Transdisciplinary approach to archaeological investigations in a Semantic Web perspective

Vincenzo Lombardo a,*, Tugce Karatas a, Monica Gulmini b, Laura Guidorzi c,e, and Debora Angelici d

a Dipartimento di Informatica, Università di Torino, corso Svizzera 185, 10149 Torino, Italy
E-mails: vincenzo.lombardo@unito.it, tugce.karatas@unito.it
b Dipartimento di Chimica, Università di Torino, via Pietro Giuria 7, 10125 Torino, Italy
E-mail: monica.gulmini@unito.it
c Dipartimento di Fisica, Università di Torino, via Pietro Giuria 1, 10125 Torino, Italy
E-mail: laura.guidorzi@unito.it
d TecnArt S.r.l., via Modena 58, 10153 Torino, Italy
E-mail: angelici@tecnart.eu
e INFN Sezione di Torino, via Pietro Giuria 1, 10125 Torino, Italy
E-mail: laura.guidorzi@unito.it

Editors: First Editor, University or Company name, Country; Second Editor, University or Company name, Country
Solicited reviews: First Solicited Reviewer, University or Company name, Country; Second Solicited Reviewer, University or Company name, Country
Open reviews: First Open Reviewer, University or Company name, Country; Second Open Reviewer, University or Company name, Country

Abstract. In recent years, the transdisciplinarity of archaeological studies has greatly increased because of the mature interactions between archaeologists and scientists from different disciplines (called “archaeometers”), where a number of diverse scientific disciplines collaborate to get an objective account of the archaeologic records. A large amount of digital data support the whole process, and there is a high value of keeping the coherence of information and knowledge, as contributed by each intervening discipline. During the years a number of representation models have been developed to account for the recording of the archaeological process in data bases and lately, some semantic model, compliant with the CRMarchaeo reference model, has been developed to account for linking the institutional forms with the formal knowledge concerning the archaeological excavations and the related findings. On the contrary, the archaeometric processes have not been addressed yet in the Semantic Web community and only an upper reference model, called CRMsci, accounts for the representation of the scientific investigations in general. This paper presents a modular computational ontology for the interlinked representation of all the facts related to archaeological and archaeometric analyses and interpretations. The computational ontology is compliant with CIDOC-CRM reference models CRMarchaeo and CRMsci and introduces a number of novel properties and classes to link the two worlds in a joint representation. The ontology is in use in the BeArchaeo, which is a methodological project for the establishing of a transdisciplinary approach to archaeology and archaeometric disciplines, interlinked through a semantic model of processes and objects.

Keywords: Archaeology, CRMarchaeo model, keyword three, keyword four, keyword five

1. Introduction

Archaeological investigations have been relying more and more on reflexive methodologies [1]. Nowadays, sense making within archaeological investigations starts its journey in the excavation site and continues up to museum curatorial practices and is accompanied by labels in exhibitions and records in digital repositories and collections. In fact, though interpretations still rely upon the expertise of the excavation
team [2], the trend is to carry reflexivity to the extreme consequences through the video recordings of initial sense making during the excavation and producing daily reports by using web-based interfaces, up to filling the data base entries for the excavation. This documentation, which can also be accessed later, reveals much of the background to the interpretations and the audiences, as well as other scientists, can query the data and evaluate conclusions.

The other methodological process that characterizes the current conduction of an archaeological investigation is the contribution of archaeometry, acknowledged by many archaeologists as an indispensable and integral part of archaeology. Archaeometry designates the development and application of natural scientific methods and concepts to the solution of cultural-historical questions. Although applications of natural sciences in archaeology have actually a long tradition (cf., e.g., “the quantitative analysis of Roman coins in 1799 by Martin Heinrich Klaproth in Berlin”), archaeometry is both archaeology by ultimate aim, but natural science by approach, including all disciplines that may contribute to archaeology (physics, chemistry, biological sciences, anthropology, geological sciences, ...), by measuring and evaluating facts and interpretations [3] [4].

