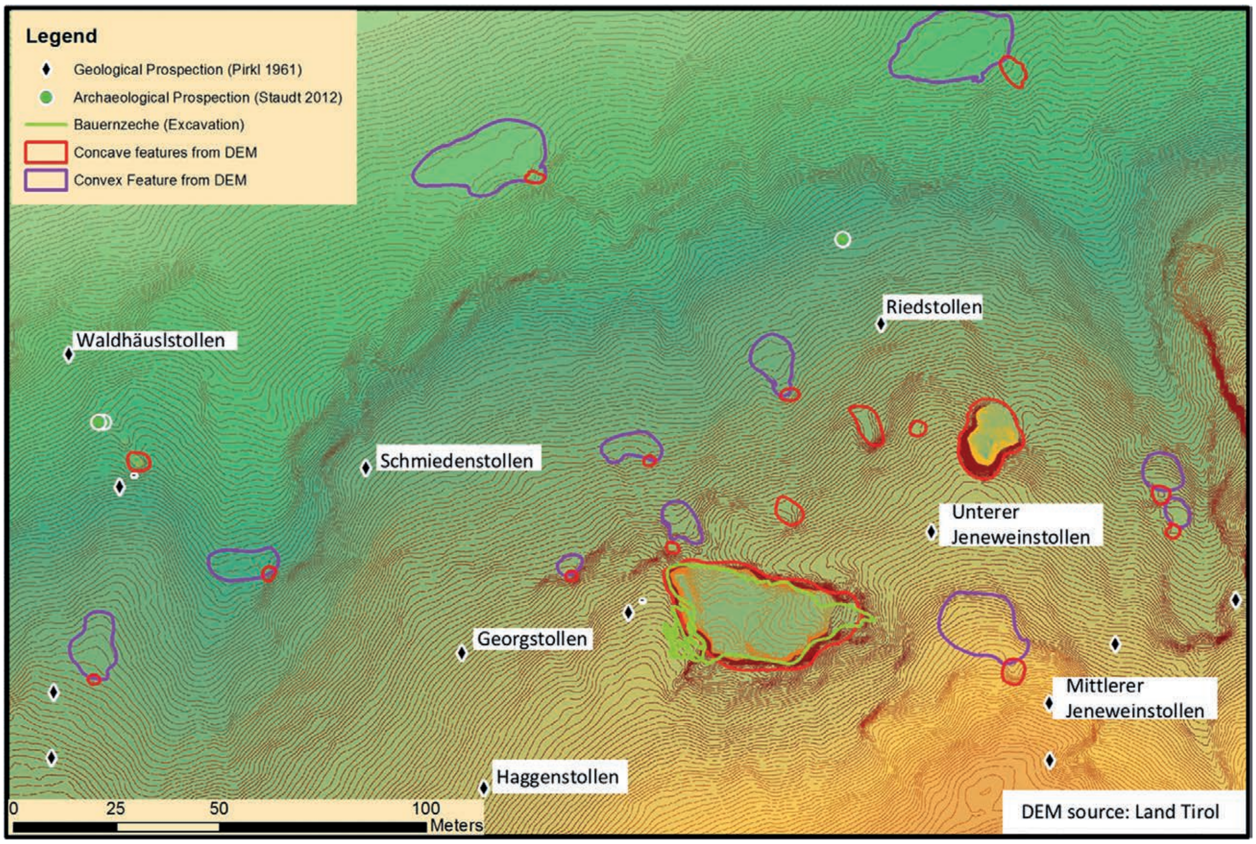

Fig. 4: Information sources that indicate physical structures related to mining activities.

