Transdisciplinary approach to archaeological investigations in a Semantic Web perspective

DOI:10.3233/SW-223016

(Article begins on next page)

Semantic Web 1 (0) 1–4
IOS Press

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"). A number of diverse scientific disciplines collaborate to get an objective account of the archaeological records. A large amount of digital data support the whole process, and there is a great value in 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. 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 the archaeological and archaeometric analyses and interpretations, also connected to the recording catalogues. The computational ontology is compliant with CIDOC-CRM reference models CRMarchaeo and CRMsci and introduces a number of novel classes and properties to merge the two worlds in a joint representation. The ontology is in use in "Beyond Archaeology", a methodological project for the establishing of a transdisciplinary approach to archaeology and archaeometry, interlinked through a semantic model of processes and objects.

Keywords: Archaeology, CRMarchaeo model, Archaeometry modeling, BeArchaeo project

1. Introduction

Archaeological investigations have been relying more and more on reflexive methodologies [1]. Nowadays, making sense of archaeological investigations

*Corresponding author. E-mail: vincenzo.lombardo@unito.it.

starts its journey in the excavation site and continues up to museum curatorial practices, accompanied by labels in exhibitions and records in digital repositories and archives. In fact, though interpretations still rely upon the expertise of the excavation team [2], the trend is to carry reflexivity to its extreme 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. The audiences, as well as other scientists, can query the data and evaluate conclusions.

The other methodological issue that characterizes the current conduction of an archaeological investigation is the contribution of archaeometry, acknowledged by many archaeologists as an essential and integral part of archaeology. Archaeometry involves 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 (e.g., "the quantitative analysis of Roman coins in 1799 by Martin Heinrich Klaproth in Berlin"), archaeometry is archaeology by ultimate aim, but natural science by approach. It includes all the disciplines that may contribute to archaeology (e.g., 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 achieve an integrated interpretation increase. The individual archaeologist interfaces with the recording structure, which supports access to reflection and dialogue with all the members of the project; additionally, the challenge is to realize a holistic view of the data, with interpretations about findings, stratigraphic units, or sites to be developed in broad contexts, satisfying historical and natural scientific constraints [5, 6]. Although problems derived from "faultlines between field and laboratory staff or from the practical separation of ever more complex forms and types of data" [7] have been acknowledged in digital integration, 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 (or DDC) has emerged as a viable workflow for the management of the related digital assets during their entire lifecycle [8]. It consists of "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 - DCC[1]). In archaeological investigations, the digital assets can be more or less formal descriptions of artifacts and of the excavation context (stratigraphic units and preliminary interpretations), curated by archaeologists, or measurements of some physical parameters that reveal some hidden property, resulting from some archaeometric investigation [9]. Data recording sheets 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. The research goes through a truly transdisciplinary endeavor, where research questions arise through the collaboration and peer-to-peer cross-fertilization of several disciplines [10]. At the same time, datasets are increasingly available online: projects such as, e.g., the Digital Archaeological Record[2], the catalogue section of the Central Institute of Cataloguing and Documentation of the Italian Ministry of Cultural Heritage[3], and the Archaeology Data Service[4] make a number of archeological data available for quantitative testing and processing, and these data are reused by other researchers in novel ways (see, e.g., [11]).
However, most datasets are actually isolated from one another; 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 [12]. 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 archaeological investigations, have been appraised as "unrealized 'great expectation' " [13].

In this scenario, the Semantic Web approach has been invoked to support the sharing of data, particularly for the transdisciplinary endeavors [14]. In recent years, some projects have provided access to col-

[1] http://www.dcc.ac.uk, visited on 15 January 2022.
[2] http://www.tdar.org/, visited on 15 January 2022.
[3] http://www.iccd.beniculturali.it, visited on 15 January 2022.
[4] http://archaeologydataservice.ac.uk/, visited on 15 January 2022.

lections of archaeological data through the integration of knowledge organization systems/services (KOSs)[5], conceptual frameworks such as the Dublin Core Metadata Initiative (DCMI)[6], the CIDOC-CRM conceptual reference model[7]. Project ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) relies on these ontological tools and models to enable the sharing and re-use of about two million archaeological datasets[8].

However, according to our knowledge, the representation of the archaeometric processes as well as a modern and transdisciplinary conception of the archaeological endeavor at large have not found their way through the Semantic Web endeavors. This paper presents a conceptual model and ontology for supporting this transdisciplinary conception of the archaeological investigations, at the crossroad of many archaeometric disciplines, 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 [15], related to ongoing EU project "Beyond Archaeology"(BeArchaeo[9]), 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[16][10] and the CRMsci model[17][11], 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 the 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. [18, 19], among others), analysis (e.g., GIS [20]), exhibition (starting from the virtual archeological reconstructions of the 1990s [21, 22] and addressing general public outreach and participation [23]).

A particular mention goes to the pioneering Çatalhöyük project, concerning a Neolithic settlement in Turkey, carried out with the goal of maintaining the data as long as possible. The Çatalhöyük Database and the Çatalhöyük Image Collection Database[12] make the documentation of the Çatalhöyük excavation site available. Custom platforms allow for the search of data uploaded during every excavation season and then made available through the Çatalhöyük Living Archive, which tells about two decades of excavations and analyses.

Project ARIADNE provides an event-centric ontological representation of the archaeological excavation relying on CRMarchaeo and CRMdig ontologies [24]. However, the legacy of the ARIADNE project, which currently continues with ARIADNEplus, is to be a web of interlinked archaeological datasets that comply with the Linked Open Data principles. The effort required to project partners is to convert and work with data in the (not always familiar) Semantic Web formats. A large amount of digital data demand for the coherence of recorded information, as contributed by each intervening discipline.

However, even across projects within single institutions, the global picture is a "rather disparate grouping, or 'archipelago', of diverse, specialized, but rather isolated and independent information systems and databases" [25]; limits concern sharing and standardization of data [26]. Also a survey made within the

[5]https://nkos.slis.kent.edu, visited on 15 January 2022.

[6]https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, visited on 15 January 2022.

[7]http://www.cidoc-crm.org, visited on 15 January 2022.

[8]Check projects in the portal https://portal.ariadne-infrastructure.eu, visited on 15 January 2022.

[9]https://www.bearchaeo.com/ (last visited on 15 January 2022).

[10]http://www.cidoc-crm.org/CRMarchaeo/, (last visited on 15 January 2022).

[11]http://www.cidoc-crm.org/CRMsciCRMsci/:, (last visited on 15 January 2022).

[12]http://www.catalhoyuk.com/research/database (last visited on 15 January 2022))

ARIADNEPlus project[13] 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 [27]. 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 [28]. Kansa and Kansa get to promote a general "data literacy" for archaeologists, who should care personally for their own data, through direct management and communication [29].

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 [30]. 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 [31]. 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 [32], on which a number of projects are based: for example, the two projects, namely EAMENA (Endangered Archaeology in the Middle East and North Africa)[14] and ASOR (American Schools of Oriental Research) Cultural Heritage Initiatives for Syria and Iraq[15], 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) with its encoding standards (e.g., Earth Observation GeoJson) and system integration interfaces (e.g., WMS – Web Map Service), which ensure compatibility with GIS applications (e.g., ArcGIS and Google Earth), common browsers, and online map services. Also, Arches includes modules for vocabulary management, such as Getty Art and Architecture Thesaurus[16].

3. Digital data curation and the BeArchaeo project

Digital data curation consists of 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 added value 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 [33]. 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 Curation Lifecycle Model [34] and I2S2 Idealized Scientific Research Activity Lifecycle Model [35]. Here we describe the digital data curation through an abstract representation of the tasks, adapted from [9].

3.1. Digital Data Curation model

The Digital Data Curation model consists of six common tasks (blue circles in Fig. 1) for the management of data directly acquired from the cultural heritage asset to the final outputs of some publication or exhibition. From left to right, we can notice an increasing abstraction of digital data, until interpretation; then data are archived as documentation (top) and/or employed in the exhibition of the results (bottom). 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

[13] D2.1 Initial Report on Community Needs https://ariadne-infrastructure.eu/wp-content/uploads/2019/11/ARIADNEplus_D2.1_Initial-Report-on-Community-Needs-1.pdf, dated 31 October 2019, visited on 15 January 2022.

[14] https://eamena.org, visited on 15 January 2022.

[15] https://www.asor.org/chi, visited on 15 January 2022.

[16] https://www.getty.edu/research/tools/vocabularies/aat/, visited on 15 January 2022.