However, as archaeology, with the growing contribution of archaeometry, becomes fragmented into specialized areas of knowledge, challenges to integrated interpretation and contextual understanding through the archaeological investigation increase. While, on the one hand, the individual archaeologist interfaces with the recording structure, which supports access to reflection and dialogue with all the members of the project, on the other, the challenge is to realize a holistic view of the data, with interpretations about findings, stratigraphic units, or whole archaeological sites to be developed in broad contexts, satisfying historical and natural scientific constraints [5] [6]. Researchers of the pioneering Çatalhöyük project, implementing digital integration of data from the field excavation, have acknowledged problems derived from “faultlines between field and laboratory staff or from the practical separation of ever more complex forms and types of data” [7], but the adoption of digital technologies and methods in the field (such as GIS and 3D visualization on tablets) has led to a maturing and expansion of the reflexive objectives.

In a number of cultural heritage areas, digital data curation has emerged as a viable workflow for the management of the related digital assets during their entire lifecycle [8]. It consists in “actively managing data [...] with the aim of supporting reproducibility of results, reuse of and adding value to that data, managing it from its point of creation until it is determined not to be useful, and ensuring its long-term accessibility and preservation, authenticity and integrity” (Digital Curation Center - DCC1). In archaeological investigations, the digital assets can be more or less formal descriptions of artifacts and the excavation context (stratigraphic units and preliminary interpretations) curated by archaeologists or measurements of some physical parameter that reveal some hidden property resulting from some archaeometric investigation [9]. Data recording sheets for findings and surrounding contexts enable the recording of excavation outcomes in archaeological databases; however, the interpretation (e.g., the classification of some artifact or the estimation of some chronology) proceeds in incremental phases and, also given the contribution of archaeometric methods, can be subject to revisions. This can lead to the implementation of a truly transdisciplinary endeavor, where research questions arise through the collaboration and peer-to-peer cross-fertilization of several disciplines [10]. Archaeology and archaeometry also engage philosophy, social sciences, and activities on the field (including the negotiation with contractors and the decision makers), and the reconstruction of the past co-exists with “articulating activist political positions in the present” [11]. At the same time, datasets are increasingly available online: projects such as the Digital Archaeological Record2, the catalogue section of the Central Institute of Cataloguing and Documentation of the Italian Ministry of Cultural Heritage3, and the Archaeology Data Service4 make available a number of archeological data for quantitative testing and processing and these data are reused by other researchers in novel ways (see, e.g., [12]). However, datasets are actually isolated from each other; some researcher also reports no connection to grey literature (the so-called unpublished excavation reports), and there is a demand on semantic interoperability between differing database structures and terminology [13]. Semantic interoperability is also called to overcome some of the limits that have been raised for IT applications in archaeology, which, while appointed to bring some data-driven theory-neutrality to archaeo-

---

1http://www.dcc.ac.uk, visited on 3 March 2021.
4http://archaeologydataservice.ac.uk/, visited on 3 March 2021.
logical investigations, have been appraised as "unrealized 'great expectation'" [14].

In this scenario, the Semantic Web approach has been invoked to support the sharing of data, particularly for the transdisciplinary endeavors [15]. This paper presents a conceptual model and ontology for supporting the transdisciplinary vision of the archaeological investigations, at the crossroad of many archaeometric discipline, contributing to its reflexive methodology in the context of an encompassing digital curation of the data. In recent work, we have proposed an ontology-based approach for the encoding of the semantic knowledge underlying the archaeological forms to be filled for the documentation of the excavation and the interpretation phases [16], related to ongoing EU project BeArchaeo, which consists in an archaeological excavation, the consequent interdisciplinary archaeometric analyses of the site and the excavated materials, the interpretation of the findings, and the dissemination of the results through physical and virtual exhibitions. Here we address the overall ontological approach, which specializes the CRMarchaeo model and the CRMsci model, of the CIDOC-CRM family.

The paper is organized as follows. In the next section, we report on the related work about the digital approach to archaeological data, with particular reference to their semantic organization. Then, we introduce the general context of the digital data curation and BeArchaeo, a DDC-born archaeological project. The core of the paper is the description of a comprehensive approach to the conceptualization of the archaeological and archaeometric domains, at the base of a transdisciplinary approach to archaeological investigations. Running examples are taken from the BeArchaeo project, carried on with a semantic organization of data in support of the coordination of all the tasks, from the excavation planning to the final exhibition of the results.