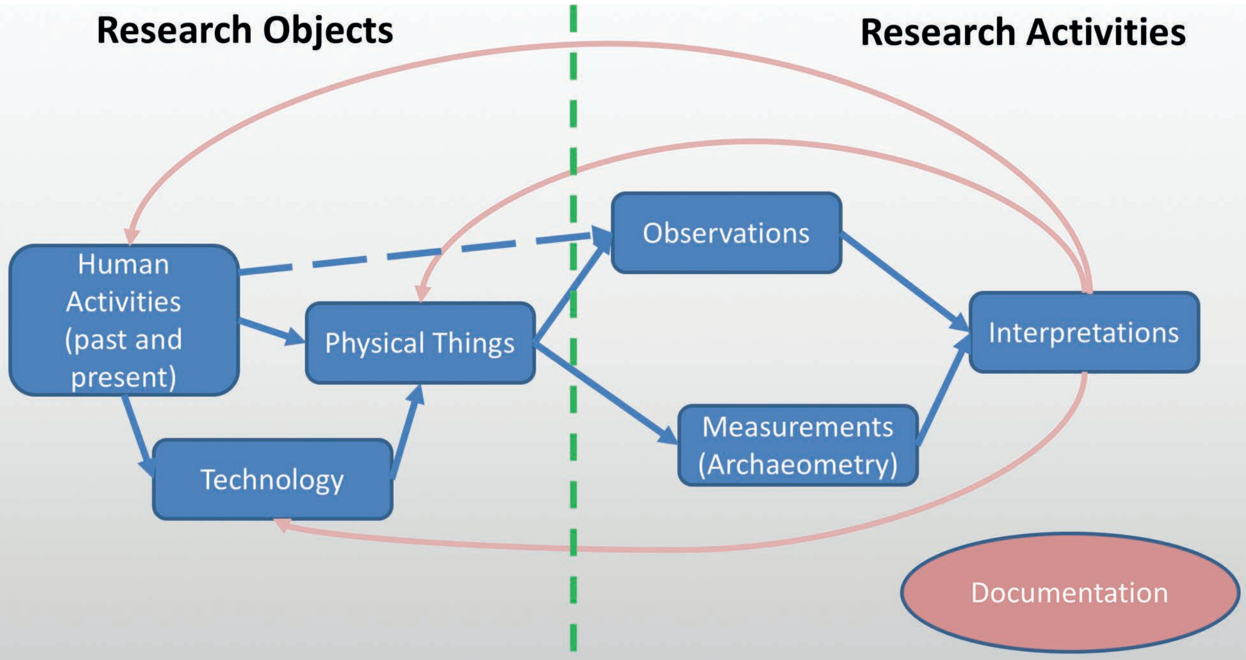

Fig. 5: Research objects and research activities in a network model.

again by RC HiMAT archaeologists and documented in field survey records. Then the high resolution digital elevation model was analysed and relevant physical structures in the vicinity of the observations made by Pirkl and the archaeologists were marked (Fig. 4). For the data sources displayed in figure 4 observations and analysis of these physical mining structures like pits, heaps, beneficiation and smelting sites are documented. The same physical structure may have been observed or analysed several times by different humans or different

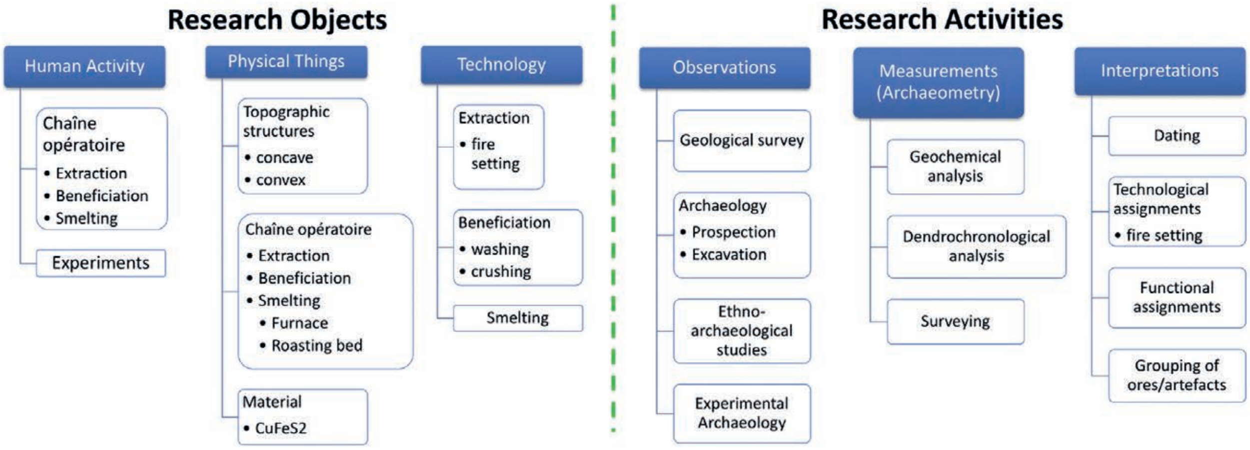

Fig. 6: Refinement of research objects and research activities in a thesaurus.