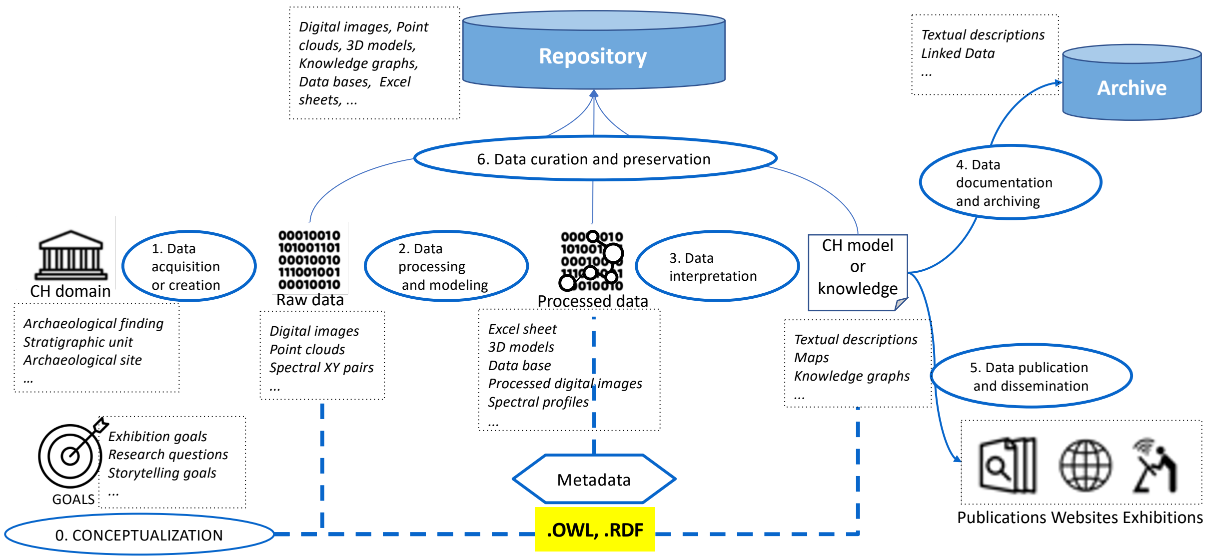

Fig. 1. Abstract representation of the digital data curation model.

archaeological finding (including fragments), a stratigraphic unit, the whole archaeological site.

Conceptualization

The conceptualization phase (numbered 0), which is the major focus of this paper, provides a knowledge framework to define the model for the digital data that are produced during the project implementation. The BeArchaeo ontology, presented here, addresses the archaeological knowledge, the archaeometric knowledge, and the design of the forms to be filled during the archaeological/archaeometric endeavor. The heritage involved and the goals of the digital curation project determine what part of the ontological model is used, providing the backbone for the database schema design that will account for the description and encoding of the digital data produced by the project.

Data creation or acquisition

Digital data curation typically starts with the data creation or acquisition (numbered 1) 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 results into digital numeric values. Archaeology usually includes operations such as laser scanning or photogrammetry, while archaeometry includes scientific tests, such as radiography or observation under an electron microscope. The growing involvement of archaeometry in the archaeological research is generating huge sets of digital entities from a variety of instrumental measurements, which can be performed either on the archaeological objects or on samples detached from them.

Data processing and modeling

The data processing and modeling phase (numbered 2) focuses on creating a conceptual model for the data to be stored in a database or spreadsheet, together with the associations between different data objects and the rules (many projects employ E-R Model and UML format). The goal is to support effective exchange of knowledge and interoperability. This phase can be iterated and/or concerning several acquired data objects. As an example, we can consider the realization of 3D models from point clouds of an archaeological finding and its chemical elemental composition. Even by employing the same scientific technique for determining the chemical elemental composition (for example, X ray fluorescence), the composition can be produced as a qualitative table, a quantitative table, or a chemical map of the surface, according to the equipment that is used for the investigation. Different digital objects are therefore produced and each of them gives different information. The role of the data processing and modeling phase is therefore crucial to clarify this point and to enhance the quality of the subsequent phase of interpretation.

Data interpretation

Data interpretation (3) is the process of making sense of data that have been collected, analyzed, and presented. This phase has a strong connection with the reflexive methodologies addressed above. Interpretation can be carried out by humans or machines; the result can be an explanatory text in natural language, a revealing diagram, or, in the case of semantic reasoning, a chain of inferences or a knowledge graph. The members of the project can access a holistic overview of the data and the interpretations can concern individual items, sets of items, or higher-order categories: the dating of an archaeological finding, with its motivation (relying on other digital data) and the maps with the paths of materials from source locations to final locations are two frequent examples.

Data documentation and archiving

The data documentation and archiving process (numbered 4 in the figure) manages the metadata about some data product (e.g., database tables) that enables one to understand and use the data. It concerns all the data that actually contribute to the interpretation and greatly supports the reflexivity. Data and documentation 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).

Data dissemination and publishing

Data dissemination and publishing (5) is the distribution or transmission of statistical data or of the knowledge arising by the overall process to end-users, made available in some online structured format or as paper publications (i.e., PDF files) based on aggregated data, as well as the exhibitions and websites of the collections owned by the cultural heritage organizations. Finally, the task of data curation and preservation (6) records all the data and metadata created during the first 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.

3.2. Application of the model to BeArchaeo example

Now we illustrate this model of digital data curation with an example that is related to some digital data generated from an archaeological finding during the BeArchaeo DDC-born archaeological project. The project carries out an archaeological excavation and the related archaeometric analyses of the Tobiotsuka 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 and other archaeological materials now stored in museums and laboratories, in Japan[17].

The project activities and outcomes are accessible to the general public through engaging media communication along the project development. In this section, we apply 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, undergoing 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 a comparison of the component materials, including elemental composition, morphological features, presence, typology and composition of inclusions such as minerals or rock fragments.

The digital curation workflow starts as soon as SH1, an archeological finding fragment, has been found. In particular, Figure 2 addresses a measurement carried out in the lab, where scientists acquired images of the fragment by Scanning Electron Microscopy (SEM),

[17]BeArchaeo website https://www.bearchaeo.com/ (last visited on 15 January 2022)

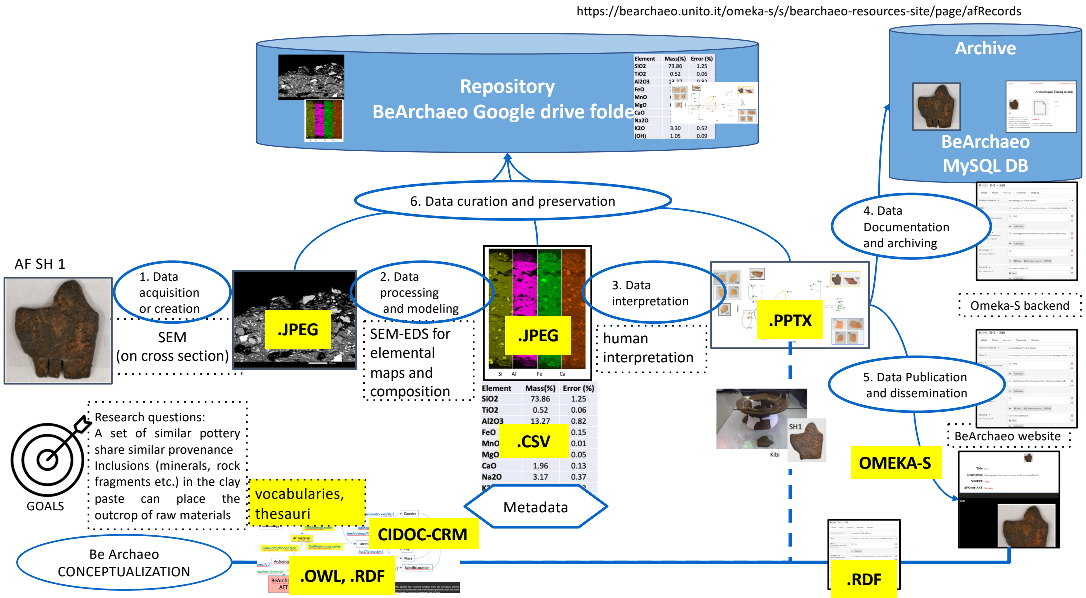

Fig. 2. Digital data curation model applied to the archaeological finding SH1 in the BeArchaeo project.

coupled with Energy Dispersive Spectroscopy (EDS). The 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 a feature of the asset at some level (e.g., the possible presence of a surface coating). Elemental maps of a portion of the sample, which are visible in the figure, highlight that the coating is depleted in Al_2O_3; later, it may suggest an enrichment in iron compounds, which would indicate that a coating was actually present. Such information derives from the combination of different scientific tests and different expertises. In a digitally–born project, the need to harmonize the procedures strongly supports the synergistic interaction. An example, which we can use for sake of simplicity, can refer to the archaeological question of defining if an archaeological finding (e.g., a pottery fragment) may share a common origin with other fragments that have been found in other archaeological sites. The question can be faced, as a first instance, by determining the elemental composition of the fragment. Presently, it has been determined by induced coupled plasma optical emission spectroscopy (ICP-OES). Raw data must guarantee interoperability and reuse; then, the acquisition step must guarantee that all the information on measuring conditions and procedures is recorded (as also stated in [36]). The processing and modeling step produces the information on the quantitative elemental composition of the sample, ensuring a high-quality base for data interpretation.

In the interpretation step, we can compare the elemental chemical composition of the fragment with the compositions of other fragments, so that the hypothesis of a common manufacture can be discarded or supported, respectively. In the latter case, we can go on with building the multidisciplinary knowledge by including, in the decision process, further items from the investigations with other scientific techniques (such as optical microscopy or mineralogical/petrographical data) which can lead to discard/support the interpretation made with elemental analyses data. A single operation of data acquisition plus processing and modeling can be included in many interpretation processes, supporting reflectivity and fertilizing interdisciplinarity. The intermediate and the final data are stored into the repository, currently a Google drive shared folder (to evolve into a more effective data repository connected to the database), through the tasks of Data curation and preservation. Moreover, the interpretation, in the format of powerpoint slides, is 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 conceptualization phase, to ease the problems of interoperability and connection between the archeological and the archaeometric data.

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 defined in a RDF file, the definition of customized vocabularies, and the construction of templates for the instantiation of filling forms [15].

Related to these concerns and potential interpretations, the database design of BeArchaeo project provides the information structure to all the digital curation phases of the project. In this case, it provides a repository while creating the archive of the archaeological findings with the related media. Media and metadata are stored in the BeArchaeo database as Archaeological Finding form, interfaced by an Omeka-S based web platform, in order to support the archaeologist's work in recording the excavation and interpretation activities.