2. Related work

Archaeological projects go digital in all their phases: data collection, curation, and visualization (see, e.g. [17, 18], among others), analysis (e.g., GIS [19]), exhibition (starting from the virtual archaeological reconstructions of the 1990s [20, 21] and addressing general public outreach and participation [22]). A particular mention goes to the pioneering Çatalhöyük project, carried out with the goal of maintaining the data as long as possible. The Çatalhöyük Database and the Çatalhöyük Image Collection Database make available the documentation of the Çatalhöyük excavation site. Custom platforms allow for the search of data uploaded during every excavation season and made available through the Çatalhöyük Living Archive, which tells about two decades of excavation and analysis at a Neolithic settlement in Turkey. However, even across project within single institutions, the global picture is a "rather disparate grouping, or 'archipelago', of diverse, specialised, but rather isolated and independent information systems and databases" [23]; limits concern sharing and standardization of data [24]. Also a survey made within the AriadnePlus project reports that researchers are not very aware of the issues of data sharing and Linked Data. Linked Open Data are also advocated to encourage the dissemination and the linking of archaeological datasets [25]. The motto "data sharing as publication" promotes an initiative to publish data and resources from archaeology after review by an editorial board and integrate data through some (simple) ontological model. Integration and sharing of data through the instantiation of acknowledged ontologies support the major challenge archaeologists have to face, namely data reuse [26]. The authors get to promote a general "data literacy" for archaeologists, who should care personally for their own data, through direct management and communication [27].

There have been some semantic approaches, especially in the context of the reflexive methodologies, hence requiring some knowledge to interconnect objects, events, and people, historical context and excavation process [28]. CIDOC-CRM ontology has been employed to deal with interpretations as events that occur from the excavation process and can occur later again, when initial interpretations are revised or integrated, in the context of the long running Çatalhöyük...
project [29]. In this case, CIDOC-CRM worked as the backbone for a digital counterpart of a more conventional print report, emphasizing the need for time-consuming data cleansing with typical archaeological datasets. One of the most relevant takeaways of the analysis was the need for a publishing platform, where the complex and massive content could be inserted and accessed through user-friendly interfaces.

An indirect use of CIDOC-CRM data model is through the Arches platform [30], on which a number of projects are based: for example, the two projects, namely EAMENA (Endangered Archaeology in the Middle East and North Africa)11 and ASOR (American Schools of Oriental Research) Cultural Heritage Initiatives for Syrian and Iraq12, which record archaeological sites and landscapes that are under threat or damaged across the Middle East and North Africa, with goals of documentation, sharing information, and planning responses. Arches manages six resource types: heritage resources (such as archaeological sites or buildings), heritage resource groups (e.g. urban districts), actors (e.g. persons or organizations), historical events (e.g. floods or epidemics), activities (e.g. investigations), and information resources (e.g. media files). The data model of Arches builds on CIDOC-CRM and other interoperability standards, such as the Open Geospatial Consortium (OGC) for the geospatial data, and includes modules for vocabulary management, such as Getty Art and Architecture Thesaurus13.

3. Digital data curation and the BeArchaeo project

Digital data curation consists in the coordination of the representation and management of the digital assets related to cultural heritage, i.e. tasks as selection, processing, preservation, maintenance, collection, and archiving of the digital, with possible value adding for subsequent exploitation [8]. The notion of digital data curation has been revised and updated several times, with a recent focus on motivations and big data [31].

To systematize goals and practices of digital data curation, a number of models have appeared in the literature from many institutions, such as, e.g., Digital Curation Center Lifecycle Model [32] and I2S2 Idealized Scientific Research Activity Lifecycle Model [33]. Here we describe the digital data curation through an abstract representation of the tasks, adapted from [9].