scientific methodologies. In addition to the observations there are interpretations of the physical structures as to their functionality and their dating based on observations and analysis. As shown in the illustration Geoinformation systems use layers to represent information. When used in archaeology physical structures, observations, analysis and interpretations are either mixed in one layer or there exist several layers for different information sources. On the represented area we find recent and historic observations of mining structures and there has been an excavation in the so called “Bauernzeche” where prehistoric mining activities using the fire setting technique have been verified. To integrate the information of observations, analysis and excavations together with the interpretations a network model is necessary that goes beyond the layer structure used in Geoinformation systems. This model should help archaeologists to make use of all the information available in one area and build a network that can be explored for further research.

Creating this network model the other information sources related to archaeometrical and archaeometallurgical analysis, experimental archaeology and ethnoarchaeology have to be taken into account.

How to integrate on a theoretical level? – The conceptual model

To integrate heterogeneous information sources a conceptual model that has the ability to represent the concepts coming from different domains and in particular from different methodologies and documentation practices is a necessity. The CIDOC CRM ontology (Le Boeuf et al., 2018) is an ISO standard in Cultural Heritage Documentation and an event centric data model that fulfills these requirements. Past mining activities and contemporary research activities (which are subclasses of events) are the essential nodes in the model that relate research objects with research activities as displayed in figure 5. The documentation of the research activities are the information sources shown in the previous chapter.

Extensions of the CIDOC CRM (2016) were used to model observations (CRMsci), interpretations (CRMinf), spatial information (CRMgeo) and digital provenance (CRMdig). The classes of the model had to be refined with a thesaurus (Fig. 6) in order to represent the detailed information of the available documentation and to answer research questions relevant for the domain. The integration of vocabularies originating from different sources is a serious challenge (Doerr, 2006). Within the DARIAH Infrastructure (www.dariah.eu) an approach was developed to integrate terms within a backbone thesaurus and thus create the ability to query upper levels without the need to reach consensus on lower level terms which is often an almost impossible task to accomplish (Dariah EU, 2016).

Figure 7 shows how to apply the model to the past human activity of ore extraction through fire setting. This mining activity with the specific technology created the physical structure of mine “Mauk E”. It was excavated and dendrochronological measurements have been conducted on wood samples, as well as a 3D documentation of the mine and the surrounding terrain took place. The interpretation from observations and measurements state that ore extraction took place between 720 to 707 BC and that the fire setting technology was used.

How to integrate on a practical level? – The implementation

The available information has to be converted in a structured format, either being a tabular format, a relational database or an XML structure. The terms used in the existing documentation have to be aligned to the thesaurus to obtain data that can be mapped to the classes of the ontology using the specialisations provided by the thesaurus. Karma (ISI, 2016), a tool of the semantic web community was used to map the information sources to