4. Transdisciplinary conceptualization of the archaeological/archaeometric investigations

Given the digital data curation schema, which involves a conceptualization addressing several disciplines, we have developed the BeArchaeo ontology, with the design principle to capture the connections between the archaeological and the archaeometric realms, respectively. Transdisciplinarity is mediated by the formal ontology, with research questions arising from the collaboration between the disciplines [37]. The BeArchaeo ontology pivots on the description of the objects, and merges the general archaeological and archaeometric entities with the fields of the catalogue records [15]. Design patterns, for connecting 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 archaeological knowledge (lower left part of Figure 3), the archaeometric knowledge (lower right part of Figure 3), and the catalogue record knowledge (upper part of Figure 3).

Figure 3 provides an overview of a sample encoding. Going left to right: the stratigraphic unit "SU 202" (content of the title field of the catalogue record for this unit) is the source of the archaeological finding "AF 59" (content of the title field of the catalogue record for this finding); the type of the finding is "Sue (ceramics style)", as selected from the Getty-AAT thesaurus and "sekki", as selected from the BeArchaeo thesaurus; the finding body[18] has undergone some chemical test for calcium oxide (*CaO*, a measurement activity), which has produced a result in wt% value. A data evaluation process assigns some dimension, namely an attribute for the body predominant composition ("Calcareous").

The realization of the BeArchaeo ontology relies on the CIDOC-CRM reference model family. The pyramidal CIDOC-CRM family of models (Fig. 4, right[19]) extends the general documentation model (entities identified with prefix *cidoc-crm*) through specialized thematic models for the needs of projects and organizations. In particular, CRMdig is a model for provenance metadata, CRMgeo is a model for spatio-temporal entities. Of particular interest for the archaeological and the archaeometric endeavors, we address the CRMsci and the CRMarchaeo models, respectively. We plan to deal with an ontological model of provenance in the future; currently, we have encoded provenance in the notes of the investigation processes (see Figure 8).

In Figure 4, we can see the overall picture. Colors distinguish the ontological module of the classes: turquoise rectangles identify CRMsci classes, ochre rectangles are CRMarchaeo classes; grey rectangles are core CIDOC-CRM classes; finally, white rectangles are BeArchaeo classes. The figure illustrates the major relationships between BeArchaeo ontology and the CRMsci and CRMarchaeo reference models, as well as the references to the two archaeological thesauri BeArchaeo-AFT (Archaeological Finding Thesaurus), for a taxonomy of Japanese history materials, built within the project, and Getty-AAT (Art and Architecture Thesaurus). The major classes are *bearchaeo/ Archaeological_Finding* and *CRMarchaeo/A8_Stratigraphic_Unit*, which describe the objects that tangibly connect all the tasks related to an

[18]Usually, for chemical tests, an archaeological finding is considered as composed a body, a coating, and an embellishment.

[19]Pyramid on the right is reported from Martin Dörr's CIDOC-CRM extension suite presentation in Nuremberg, Germany, May 19, 2015, https://slidetodoc.com/cidoc-crm-family-harmonized-models-for-the-digital/.

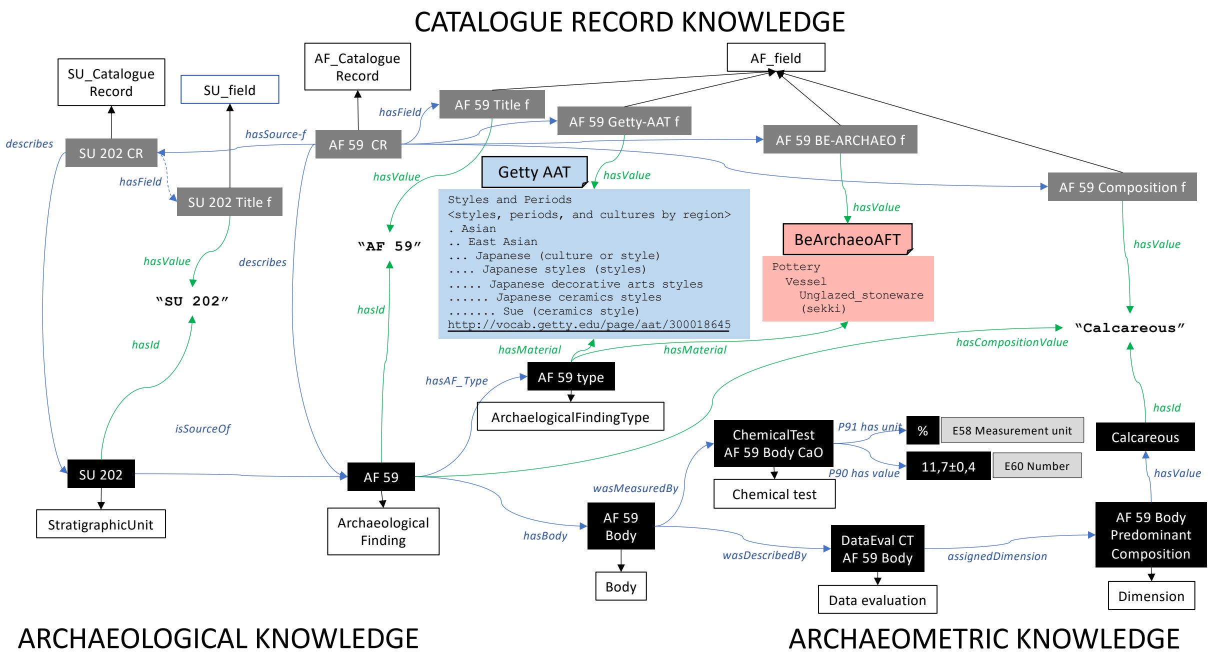

Fig. 3. Modeling of the archaeological finding “AF 59”, exemplifying archaeological and archaeometric knowledge, respectively, and the corresponding fields in the archaeological finding record. The rectangles in grey or black are the individuals; the white rectangles are the classes; object properties 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.

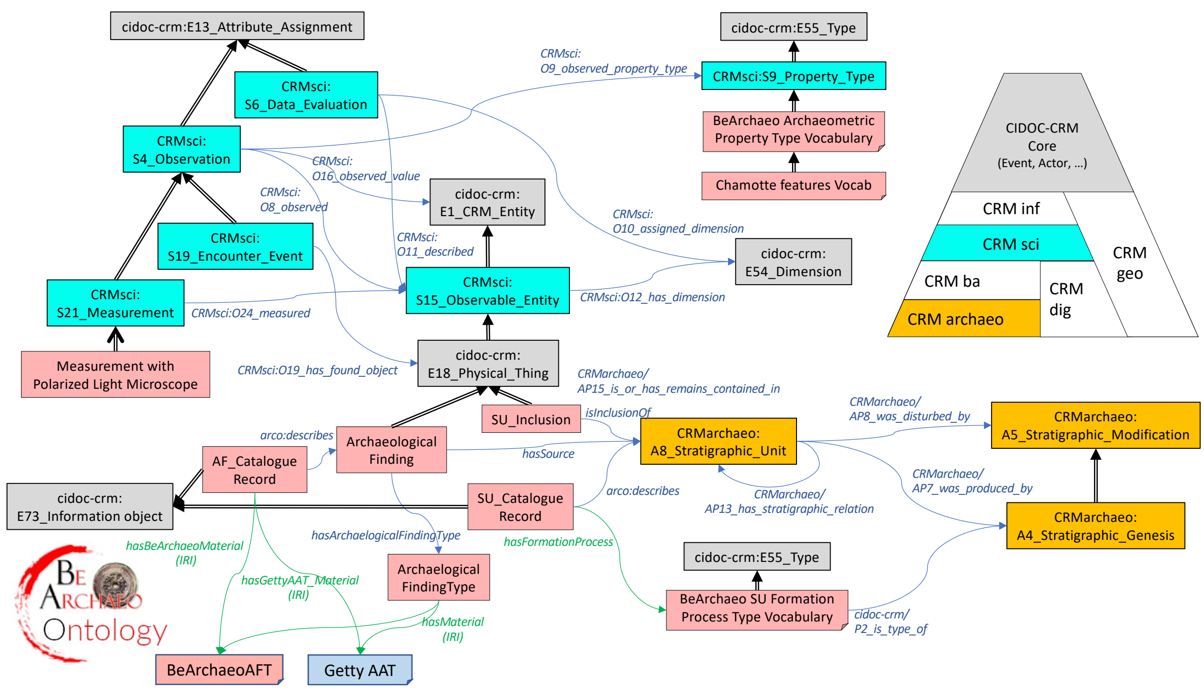

Fig. 4. Major relationships between BeArchaeo and CIDOC-CRM family. Colors are employed to distinguish the ontological modules.

archaeological investigation (a stratigraphic unit is the source of some archaeological finding or at least of some inclusion, a fragment of some material that is relevant for the investigation). They are connected with the related catalogue records (*bearchaeo/AF_-Catalogue_Record* and *bearchaeo/SU_Catalogue_-Record*), which describe the respective objects. Class *bearchaeo/Archaeological_Finding* specializes class *cidoc-crm/E18_Physical_Thing* and has a type, which refers to the specialized vocabularies, Getty-AAT and BeArchaeo-AFT.