The Digital Data Curation model consists of six common tasks (in blue circles Fig. 1) for the data management cycle from the cultural heritage asset to the final outputs of the digital curation process. Each task is exemplified with tools and components (bordered by dotted lines in the figure). In the archaeological case, the cultural heritage (CH) item can be an archaeological finding (including fragments), a stratigraphic unit, the whole archaeological site. Conceptualization provides a knowledge framework to define the model for the digital data that are produced during the project implementation. The conceptualization depends on both the cultural heritage domain (an archaeological site, a specific finding, ...) and the goal intended to be achieved by the project (the chronology of the finding, the material provenance, ...). The conceptualization provides the backbone for the database schema design that will account for the description and encoding of the digital data produced by the project. Digital data curation starts with the data creation or acquisition by focusing on what data are acquired, how, and why. Data acquisition brings data that have been created by a source outside some organization into the organization, for production use. This means that a number of activities, supported by tools, must be carried out, namely identifying, sourcing, understanding, assessing, and ingesting raw data. Instead, data creation is the process that samples signals that measure real world physical conditions and converts the resulting samples into digital numeric values. The data processing and modeling phase focuses on creating a conceptual model for the data to be stored in a database, together with the associations between different data objects and the rules (many projects employ E-R Model and UML format) in order to support effective exchange of knowledge and interoperability. Data interpretation is the process of making sense of data that has been collected, analyzed, and presented. All the members of the project can access a holistic overview of the data and the interpretations about individual items, sets of items, or higher-order categories. The task of data curation and preservation records all the data and metadata created during these three phases. The semantic relations between artifacts and their constituent parts is crucial in this step as well as aspects regarding authorization, persistent identification, data curation and long-term archiving. The data documentation and archiving process manages the
metadata or information about some data product (e.g., data table, database) that enables one to understand and use the data. For example, a database can be classified by the type of content included in it (e.g., bibliographic, statistical, document-text) or by its application area (e.g., biological, geological, etc). On the other hand, data dissemination and publishing is the distribution or transmitting of statistical, or other, data to end-users in order to make data available (i.e. electronic format or paper publications such as PDF files based on aggregated data as well as the exhibitions and websites of the collection of the cultural heritage organizations).

Now we illustrate the model with an example of digital curation of data which is related to some digital data generated from an archaeological finding as recorded during an archaeological excavation and analyzed afterwards in order to track the activities carried out. BeArchaeo is a DDC-born archaeological project, consisting of the archaeological excavation and archaeometric analysis of the Tobotsuka Kofun located in Soja city in Okayama Prefecture of Japan. Together with other Kofun burial mounds and the related archaeological material in ancient Kibi and Izumo areas, researchers aim to develop a transdisciplinary vision in studying the archaeological site; the project activities and outcomes are accessible to the general public through engaging media communication along the project development. In this section we will examine the proposed digital curation operational framework for ongoing activities of the archaeological discoveries, scientific interpretations and the related database.

Fig. 2 instantiates the general model above on one operational workflow addressing the digital data originated since the discovery of the archeological finding named SH1 and addressing a specific investigation path, at the current stage of development. As we have seen above, interpretations are recorded in some digital format and then revised or updated, also encoding other formats, going formally when possible.

The conceptualization of the knowledge in the BeArchaeo project is driven by the design principle of recording the archaeological/archaeometric activities and the collected data that occur both on the archaeological site and in the lab. The data are recorded in a database filled by the scientists in order to be employed in interpretation processes and exhibition organization. The goal of the digital data curation is to support the scientific research on the composition of the findings and to examine their relation with the question of their similarities and differences. In this specific example, the research question is to find the provenance of a set of similar potteries through comparison of the component materials, including elemental composition, morphological features, presence, typology and composition of inclusions such as minerals or rock fragments. The fragment SH1 is an archeological finding fragment. The digital curation workflow starts as soon as the fragment SH1 has been found. In particular, the figure addresses a measurement carried out in the lab, where scientists analyze the structure of the fragment through a process of data acquisition. Images of the cross-section of the fragment are analyzed by...
means of a SEM-BSE (Scanning Electron Microscope-backscattered electrons). This process generates raw data (a magnification is shown in the figure, jpeg file format). The task of data modeling and processing enriches raw data with metadata that reveal an interpretation of the asset at some level (e.g., measure size of the fragment or assign label "clay"). Elemental maps of a portion of the sample, which are visible in the figure, highlight that the coating is depleted in Al$_2$O$_3$; later, it may suggest an enrichment in iron compounds, which would indicate that a coating was present. The outcome of this step, namely the processed data, are the numeric values concerning the elemental composition determined after a set of analyses, enriched with metadata. For example, according to scientists, it is evident that the surface is made of a flaking brown layer (and a conservation treatment with Binder 17 - acrylic acid ester copolymer has been applied in the past), which may represent the residual of a surface coating. Also, elemental maps of a portion of the sample are added as metadata (actually, elemental maps have been acquired before, within the process of Energy Dispersive X-Ray Analysis (EDX) together with results from SEM analysis, but logically belong to this phase). Together with metadata-enriched processed data, these data undergo the data interpretation process, becoming a proper model of the cultural heritage domain.