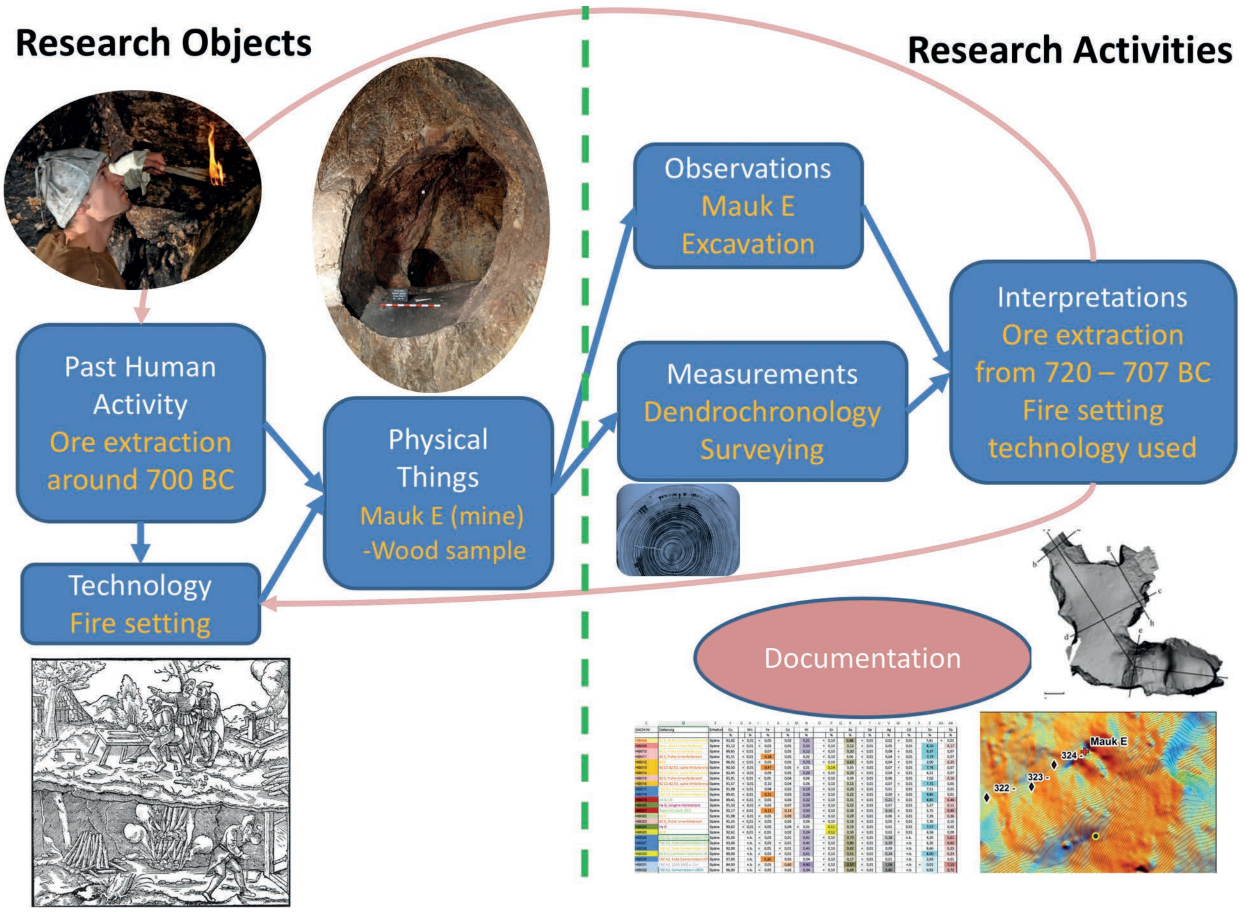

Fig. 7: Example of the modelling for the mine "Mauk E".