CRMarchaeo reference model takes inspiration from Harris' model [38], which accounts for the stratified arrangement of an archaeological excavation. The excavation model includes the description of the dichotomy between the (natural or human) phenomena that produced the stratification (centered around the class *CRMarchaeo/A1_Excavation_Process_Unit*) and the units that are the outcome of the generation/modification process (centered around the class *CRMarchaeo/A8_Stratigraphic_Unit*). Stratigraphic units contain some remains, classified as physical objects (centered around the class *cidoc-crm/E18_Physical_Thing* of the core ontology). Stratifications and their contents are analyzed and interpreted to determine the relative chronological order of the strata, then the classification and functionality of the objects therein, up to the high-level reconstruction of the beliefs and behaviors of some group of people in the past in that place. A stratigraphic unit, produced by some genesis process (*CRMarchaeo/A4_Stratigraphic_Genesis*), can also be modified by a *bearchaeo/A5_Stratigraphic_Modification*, of which formation process types, acknowledged by the official excavation recording forms, are a specific vocabulary.

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).

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. As observed through the example in Figure 3, the development of the BeArchaeo ontology comprises three modules, the archaeological knowledge, the archaeometric knowledge, and the catalogue record knowledge, with connections to standard ontologies and the inclusion of non-ontological resources. In particular, the third module concerns the form through which the first two modules are recorded for the digital data curation process. In the rest of this section, we address the major decisions for the ontology modeling process and then we provide an overview of the classes and properties of the BeArchaeo ontology.

5.1. BeArchaeo ontology modeling process

Here we go through the methodology addressed, the technical structure of the ontology, its alignment with standard models, the logical profile implemented, and the technicalities and documentation of the released model.

Methodology

Given the three knowledge sources we are addressing, we have employed a number of scenarios from the NeOn methodology [39]. In particular, the development of the catalogue record ontology falls in the Scenario 1, going from the specification of the form entries to the development of the ontology from scratch. We analyzed the materials provided by the national institutions (check details in [15]) to conceive a set of classes and properties that describe the fields that form the catalogue records and how they are connected with the archaeological and the archaeometric knowledge. The goal was to employ a semantic database and a semantics-based web-publishing platform to implement the form filling operations. The semantic relations of the database underlying the forms are connected to the archaeological and archaeometric knowledge sources.

Scenario 2, which concerns the inclusion of non-ontological resources into the formalization, manifested in the work with a number of small and large vocabularies, such as, e.g., the 5-termed Compaction value vocabulary used by the archaeologists and the large Munsell color system, used by the ar-

chaeometrists (especially pedologists), respectively, to single out a stratigraphic unit.
The reuse and merge of CMRarchaeo and CRMsci standard resources as well as the WGS84 vocabulary fall under the Scenario 5, i.e. the re-use and merge of other ontological resources; actually, a number of other resources should be integrated to represent historical epochs and chronology. However, in these cases, we have deferred the alignment to a future work, because there are many conventions used in the archaeological research documentation that require more time to be addressed correctly.
Scenario 9, useful for the adaptation of the ontologies to other languages and cultures for the production of a multilingual ontology, has been implemented in the development site for the Japanese archaeologists (who did not feel comfortable with English-based terms) and is currently under testing[20].

Modularization

The ontology consists of three subontologies: Catalogue record structure (split into sections), Archaeological knowledge and Archaeometric knowledge. The three modules have some interfaces, namely, the major archaeological categories of Stratigraphic units and Archaeological findings. For practical reasons, for the implementation of the web interface to the forms, we split in turn the Catalogue record knowledge about the stratigraphic units into further five subontologies, as implemented by the forms of the Italian Ministry of Culture [15]: the "registry" section (identifiers and spatial information such as room, trench, area, ...), the "description" section (with inclusions and soil attributes), the "stratigraphy" section (for the relations with other stratigraphic units), the "dating" section (for elements relevant for chronology), and the "sampling" section (data about the excavation process).

Alignment

Alignments concern mostly the Archaeological knowledge of BeArchaeo with CRMarchaeo model and the Archaeometric knowledge with CRMsci model, respectively. Both the archaeological module and the archaeometric module, together with the catalogue record module are aligned with the core CIDOC-CRM model. Figure 4 shows these alignments: Archaeological findings and the Inclusions of the stratigraphic units are subclasses of the physical things in CIDOC-CRM core model. Catalogue records are subclasses of the information objects, again in the CIDOC-CRM core model. BeArchaeo stratigraphic unit is the same class as CRMarchaeo stratigraphic unit, and the BeArchaeo formation process is a subclass of the stratigraphic genesis class of the CRMarchaeo model. Archaeometric classes are generally subclasses of the CRMsci classes: measurements are specialized into several subclasses of measurements (e.g., with Polarized Light Microscope) and property types into specialized vocabularies (e.g., Chamotte features vocabulary).

Logical profile

The current development of the BeArchaeo ontology is expressed in OWL2 EL language. There are a few axioms that represent the necessary and sufficient conditions for some specific classes, related to the catalogue records. Possibly, the archaeological and archaeometric modules should require some more expressive axioms, in order to check the consistency of the conclusions reached within the archaeological realm with the knowledge from the archaeometric analysis and evaluations.

Technicality and documentation

Classes and properties are commented extensively and a LODE implementation provides the documentation of the merged BeArchaeo ontology[21]. The catalogue record model has been described with a number of subontologies concerning the five sections of the stratigraphic unit record (SU catalogue record) and one subontology for the archaeological finding record (AF catalogue record); then, one module for the archaeological knowledge and one module for the archaeometric knowledge. The several subontologies of the SU record concern the sections, which in turn contain a number of fields. The class *SU_CatalogueRecord* is connected to the sections with the property *hasSection*; each section class is connected to its field with the property *hasField* (see instantiated case in Figure 3). The ontologies for the records are connected to the archaeological knowledge through the property *arco/describes*, as introduced by project ArCo[22] for the relationship between an entity that describes another entity in the field of cultural heritage [40]. The ontol-

[20]See the experimental Japanese version of the database, https://bearchaeo.di.unito.it/omeka-s/s/jtoppage/page/welcome, visited on 15 January 2022.

[21]http://purl.org/bearchaeo/bearchaeo_lode, visited on 15 January 2022.

[22]http://wit.istc.cnr.it/arco/

ogy is expressed in OWL/RDF formats and published at two permanent addresses[23].

5.2. Overview of BeArchaeo classes and properties

Now we provide an overview of the archaeological and archaeometric modules; the classes and properties of the catalogue record module, sketched in Figure 3 reflect the entities presented here and are accessible through the web platform interface implemented for the scientists to insert their data during the excavation and the laboratory work (Figure 11).

The Archaeological module

In the figures 5 and 6 there are the classes, vocabularies, and properties concerning the description of the stratigraphic unit and the archaeological finding, 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 have 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 taken during the excavation, with a particular concern for Elevation. The distinguishing criterion determines how this unit has been identified: the terms that concern this attribute are three (Color, Composition and Compaction) and there are other three properties that possibly specify the actual values for such attributes (namely 6-valued soil/matrix term for composition, 5-valued term for compaction, and a free string for color). Color, in the relationship with archaeometrists (specifically, the soil scientists) can be recorded with the encoding provided by the well-known Munsell color system, in use in pedological studies[24]. Finally, the formation process concerns a specialization of the processes that are responsible for the creation and modification of the stratigraphic unit, with a frequent term vocabulary, which can be further augmented with free text insertion. The properties in the center of the figure specialize the stratigraphic relation property (*CRMarchaeo/AP13_has_stratigraphic_relation*):

- *sameStratumAs*, for two stratigraphic units that are claimed to belong to the same stratum of soil interrupted by some intervening unit[25];
- *isBoundTo*, for a stratigraphic unit that is a limit for another one;
- *abuts/isAbuttedTo*, for a stratigraphic unit that edges another one;
- *cuts/isCutBy*, for a stratigraphic unit that introduces a discontinuity into another one;
- *covers/isCoveredBy*, for a stratigraphic unit that covers (stands over) another one;
- *fills/isFilledBy*, for a stratigraphic unit that has filled a cut (see above);

Also, there are two temporal relations, *laterThan* and *earlierThan*, resulting from the interpretation of the stratigraphy. The latter terms, which originate from the terminology reported in the institutional records of the excavation recording, shall be later aligned with some general temporal ontology.

An archaeological finding (Figure 6) 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 (and many other projects), because of the employment of the ontology into the design of the final exhibition. The archaeological finding has a reference type and some component material. Types refer to terms in the previously mentioned Getty-AAT thesaurus and the BeArchaeo-AFT thesaurus, the latter encoding knowledge from an authoritative Japanese reference [41]. Also the component material has a type (referred again in Getty-AAT) and the information about the administrative location. Finally, an archaeological finding is marked with its chronology, currently limited to a free text insertion,

[23]URL http://purl.org/beArchaeo/beArchaeo_merge_all merges all the other sub-ontologies. Also a GitHub repository is accessible through the other permanent URL https://w3id.org/bearchaeo.

[24]Munsell 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 https://munsell.com/about-munsell-color/how-color-notation-works/, visited on 30 September 2021.

[25]This term represents the relationship between two stratigraphic units that belong to the same stratum. While the other terms in this list come from the institutional documentation on archaeological excavations, the term officially used for this equality relationship, namely *isEqualTo*, looked awkward in the Semantic Web community and certainly does not coincide with OWL property *sameAs*. However, we preserved the term *isEqualTo* in the forms, to ease the archaeological practice.