Based on the findings, the resulting statement is that the layer is formed by very thin and purified clay bearing oxidized iron compounds. In general, such statements may include the scientific context underlying the data as well as the additional questions and interpretations. As the morphology of the fragment would arise more questions, some further samples of the same typology in a better conservation state would clarify this point.

The interpretation shows that the fragment can be a part of the Kibi area, because of the cluster that tentatively set together the fragments SH1, SH3 and SH4, although it seems rather scattered. So, the current conclusion is that SH1 can originate in the Kibi area (interpretations reported through powerpoint slides). The intermediate and the final data are stored into the repository, namely a Google drive shared folder, through the tasks of Data curation and preservation. Moreover, the interpretations in the format of powerpoint slides are also selected and stored, as part of the Data documentation and archiving task, into the BeArchaeo Archive, namely a MySQL database, underlying an Omeka-S installation, which also works as centralized database for the coordination of digital data curation. The model will also be enriched with further metadata (e.g., the digital image also receives the identifier of the physical fragment). The database schema design as well as the
organization of the google drive folders are based on
the proposed semantic model worked out after the con-
ceptualization phase, to ease the problems of interoper-
ability and connection between the archeological and
the archaeometric data (see below). Finally, in order
to make the knowledge available to the archaeologists
on the field, a BeArchaeo project website, based on
the mentioned installation of the Content Management
System (CMS) Omeka-S, is available. The recording
of the archaeological findings and forms as templates
are made possible through a web-publishing platform
that allows for the import of semantic properties de-
\[...\]

4. Transdisciplinary conceptualization of the
archaeological/archaeometric investigations

Given the digital data curation schema above, which
involves a conceptualization addressing several disci-
plines, we developed an ontology, called BeArchaeo
ontology, with the design principle to capture the con-
nections between the archeological and the archaeo-
metric realms, respectively. Transdisciplinarity is me-
diated by the formal ontology, with research questions
arising from the collaboration between the disciplines
[11]. The BeArchaeo ontology pivots on the descrip-
tion of the objects, and merges the general classes and
properties belonging to the archaeological and the ar-
cheometric realms with the fields of the archaeologi-
cal catalogue records [16]. Design patterns, for con-
necting these knowledge domains, are not available (to
the best of our knowledge). The result is an application
ontology that merges three types of knowledge: the ar-
cheological knowledge (lower left part of Figure 3),
the archaeometric knowledge (lower right part of Fig-
ure 3), and the catalogue record knowledge (upper part
of Figure 3).
Fig. 3. Modeling of the archaeological finding “AF 99”, exemplifying archaeological and archaeometric knowledge, respectively, and the corresponding fields in the archaeological finding record. The rectangles in grey or black are the individuals; the white rectangles are the classes; object properties are depicted as blue lines, while datatype properties are depicted as green lines; the three elements in Courier font are the strings that are actually written in the final form interface.

Fig. 4. Major relationships between BeArchaeo and CIDOC-CRM family. Colors are employed to distinguish provenances.
Archaeological findings, as physical things, can be the object of a task CRMsci/S19 Encounter event (an archaeologist encounters a finding in a stratigraphic unit). Physical things are a subclass of observable entities (class CRMsci/S15 Observable Entity), which can be observed (specifically measured), producing values (any cidoc-crm/E1 CRM entity) for some property type (class CRMsci/S9 Property Type). The data collected can be evaluated (class CRMsci/S6 Data Evaluation) for the assignment of some dimension (property CRMsci/O10 assigned dimension) to the archaeological finding (check the description of the digital data curation for the example SH1 above).