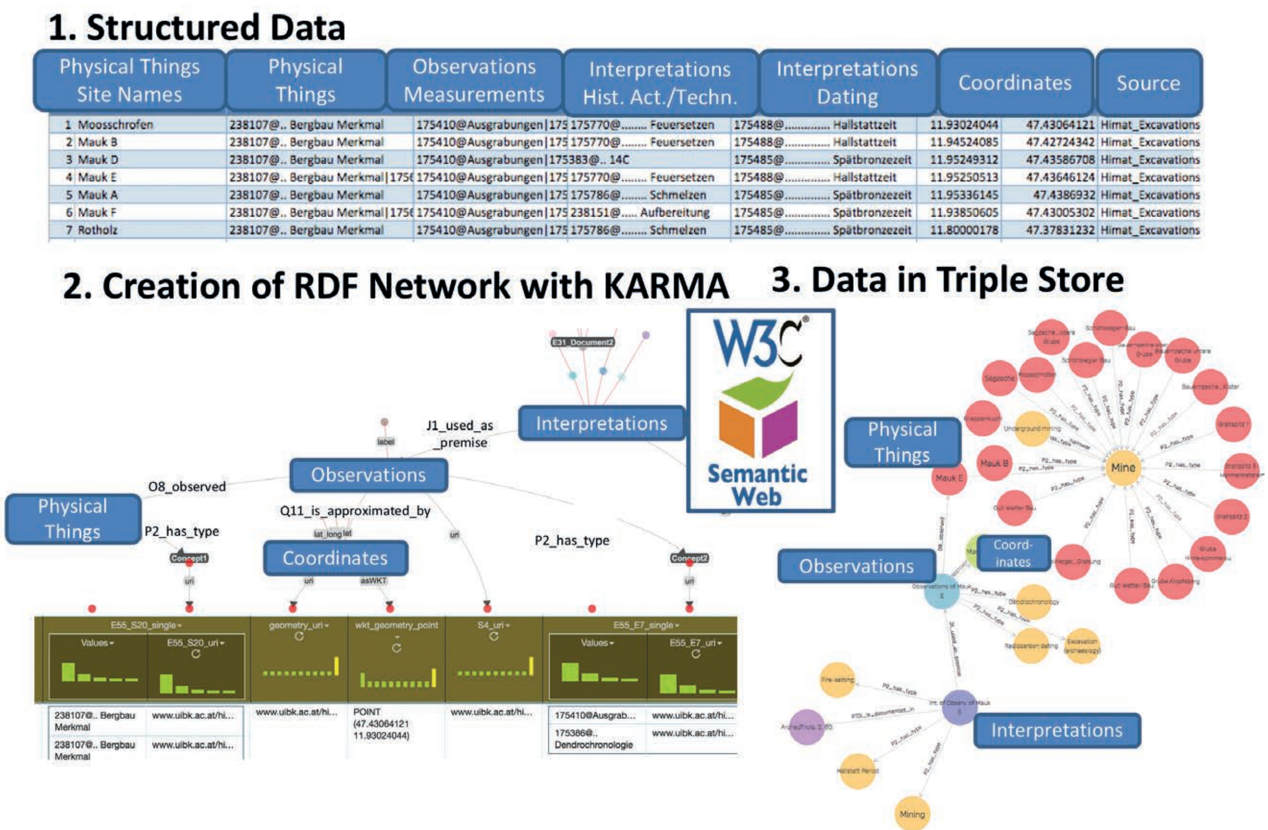

Fig. 8: Using Karma to map structured data to the formal definitions of the CIDOC CRM.