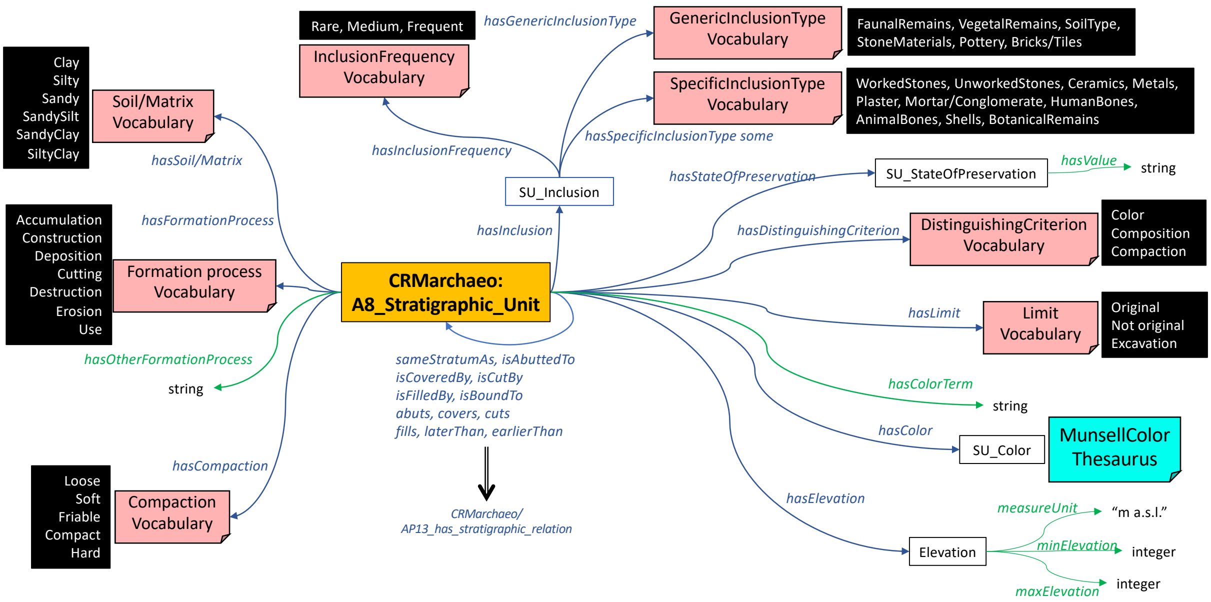

Fig. 5. Conceptual model of the stratigraphic unit knowledge (including references to thesauri and vocabularies (with list of terms)).

together with its motivation, but with the idea of providing an encoding in the terms of a time ontology, with possibly many alignments, depending on the disciplinary traditions in both archaeology and archaeometry.

The Archaeometric module

Archaeometry is a vast endeavor. As far as we know, this is the first attempt to model the archaeometric investigation in a digitally-born archaeological project. We want to keep record, in the digital data, of the decisions made during the analysis (going from acquisition to processing and interpretation) and to relate the archaeometry-based interpretation with the evaluations, data, and interpretations conveyed by the archaeologists. The focus of the project is on the documentation and dissemination of the results; in the future, we plan to also address consistency and inference between the disciplines participating into the endeavor, with the semantic web encoding.

The current development of the BeArchaeo archaeometric module implements a trade-off between a wide appraisal of the archaeometric domain, with its processes and data formats, and the needs of the BeArchaeo project, which addresses a restricted set of archaeometric investigations in detail. However, the alignment of the archaeometric module with the CRMsci standard model and the richness of the multidisciplinary team working on the project provides us a wide scope. Now, we first address the conceptualization of the archaeometric model; then, we give an insight on the ontological model; finally, we illustrate two paradigmatic examples.

Conceptualization of the archaeometric model

The goal of the conceptualization phase for the archaeometric module is to provide a coherent and cohesive structure for all the archaeometric investigations, which work in a transdisciplinary setting, mutually influencing one another. The several disciplines specialize the CRMsci reference model through the specific processes and the corresponding digital data formats. The disciplinary researchers have been asked to speculate on the procedures and results concerning the stratigraphic units and the archaeological findings, in order to single out the 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 core CIDOC-CRM and CRMsci models.

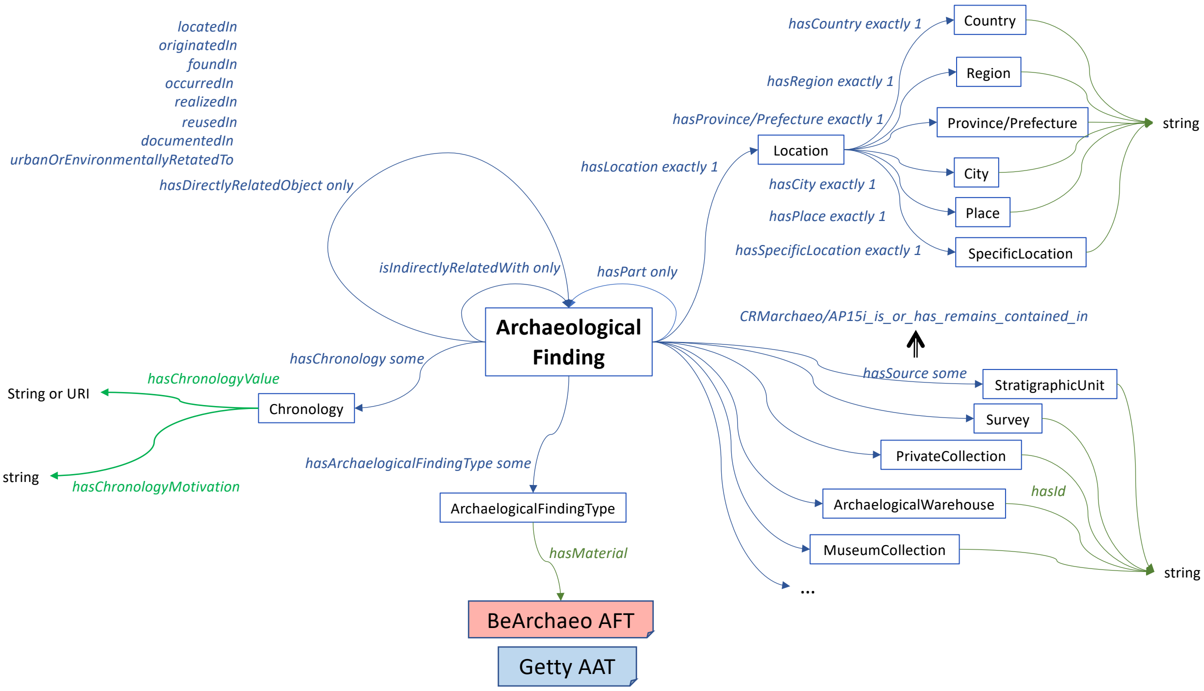

Fig. 6. Conceptual model of the archaeological finding.

Figure 7 shows a portion of the upper level structure of the measurements that occur in the archaeometric domain, when dealing with the archaeological findings. BeArchaeo archaeometric measurements are all subclasses of *CRMsci/S21_Measurement*; classes are distinguished by the object measured (archaeological finding or stratigraphic unit), the measurement technique (e.g., Polarized Light Microscope, Thermoluminescence, Archaeomagnetism, Metabarcoding of microbial taxonomic diversity), and the material addressed (e.g., pottery, glass, organic remains). Specialized vocabularies identify the observed property types and, for each measurement, the observed values. Measurements are typed and also connected to some entry in the Getty AAT thesaurus (if this exists). For example, Figure 8 shows an instance of a measurement class concerning the X-ray Fluorescence Spectrometry (XRF), applied to the Archaeological finding "BA18". XRF has a type in the Getty AAT (300224161).

All measurements rely on a number of factors, such as environmental conditions, the actual device, with its settings and calibrations, precision, and scale. Following the indications provided by the CRMsci reference, this information is reported in a note, currently a string datum, connected through the *cidoc-crm/P3_has_note* property. Figure 8 reports the note for the XRF measurement, consisting of, e.g., the instrument that made the measurement, the voltage utilized, the beam size, and the number of acquisitions that have been done. As noticed, measurements address the acquisitions in the digital data curation pipeline, producing the so-called raw data (Fig. 1). So, we include such information into the catalogue record designed for the object. The same considerations hold for the processed data, where algorithms and software libraries are determinant for the achievement of the results. We are aware that a note is not the best solution for these relevant metadata and the connection to data provenance ontologies, such as CRMdig or PROV-O, is to be deployed in the near future.

We have currently developed classes and properties for archaeometric analyses such as: Polarized Light Microscopy, elemental chemical analysis by X-ray fluorescence (XRF) and induced coupled plasma optical emission spectroscopy (ICP-OES), molecular chemical analyses by Raman spectroscopy and Diffuse Reflectance Spectroscopy, Thermoluminescence dating, Archaeomagnetism, Soil morphological assessment, Radiography, Tomography, and Metabarcoding of microbial taxonomic diversity. In each case, we have developed specific vocabularies, geared to the project

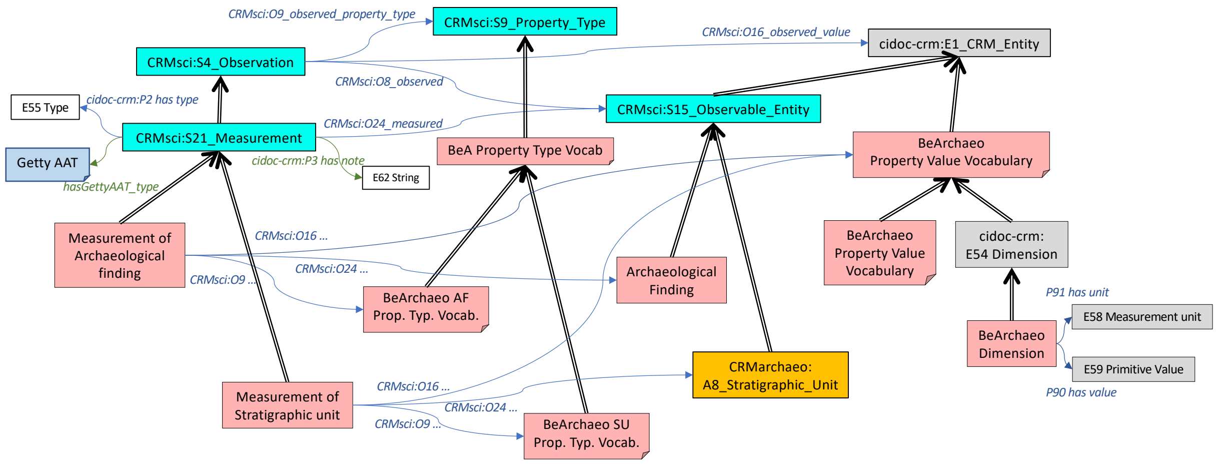

Fig. 7. Overall model of the BeArchaeo archaeometry ontology.