As an example (see Figure 5), illustrating an example from the current documentation of CRMsci\(^\text{15}\), we look at the Temple of Hercules in Amman\(^\text{16}\). The overall height (low, right, type cidoc-crm/E54 Dimension) of the statue of Hercules (low, center, type CRMsci/S15 Observable Entity) in the Temple of Hercules in Amman has been assigned (CRMsci/O12 has dimension) through a data evaluation activity (low, left, CRMsci/S6 Data Evaluation) from the measurement (middle, left, CRMsci/S21 Measurement) of the size (top, center, CRMsci/S9 Property Type) of a fragment of a finger (middle, center, type CRMsci/S15 Observable Entity).

5. The BeArchaeo ontology

The conceptualization described above has been enriched with specialized vocabularies for supporting the digital data curation process of an archaeological investigation. Here we distinguish between the objects and the processes, each category yielding its digital data to be curated.

In the figures 6 and 7 there are the major relations of the stratigraphic unit class and the archaeological finding class, respectively. Going clockwise, a stratigraphic unit has inclusions (i.e., entities that are contained in the stratum), which are of some type, that can be generic or specific, and has a frequency of occurrence in the unit, qualitatively valued as rare, medium, or frequent. Inclusions have types that are taken from partially overlapping vocabularies, based on the practical experience of the archaeologists (these may change and should be aligned with the types included in the thesauri for the archaeological findings). Some informal properties, noted as free text, are the state of preservation of the unit and the measurements


Fig. 5. The calculation of the overall height (ES4) of the statue of Hercules (S15) in the Temple of Hercules in Amman from the measurement of the size of the fragment of the finger.

The archaeological finding (Figure 7) can be part of another archaeological finding (frequent is the case of fragments to be composed afterwards) and is sourced by some stratigraphic unit as well as museum collection or other places. This variety of sources concerns the goals of the BeArchaeo project, because of the employment of the ontology into the design of the figure. Each monodisciplinary team has thus deeply reflected on their own procedures, data formats, and knowledge contributions. After that, the broad group of researchers have discussed the links that could have been set among the diverse monodisciplinary outcomes, in order to enhance the overall knowledge in a transdisciplinary perspective. So, they carefully selected the entities supporting the inferential processes from data, in order to include them into the conceptual model. Finally, they tackled the challenge of conceptual modelling according to a common formal structure based on CIDOC-CRM and CRMsci.

17Munsell color system is based on the three-dimensional model, where each color is defined by a triple of hue (the color of the color), value (how light or dark is the color), and chroma (or saturation/brilliance of the color), set up as a numerical scale with visually uniform steps, visited on 3 March 2021.

The specialization of the CRMsci reference model through the processes and the digital data produced by the various disciplines proceeded by engaging the disciplinary researchers, who have been asked to speculate on the procedures and results concerning the archaeological sites or findings, in order to single out the relevant concepts that are related to their disciplinary contribution to the overall investigation. Each monodisciplinary team has thus deeply reflected on their own procedures, data formats, and knowledge contributions. After that, the broad group of researchers have discussed the links that could have been set among the diverse monodisciplinary outcomes, in order to enhance the overall knowledge in a transdisciplinary perspective. So, they carefully selected the entities supporting the inferential processes from data, in order to include them into the conceptual model. Finally, they tackled the challenge of conceptual modelling according to a common formal structure based on CIDOC-CRM and CRMsci.
Fig. 6. Conceptual model of the stratigraphic unit knowledge (including references to thesauri and vocabularies (with list of terms)).