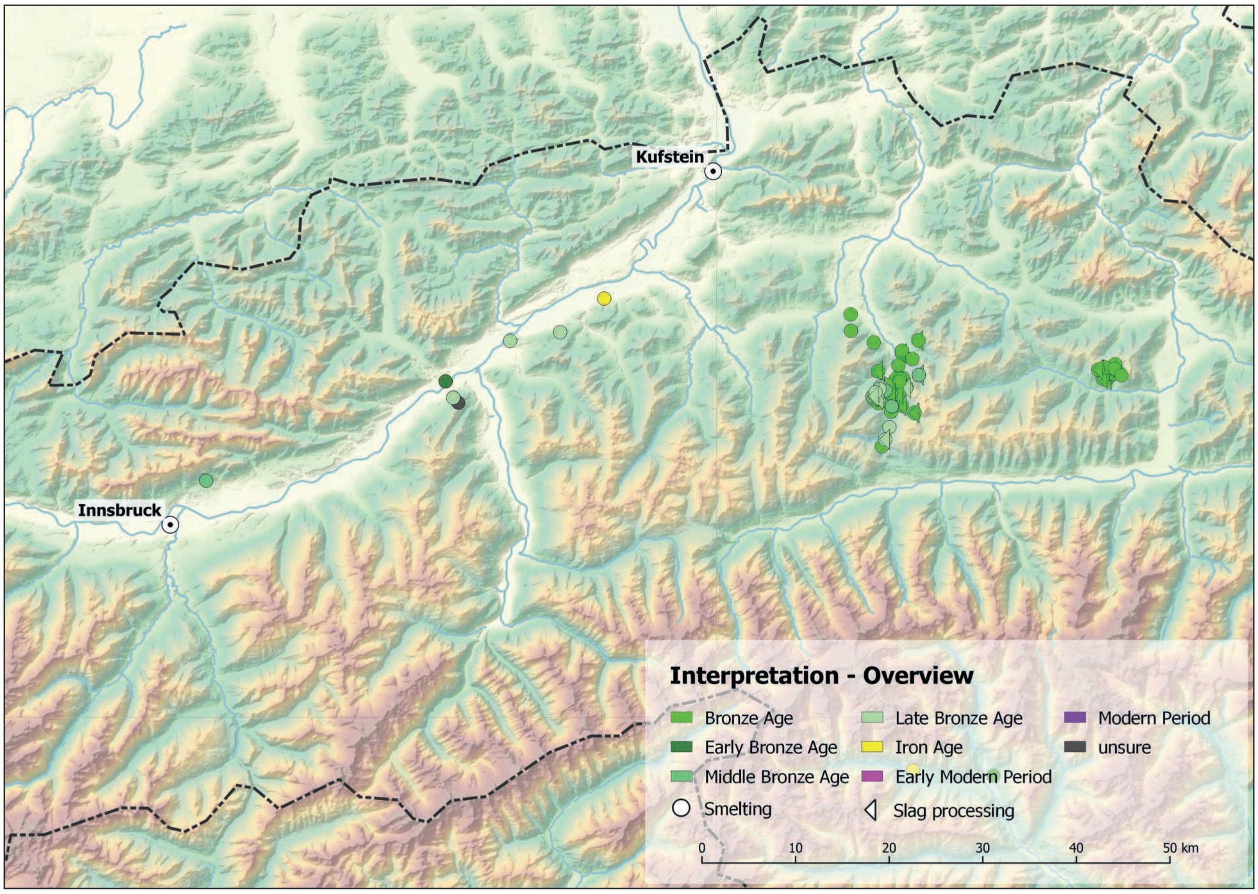

Fig. 9: Map of known archaeological sites in the eastern part of Northern Tyrol and Salzburg with interpretations of smelting and/or slag processing.

the data model. Figure 8 shows how the original data of the documentation, enriched with thesaurus terms is first structured in a tabular format. In a second step the tabular data is mapped to the formal definitions of the CIDOC CRM ontology with Karma. It can be exported in RDF (Resource Description Framework), a data format able to relate logical statements within a network (W3C, 2014). The thesaurus was created with the Karma tool as well and represented in SKOS (Simple Knowledge Organization System), a data model of the semantic web community for sharing and linking knowledge organization systems, such as thesauri, taxonomies, classification schemes and subject heading systems (W3C, 2009). The product is a knowledge graph representing the information of the structured data with the concepts of the CIDOC CRM and the terms of the thesaurus.

After mapping the different information sources and the thesaurus to the common data model the created RDF structure is ingested in a triple store, which is a database to store RDF data. In the triple store the linking of the resources (single information source elements like a specific underground mine or a concept like the Early Bronze Age) takes place performing the actual integration. Resources are either linked on a class level (because they belong to the same CIDOC CRM class, e.g. Observation), on the SKOS concept level (because the same thesaurus term was attributed to them, e.g. "Early Bronze Age") or on an individual level (because they describe the same material structure object or observation, e.g. "Barbarastollen"). Linking on an individual level is also known as coreference or entity matching and may involve additional processes to assess the identity of individuals if no common identifier is available in the different data sources, which is often not the case.

Information provision

The integration of the information and the subsequent provision in adequate tools can help researchers of prehistoric mining archaeology significantly to answer their research questions.

The RDF network of the triple store can be queried using the SPARQL (W3C, 2013) query language. To show the potential of the integration one of the research questions concerning the location and dating of archaeological sites in the eastern part of Northern Tyrol with interpretations of smelting and/or slag processing is illustrated in a map (Fig. 9). Similar maps can be produced showing the physical structures or observations/measurements that lead to these interpretations. To create the map the results of the query were loaded into a Geoinformation system.

Conclusion and outlook

An approach to integrate information related to the field of prehistoric mining archaeology coming from various sources was developed through a common data model and using tools and specifications of the semantic web community to perform the actual integration. With a specific information provision example it was shown that the integration process works and that the triple store can be used to answer specific research questions.

In the current implementation several data sources from geology, surveying and archaeology in the area of Northern Tyrol are integrated. Further research will apply the methodology to sources from geochemical analysis of ore deposits and prehistoric metal artefacts, from ethnoarchaeology and experimental archaeology. The area of investigation will be enhanced to other alpine areas with prehistoric mining research specialized in copper mining. When the documentation of other research groups is targeted the usability of the methodology will be tested further.