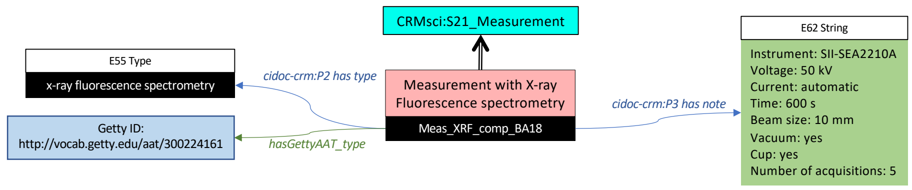

Fig. 8. Instance of the XRF measurement acquisition.

specificity. The alignment with external, comprehensive resources is planned for the near future.

To illustrate the depth of the knowledge encoding, we show the ontology developed for modeling archaeological pottery investigation by means of morphological qualitative methods (Figure 9), in particular polarized light microscopy. Analogous ontological models have been deployed for the other archaeometric processes mentioned above; below, we also show how the several investigations converge on the evaluation for achieving an interpretation.

The model is based on the annotation structure suggested by Quinn for the investigation of pottery prepared as thin sections [42]. The transdisciplinary 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. In particular, the model fleshes out the similarities spanning the diverse disciplinary procedures, by replicating the same major structure developed for modeling the analyses of thin sections in pottery investigation to other scientific tests. It models, in particular, 1) the investigation in cross section of pottery, 2) the determination of quantitative chemical composition by XRF in 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 polarized optical microscope of the archaeological ceramics in thin section reveals the complexity of these materials (Figure 9). 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*, *_Matrix*, and *_Voids*). The representation of how pottery thin sections are analyzed by means of optical microscope under polarized light consists of 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 heterogenous, or mineral/petrographic component, with subtypes 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).

Finally, connected to the data interpretation of the digital data curation schema is the modeling of data

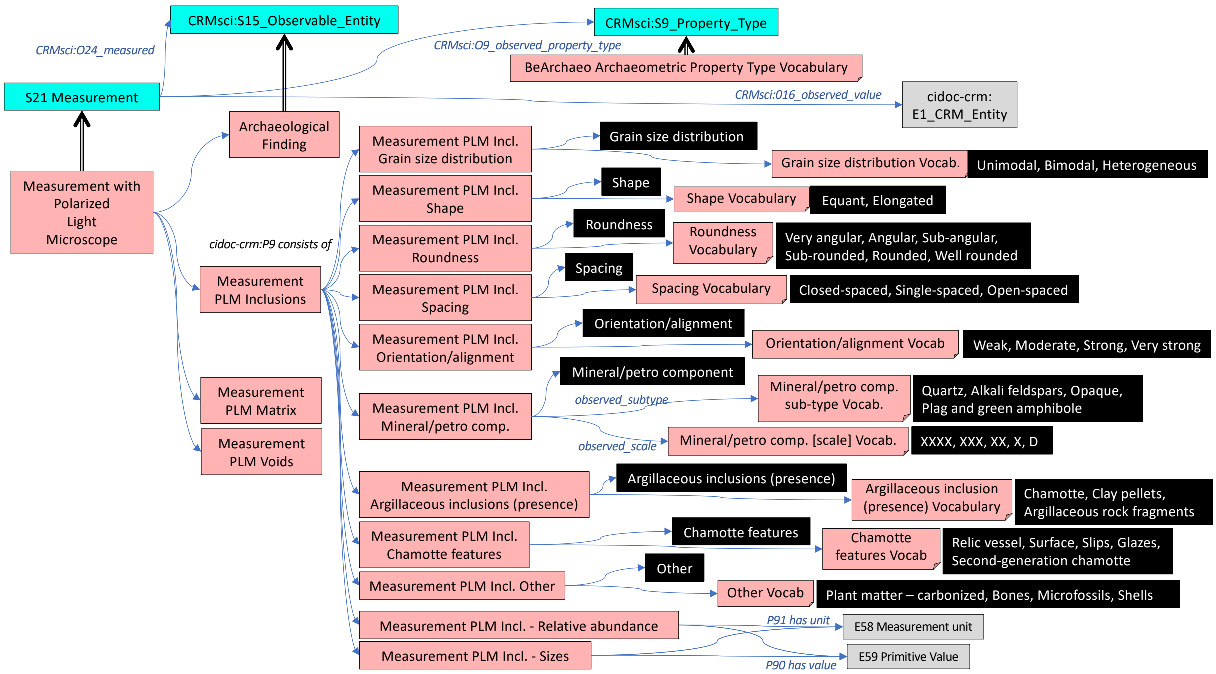

Fig. 9. Investigation of archaeological pottery prepared as thin sections through polarized light microscopy.

evaluation (class *cidoc-crm/S6_Data Evaluation*) that follows acquisitions/measurement and processing. In the instantiated model reported in Figure 10, the results from Thermoluminescence, 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 are obtained from other scientific techniques, initially used to obtain other type of information. They also produce data that can be exploited to cross-check knowledge in an interdisciplinary environment as each contribution independently suggests specific temperature ranges. In the next section, we see how this information is annotated by the BeArchaeo archaeometric team in the database to reflect such a transdisciplinary approach.

6. Preliminary evaluation of the model in the BeArchaeo Project

The digital data curation of a few findings in the BeArchaeo project forms a preliminary evaluation of the BeArchaeo ontological model. As the conceptualization and modeling of the archaeological and the archaeometric knowledge proceeds, we have developed a web platform for the form filling of the scientists, based on the catalogue record model. So, we can report on some preliminary evaluations of the approach.

6.1. *Deployment of BeArchaeo ontology for the Tobiotsuka kofun excavation*

Project Beyond Archaeology (BeArchaeo) consists of the archaeological excavation, archaeometric analyses, interpretation of the findings, and eventually dissemination of the results about the Tobiotsuka Kofun (Soja city in Okayama Prefecture), and other archaeological materials of the ancient Kibi and Izumo areas now stored in museums and laboratories, in Japan. The ontology described above underlies a semantic database for the encoding and storing of the digital data concerning the documentation of the archaeological excavation and the account of metadata that arise from the archaeometric tests and interpretations[26]. In particular, the project has drawn inspiration from the forms distributed by national authorities, which have

[26]https://bearchaeo.unito.it/omeka-s (last visited on 15 January 2022).

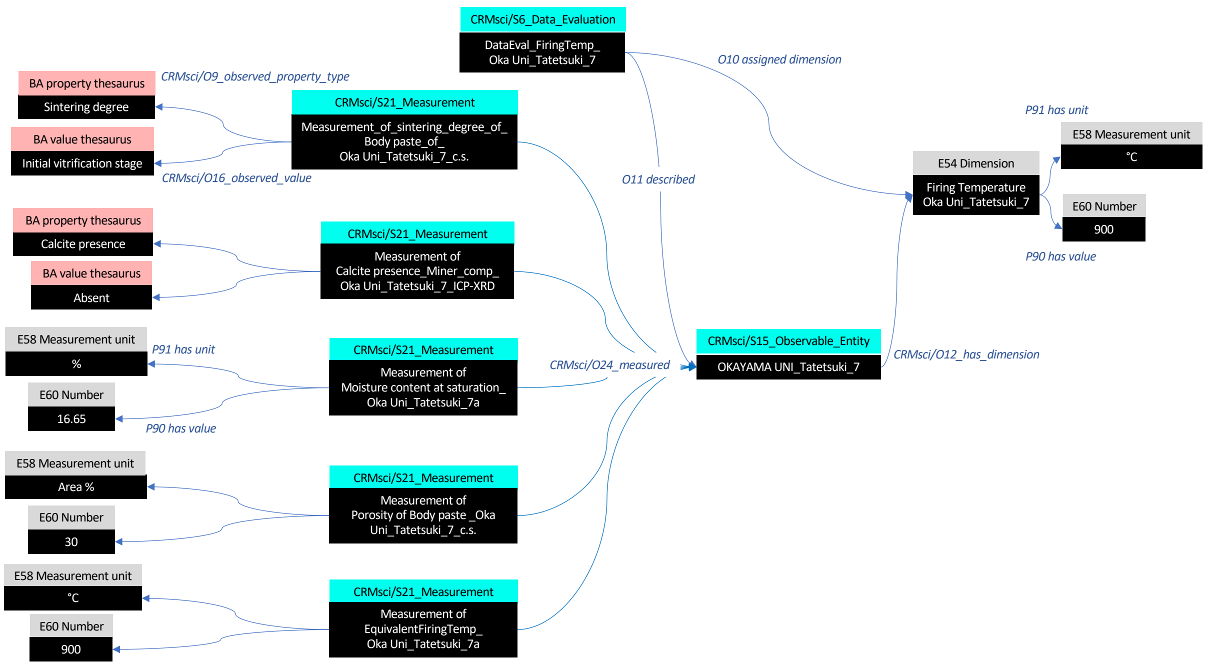

Fig. 10. Evaluation of data for the assignment of a dimension.

informed the classes and properties of the catalogue record module of the BeArchaeo ontology. The vocabularies addressed above have been encoded as custom vocabularies into an installation of the semantics-based Content Management System Omeka-S [27]. As seen above, the catalogue record module is connected to the archaeological and the archaeometric knowledge, and the plan is to perform inferences and consistency checking of the interpretations in the future.