Fig. 7. Conceptual model of the archaeological finding.
Conceptualization, after an initial account based on the investigation techniques, has shifted toward material-related processes (such as glass-, pottery-, stone-... related). Then, we produced a number of intra- and inter-disciplinary schemata, to be merged after further fruitful discussions, also highlighting the fields to be reported in the filling forms. The approach has made possible the emergence of the common features of the various procedures/techniques and supported the avoidance of redundancies while enabling the inter-disciplinary connections of the diverse knowledge forms. About archaeological findings, scientific data include significant details related to pretreatments and instrumental parameters; these data are normally neglected in databases developed according to an archaeological account, as they appear useless for non-archaeometric professionals. However, since they are crucial for stunning data with scientific significance, we have included all the procedures that, to the best of the present knowledge, have significance for the archaeometric investigations on the archaeological findings and on the archaeological site. This has considered pottery, glass and organic remains, and has included the encoding of the scientific testing procedures and of 2D/3D imaging techniques (non-invasive and invasive procedures, included sampling): namely, elemental, molecular, structural and magnetic analyses (both qualitative and quantitative), micro-morphological imaging (2D and 3D), dating.

Given the principles outlined above, the conceptual model for the archaeometric investigation identifies the entities that encode the data recorded by the scientists during their investigations. These entities have been then encoded, while keeping a link to the archaeological domain. Again, though the model is strongly interconnected, for the sake of exposition, we keep a distinction between the archaeological finding domain and the archaeological site/stratigraphic unit domain. The outcome of this work has been a set of schemata that extend CRMsci model, where all the samples obtained from the findings and the soil have become individual entities of the model and undergo specific observations (depending on disciplines), while the data obtained from observations undergo some specific data evaluation and inference making. In particular, following the tenets of CRMsci ontology, details on scientific methods and devices are best handled in each scheme as free text (cidoc-crm/P3 has note), whereas basic scientific techniques (such as "carbon 14 dating") are encoded using the type property (cidoc-crm/P2 has type), by referring to a thesaurus under development.

In the remainder of this section, we address the schema realized for modeling archaeological pottery investigation by means of morphological qualitative methods (Figure 8), in particular polarized light microscopy. The conceptual model is based on the annotation structure suggested by Quinn[36] for the investigation of pottery prepared as thin sections. The trans-disciplinary value of the conceptualization is that the scheme has been adjusted to match the investigations carried out by the many disciplines involved in the archaeometric investigation of pottery findings (see below on data evaluation). In particular, the conceptualization emerging similarities spanning the diverse disciplinary procedures, by replicating the same major structure for the thin sections in pottery. It models, specifically 1) the investigation in cross section of pottery, 2) the determination of qualitative chemical composition by XRF glass and pottery, 3) the investigation of inclusions by scanning electron microscope in glass, 4) the spectroscopic investigation of glass through Diffuse Reflectance Spectroscopy.

The analysis by optical microscope of archaeological ceramics in thin section reveals the complexity of these materials (Figure 8). They are composed of three main components (inclusions, matrix and voids), each one investigated by a section of main process (classes bearchaeo/Measurement PLM Inclusions, bearchaeo/Measurement PLM Matrix, and bearchaeo/Measurement PLM Voids) Ontology BeArcheo architectural know-how on pottery thin sections are analyzed by means of optical microscope under polarized light results in attribute values along some dimensions (e.g., relative abundance and sizes of inclusions) and terms from specialistic vocabularies (e.g., grain size distribution, valued as unimodal, bimodal, or heterogeneous, or mineral/petrographic component, with subsets such as quartz presence or alkali feldspars presence, valued as XXXX, i.e. > 50%, XXX, i.e. 50-30%, XX, i.e. 30-10%, X, i.e. <10%, D, i.e. detectable).

Following the same approach exemplified for the polarized light microscope, the overall effort on archaeological findings has defined the guidelines for the procedures listed below, namely Sampling, Sample pre-treatment for thin sections or cross sections, Structural analyses by X-ray powder diffraction, Qualitative and Quantitative chemical analyses, Thermoluminescence dating for pottery, Archaeomagnetic tests for pottery, 2D/3D imaging including X-ray Radiography, X-ray Tomography, Scanning electron microscopy, EDX elemental mappings, Photometry, Radio-carbon dating through accelerator mass spectrome-
try (AMS) for organic remains. Finally, the conceptual model for measurements has been complemented by the conceptualization about data evaluation (class CRMsci/S6 Data Evaluation), which takes into account the data produced through measurements to provide a suitable interpretation. In the instantiated model reported in Figure 9, the results from termoluminescence, archaeomagnetism, X-ray powder diffraction and Scanning electron microscopy (on the left) are combined to infer the firing temperature of a pottery shard (namely, the sample No. 7 from Tatetsuki area). In particular, the numerical value obtained from archaeomagnetism analyses can be confirmed by the observations of other parameters (i.e. moisture content at saturation, presence/absence of calcite, porosity and sintering degree of body paste), which independently suggest specific temperature ranges that are congruent with the equivalent firing temperature estimated by archaeomagnetism in particular.