Acknowledgements

The research presented here was financed by the Austrian Science Fund in the course of an Erwin Schrödinger scholarship (J 3646-N15) and by the University of Innsbruck in the course of a graduate scholarship (240346).

Bibliography

DARIAH EU, 2016. *DARIAH Backbone Thesaurus (BBT) – Definition of a model for sustainable interoperable thesauri maintenance*, Produced by the Thesaurus Maintenance Working Group, VCC3, DARIAH EU. Available at: <http://83.212.168.219/DariahCrete/sites/default/files/dariah_bbt_v_1.2_draft_v4.pdf> [Accessed 4. January 2017].

Doerr, M., 2006. Semantic Problems of Thesaurus Mapping. *Journal Of Digital Information*, 1(8). Available at: <https://journals.tdl.org/jodi/index.php/jodi/article/view/31/32> [Accessed 3. November 2016]

GBA, 2014. *Digitale Datensätze des Bergbau/Haldenkatasters betreffend ausgewählter Bergbaugebiete im Raum Schwaz-Brixlegg und Kitzbühel-Jochberg*. Fachabteilung Rohstoffgeologie der Geologischen Bundesanstalt.

Hiebel, G., Hanke, K. & Hayek, I., 2013. Methodology for CIDOC CRM Based Data Integration with Spatial Data. In: F. Contreras, M. Farjas, and F. J. Melero, eds. 2010. *Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010 - BAR International.* British Archaeological Reports, pp.547-554.

Hiebel, G. & Hanke, K., 2017. *Digital Workflows for a 3D Semantic Representation of an Ancient Mining Landscape*. The International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences XLII-2/W5, pp.355-359.

ISI, 2016. *Karma: A Data Integration Tool*. Available at: <http://www.isi.edu/integration/karma/> [Accessed 3.November 2016].

CIDOC CRM, 2016. *CIDOC CRM Compatible models & Collaborations*. Available at: <http://www.cidoc-crm.org/collaborations> [Accessed 9. January 2017].

Koch Waldner, T., 2017. *Räumliche und zeitliche Struktur des prähistorischen Bergbaus in der Region Kitzbühel. Charakterisierung einer alpinen bronzezeitlichen Bergbaulandschaft anhand von archäologischen Befunden, Geländeanalysen und überregionalen Vergleichen.* Unveröff. Dissertation. Universität Innsbruck.

Le Boeuf, P., Doerr, M., Ore, C. E. & Stead, S. 2016. *Definition of the CIDOC Conceptual Reference Model.* Available at: <http://www.cidoc-crm.org/official_release_cidoc.html> [Accessed 6. April 2016].

Pirkl, H., 1961. *Geologie des Trias-Streifens und des Schwazer Dolomits südlich des Inn zwischen Schwaz und Wörgl (Tirol)*. Jahrbuch Geol. B. A., Bd. 104, 1. Heft. Wien : Geolog. Bundesanstalt, 1961.

Scherer-Windisch, M.; Brandner, D. & Tiefengraber, G., 2018. KG Viehhofen, OG Viehhofen, PB Zell am See. *Fundberichte aus Österreich* 56 (in press).

Scherer-Windisch, M.; Brandner, D. & Tiefengraber, G., 2019. KG Viehhofen, OG Viehhofen, PB Zell am See. *Fundberichte aus Österreich* 57 (in press).

W3C, 2009. *SKOS Simple Knowledge Organization System Reference.* Available at: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/> [Accessed 19. June 2016].

W3C, 2013. *SPARQL 1.1 Overview*: Available at: <https://www.w3.org/TR/sparql11-overview/> [Accessed 9. January 2017].

W3C, 2014. *Resource Description Framework (RDF).* Available at: <http://www.w3.org/RDF/> [Accessed 19. June 2016].

Authors

Gerald Hiebel – Unit for Surveying and Geoinformation, University of Innsbruck
Gert Goldenberg – Department of Archaeologies, University of Innsbruck
Caroline Grutsch – Department of Archaeologies, University of Innsbruck
Klaus Hanke – Unit for Surveying and Geoinformation, University of Innsbruck
Markus Staudt – Department of Archaeologies, University of Innsbruck

Correspondence and material requests should be addressed to: gerald.hiebel@uibk.ac.at