The forms have been deployed as "Resource Templates", with the fast prototyping of user interfaces for both the back-end of the system, accessible by the archaeologists and the archaeometrists, and the front-end, where supervisors and stakeholders check the development of the archive and the related findings. Also, considering the multi-cultural and multi-lingual issues of the Be-Archaeo project, knowledge interoperability between Japanese and English researchers as well as data terminology have been addressed by providing also Japanese resource templates for the Archaeological Finding and Stratigraphic Unit records, respectively (currently, in the development site[28]). Also, we have uploaded rich media materials (photos and 3D models acquired from photogrammetry and scanning), that are being used for interpretation and will be the basis for the final exhibition. Figure 11 reports two images, from the back end and the front end, respectively, of the production website[29].

During the development of the BeArchaeo project, we could observe the behavior of the archaeologists and the archaeometrists, respectively. Archaeology and archaeometry are at a different stage of development with what concerns the curation of the digital data. The archaeologists have found the model accurate, mostly because of the connection of the model to the forms that are already in use, being the latter a conceptualization effort made by national authorities; so, the alignment of the catalogue record module with the archaeological knowledge resulted to be effective. The categorization of the data inserted through the form fields and the possibilities offered by the web platform to introduce and motivate different annotations has led to discussions between the team members, with an impact on the reflectivity issues mentioned at the beginning. Again, by relying on a web platform, the several roles of the users, namely Authors, Reviewers, and Editors, have contributed to a fruitful awareness of the re-

[27] https://omeka.org/s/

[28] https://bearchaeo.di.unito.it/omeka-s (last visited on 15 January 2022)

[29] https://bearchaeo.unito.it/omeka-s

sults of the project. The work with each archaeometric disciplinary team tackled the task of conceptualization within the small group at the beginning, focussing on the use of a specific investigation technique, and then extending it within larger disciplinary groups. The final broad discussion sessions have lead to the final procedures adopted within the whole multidisciplinary team. The modeling phase, which continuously enlarges its coverage, takes advantage of this transdisciplinary account of the data and the whole archaeometric team is gaining a great awareness of the similarities and differences of the procedures adopted within the disciplinary accounts, in a holistic perspective. The integration of the archaeometric and the archaeological knowledge, through a centralized database, has triggered an effort in the alignment between the interpretations provided by the different members of the team. In particular, the system has triggered discussions within the several disciplines of the archaeometric team and between the archaeological and the archaeometric teams, respectively.

6.2. *Workshop evaluation*

For evaluating the model, we have organized a two-half day workshop. Eighteen researchers, including the authors, participated. Nine were part of the BeArchaeo team, while the other nine were researchers working in archaeology and other related cultural heritage domains. The audience was international, with participants from Italy, Portugal, Brazil, Ukraine and Turkey, and multidisciplinary, with four archaeologists (with different period/location backgrounds), two museologists, one information scientist, one 3D modeler, one dating expert and nine archaeometers (with backgrounds in chemistry, biology, physics and Earth sciences, respectively). After a short introduction aimed at presenting the major theoretical and contextual background of the BeArchaeo database and the digital data curation schema, the audience were encouraged to employ the back-end interface provided on a development site (where experimental annotations and software modules are tested before being implemented in the production site). Also, they were asked to comment on the annotation schema while a moderator (one of the authors) was carrying on form filling activities, starting on exemplary findings and moving to novel archaeometric cases, to suggest individual encodings on the web platform.

A first general statement was that the semantic approach to the database led interdisciplinary teams to appraise the core on the encoding process and mediate between the various habits and practices related to established national or disciplinary procedures. Going cross-countries, in the team of the archaeologists, some supported the requirement of some national authorities for mandatory entries (encoded through object and datatype properties), while others have pointed out that other national authorities are less committed. The solution agreed was to leave semantic properties to be optionally valued, while developing specific interfaces for the national contexts (currently, we have a European interface (in English, based on the Italian Ministry of Culture forms) and a Japanese interface (only in the development site yet). Going cross-disciplines, the archaeometric areas that were not engaged in the current development of the archaeometric knowledge, for example the biologists, were able to catch the tenets of the semantic encoding; in practice, the workshop could trigger the process for the extension of the archaeometric encoding as well as identify the entities, namely the stratigraphic units for biologists, that can pivot the form filling process in synergy with the archaeological recordings. The issue of having some mandatory property also emerged for the archaeometric investigation. In particular, it seems that the property concerning "the acquisition details" should be mandatory, as it has been often stressed that instrumental details and sample treatment is very relevant information to be linked to scientific data. In the immediate future, we decided to act mostly on the interface of the filling forms, by providing a message that illustrates the importance of the acquisition details and the necessity of inserting such information in the individual entries of the archaeometric investigations.

All researchers acknowledged that being educated about the digital data curation schema underlying the semantic encoding was very helpful in understanding the form filling process, especially in the relationship between the archaeological annotations and interpretations and the archaeometric investigations and interpretations. In fact, while the current model is very inclusive in terms of the media and data to be included in the representation for a proper documentation of the outcomes of the on-the field and of the in-the-lab activities, respectively, there is an ongoing discussion in the archaeological disciplines on how to be effective in the report of selected information in the repository and how to deal with the interdisciplinary knowledge, in order to include and link the different cues that come from the different approaches. For example, one archaeologist pointed out that the representation must in-

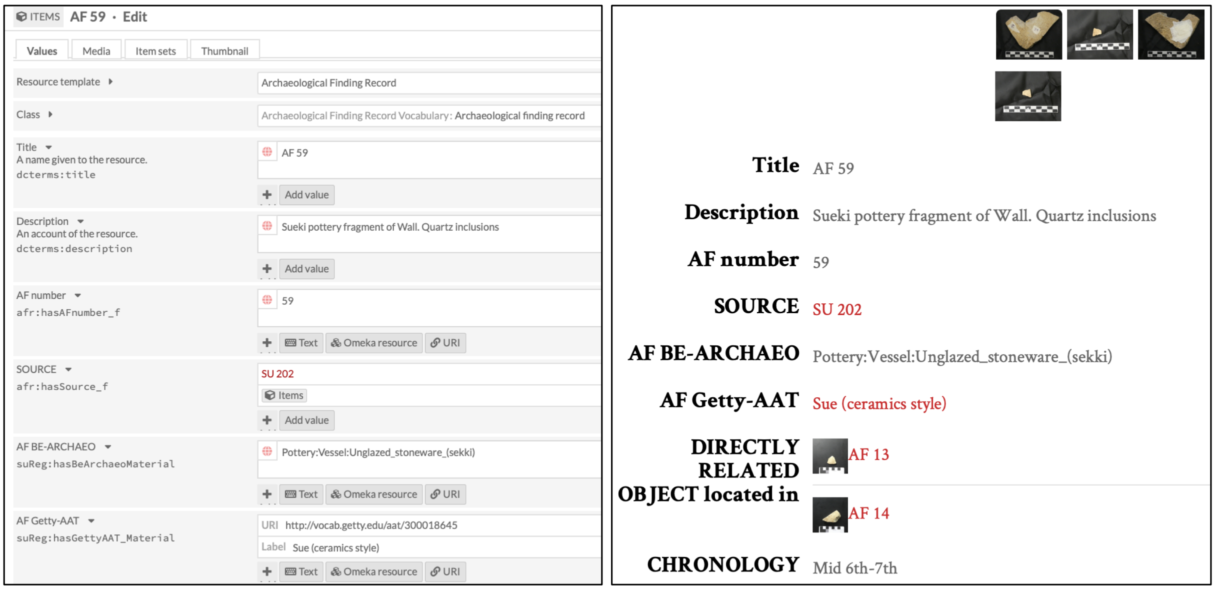

Fig. 11. Screenshot from the BeArchaeo resources website, concerning the Archaeological finding no. 59, with the related fields and media. On the left, the back end; on the right, the front end. Elements in red are links to other elements of the documentation (e.g., Stratigraphic Unit 202) or to some external knowledge source (e.g., Getty AAT thesaurus).

clude the Harris matrix to support the identification of the stratigraphic units; however, going back to the national issues above, some other noticed that the Harris matrix is not generally adopted in the Japanese archaeological studies. Indeed, a number of interesting issues also rose from the different excavation techniques that pertain the two schools of archaeology. Most of the archaeological knowledge available relies on concepts and terms, such as trenches, sections, and rooms, that have slightly different definitions according to the two traditions (e.g., in terms of depth of a trench accepted as a default); so, the ontological model should be adequately updated to include such differences and promote more fruitful collaboration for the international teams. However, the current representation has been deemed particularly valuable in supporting the construction of new knowledge through the many interpretations of the data that are linked to archaeological entities, together with the acquisition and processing phases that report on the setting and tools employed. In particular, some archaeologist reported that the organized repository could effectively support the comparison of the interpretations as they emerge while information grows from data production and modelling during the ongoing project activities. This is particularly appreciated in the context of the reflexive attitude in archaeology.