From a technical point of view, the model has been described as a number of subontologies concerning the stratigraphic unit (the correspondent CRMarchaeo class) and the archaeological finding record (subclass of cidoc-crm/E18 Physical Thing), respectively, and the archaeometric processes as subclasses of the CRM-sci reference model. The ontology is expressed in OWL/RDF formats and published at the permanent address /purl.org/beArchaeo18.

6. Conclusion

We have presented a transdisciplinary ontology-based approach to the encoding of archaeological and archaeometric knowledge. In particular, we have setup a procedure for addressing the transdisciplinary endeavor and we developed a prototype ontology of the interconnected archaeological and the archaeometric domains, respectively. These issues are particularly relevant for the digital data curation of an archaeological investigation; we have also designed both a practical workflow and the form interfaces for collecting the data as the excavation goes on, to be continued in the analysis labs, and eventually with the design of the exhibition. We have identified the major entities that are required for a reflexive methodology of archaeology, especially in its relationship with archaeometric knowledge. The conceptual model has been the outcome of several discussions carried out by the members of the archaeometric team, representing the several18File “BeArchaeo_merge_all.owl” merges all the other sub-ontologies.
Fig. 9. Evaluation of data for the assignment of a dimension.

eral disciplines involved. The conceptualization has been developed in support of a digital data curation framework that serves the needs of an ongoing archaeological investigation.

The encoding of the archaeological knowledge in an ontology that is compliant with CIDOC-CRM leads the project database to adhere to the Semantic Web paradigm and address data sharing and interoperability. The archaeological knowledge ontology is the base for a CMS-based web platform, in support of the archaeologists’ work in recording the excavation and interpretation activities. The ontology and the availability of such a widespread CMS can be easily replicated in further projects. The conceptual model and the ontology of the archaeometric knowledge serve the design/implementation of the interface forms for both archaeological and archaeometric filling, in order to enable researchers operating on the field and afterwards in the labs to load their results into the database. As far as we know, BeArchaeo is the first archaeological project that assumes a Semantic Web approach from the start. In fact, the multi-disciplinary, multicultural, and multi-lingual characters of Be-Archaeo raise a high demand of interoperability of knowledge and data. The alignment with CIDOC-CRM is pursued at the disciplinary level, by aligning the archaeologi-cal and the archaeometric descriptions through the CRM-Marchaeo and CRMsci models, where possible.

In the next future, we will provide a thorough encoding of all the aspects related to archaeometry and the strict connection with the archaeological interpretations, in terms of automatic reasoning on the data collection. As the project database will be growing in the collection of data, we are going to improve the interfaces for engaging a higher number of diverse researchers and promote the usage of the conceptual model in other archaeological/archaeometric projects. Finally, we are going to evaluate the contribution of the centralized semantics-enhanced digital data curation in its impact onto the final exhibition.

7. Author statement and acknowledgements

All authors worked on the paper topics and revised the paper. Vincenzo Lombardo carried out the design and implementation of the ontology and wrote the core sections of this paper. Tugce Karatas worked on the project digital data curation model. Monica Gulmini, Laura Guidorzi, and Debora Angelici worked on the conceptualization of the archaeological knowledge.

The BeArchaeo project is funded by the European Union’s Horizon 2020 research and innovation pro-
programme under the Marie Skłodowska-Curie, Grant Agreement No 823826.

We also thank Claudio Mattutino for the maintenance of the Omeka-S installation and Carmine Montefusco and Angelo Saccà for the UniTo hosting service of BeArchaeo database.

References