A missing feature of the current semantic model is the encoding of the sampling procedures, which are well described in the CRMsci model, as prominent in scientific investigations. In fact, it is customary to produce samples from some finding, in order to perform some individual measurements that are then compared to provide some parameter evaluation for the whole finding (this happens, e.g., for archaeomagnetism researchers). However, our efforts in the conceptualization process have given priority to the representation of objects that are composed from a number of fragments retrieved individually and subsequently analyzed to discover that they were part of a single object. Both fragments and composed object have the status of entities in the representation, with archaeological data and archaeometric investigations attached to them. For the immediate usage within the BeArchaeo project, the current representation of composed objects can be immediately adapted to the sampling issue, when limited to cases where the samples have the status of recorded items and not simply samples taken for measurements and then considered only a support of the interpretation process. Further developments are needed in the future to address this specific feature to provide a consistent representation of the archaeometric investigations.

7. 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 devised how the knowledge is linked to 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 the archaeometric knowledge. The conceptual model is the outcome of several modeling sketches and subsequent discussions carried out by the members of the archaeological and the archaeometric teams, representing the several 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 conceptual model and the ontology of the archaeometric knowledge serve the design and 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 born-semantic project that assumes a joint archaeological/archaeometric perspective from the start. In fact, the multi-disciplinary, multi-cultural, and multi-lingual characters of BeArchaeo raise a high demand of interoperability of knowledge and data. The alignment with CIDOC-CRM is pursued at the disciplinary level, by aligning the archaeological and the archaeometric descriptions through the CRMarchaeo and CRMsci models, where possible.

The realization of an overall approach, together with the adherence to well known standards and with an implemented workflow from the excavation design to the exhibition, can greatly contribute to the replication of the method across other projects. The BeArchaeo archaeological team is a proper representative of the "archaeological community": the Japanese archaeologists are strictly linked to the Japanese Research Institute for the Dynamics of Civilization[30], the Portuguese archaeologists are part of the Centro de Arqueologia de Universitade de Lisboa[31], and the Italian archaeologists are set within the International research Institute for Archaeology and Ethnology[32]. Also, after BeArchaeo, the model is going to be adopted in further initiatives in Europe (e.g., check the networking session of the UNITA project on October 2021[33]).

In the next future, we continue the encoding of further archaeometric aspects and the strict connection with the archaeological interpretations, to implement some form 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. The Omeka-S frontend, which has been an immediate solution for monitoring the project initial database schema (given some previous experience with the tool), will be replaced by a customized interface, while continuing to serve as a backend to the database monitoring. We are also working on a novel repository (currently a Google drive folder) for the media supporting the archaeometric analyses and interpretations. In particular, we are currently in the phase of analyzing the requests about the possible uses of the data in the future, in order to devise the best repository solution.

Finally, we are going to evaluate the contribution of the centralized semantics-enhanced digital data curation in its impact onto the final exhibition.

8. 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 archaeometric knowledge and the storing of the data.

[30]RIDC, https://ridc.okayama-u.ac.jp/english/, visited on 13 September 2021.

[31]UNIARQ, https://www.uniarq.net/, visited on 13 September 2021.

[32]IRIAE, https://membership9.wixsite.com/iriae, visited on 13 September 2021.

[33]http://www.univ-unita.eu/, visited on 13 September 2021

The BeArchaeo project is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie, Grant Agreement No 823826. The content of this paper represents the views of the authors only and is their sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.

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

[1] I. Hodder, *The Archaeological Process: An Introduction.*, Oxford: Blackwell, 1999.

[2] M. Olsson, Making sense of the past: the embodied information practices of field archaeologists, *Journal of Information Science* **42**(3) (2016), 410–419.

[3] G. Artioli, *Scientific Methods and Cultural Heritage: An introduction to the application of materials science to archaeometry and conservation science*, Oxford Scholarship Online, 2010.

[4] M. Reindel and G.A.W. (Eds.) (eds), *New Technologies for Archaeology: Multidisciplinary Investigations in Palpa and Nasca, Peru*, Springer-Verlag, Berlin Heidelberg, 2009.

[5] M.S. Tite, Archaeological Science - Past Achievements and Future Prospects, *Archaeometry* **33**(2) (1991), 139–151. doi:https://doi.org/10.1111/j.1475-4754.1991.tb00695.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4754.1991.tb00695.x.

[6] M. Carver, *Archaeological Investigation*, Routledge, 2009.

[7] A. Berggren, N. Dell'Unto, M. Forte, S. Haddow, I. Hodder, J. Issavi, N. Lercari, C. Mazzucato, A. Mickel and J. Taylor, Revisiting reflexive archaeology at Catalhoyuk: Integrating digital and 3D technologies at the trowel's edge, *Antiquity* **89** (2015), 433–448. doi:10.15184/aqy.2014.43.

[8] E. Yakel, P. Conway, M. Hedstrom and D. Wallace, Digital Curation for Digital Natives, *Journal of Education for Library and Information Science* **52** (2011), 23.

[9] T. Karatas and V. Lombardo, A Multiple Perspective Account of Digital Curation for Cultural Heritage: Tasks, Disciplines and Institutions, in: *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020*, T. Kuflik, I. Torre, R. Burke and C. Gena, eds, ACM, 2020, pp. 325–332. doi:10.1145/3386392.3399277.

[10] B. Nicolescu, Methodology of Transdisciplinarity – Levels of Reality, Logic of the Included Middle and Complexity, *Transdisciplinary Journal of Engineering & Science* **1:1** (2010), 19–38.

[11] F. Silva and M.V. Linden, Amplitude of travelling front as inferred from 14C predicts levels of genetic admixture among European early farmers, *Scientific Reports* **7** (2017).

[12] J. Richards and C. Hardman, Stepping back from the trench edge: an archaeological perspective on the development of standards for recording and publication, in: *The virtual representation of the past*, M. Greengrass and L.H. (Eds.), eds, Farnham: Ashgate, 2008, pp. 101–112.

[13] F. Niccolucci, S. Hermon, and M. Doerr, The Formal Logical Foundations of Archaeological Ontologies, in: *Mathematics and archaeology*, J. Barcelo and I. Bogdanovic, eds, Boca Raton: CRC Press, 2015, pp. 86–99.

[14] K.-H. Lampe, K. Riede and M. Doerr, Research Between Natural and Cultural History Information: Benefits and IT-Requirements for Transdisciplinarity, *ACM Journal on Computing and Cultural Heritage* **1**(1) (2008).

[15] V. Lombardo, R. Damiano, T. Karatas and C. Mattutino, Linking Ontological Classes and Archaeological Forms, in: *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II*, J.Z. Pan, V.A.M. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne and L. Kagal, eds, Lecture Notes in Computer Science, Vol. 12507, Springer, 2020, pp. 700–715. doi:10.1007/978-3-030-62466-8_43.

[16] A.F. Martin Doerr, S. Hermon, G. Hiebel, A. Kritsotaki, A. Masur, K. May, P. Ronzino, W. Schmidle, M. Theodoridou, D. Tsiafaki, E. Christaki, C.-E. Ore et al., Definition of the CRMarchaeo: An Extension of CIDOC CRM to support the archaeological excavation process, Technical Report, Version 1.5.0, Proposal for approval by CIDOC CRM-SIG, February 2020.

[17] M. Doerr, A. Kritsotaki, Y. Rousakis, G. Hiebel, M. Theodoridou et al., Definition of the CRMsci: An Extension of CIDOC-CRM to support scientific observation, Technical Report, Version 1.2.9, Proposal for approval by CIDOC CRM-SIG, February 2021.

[18] C.H. Roosevelt, P. Cobb, E. Moss, B.R. Olson and S. Ünlüsoy, Excavation is digitization: advances in archaeological practice, *Journal of Field Archaeology* **40** (2015), 325–46.

[19] N. Lercari, E. Shiferaw, M. Forte and R. Kopper, Immersive Visualization and Curation of Archaeological Heritage Data: Çatalhöyük and the DigIT App, *Journal of Archaeological Method and Theory* (2017). doi:10.1007/s10816-017-9340-4.

[20] J. Conolly and M.W. Lake, *Geographical Information Systems in Archaeology*, Cambridge University Press, 2006.

[21] P. Reilly, Towards a virtual archaeology, in: *Computer Applications in Archaeology*, K. Lockyear and S. Rahtz, eds, Oxford: BAR 565, 1990, pp. 133–139.

[22] J.A. Barcelo, M. Forte and D.H. Sanders, *Virtual Reality in Archaeology*, Oxford: ArcheoPress, 2000.

[23] L. Richardson, A digital public archaeology?, *Papers from the Institute of Archaeology, London: UCL* (2013).

[24] C. Meghini, R. Scopigno, J. Richards, H. Wright, G. Geser, S. Cuy, J. Fihn, B. Fanini, H. Hollander, F. Niccolucci, A. Felicetti, P. Ronzino, F. Nurra, C. Papatheodorou, D. Gavrilis, M. Theodoridou, M. Doerr, D. Tudhope, C. Binding and A. Vlachidis, ARIADNE: A Research Infrastructure for Archaeology, *Journal of Computing and Cultural Heritage* **10**(3) (2017). doi:10.1145/3064527.

[25] P. Cripps, A. Greenhalgh, D. Fellows, K. May and D. Robinson, Ontological Modelling of the work of the Centre for Archaeology, CIDOC CRM technical paper, Centre for Archaeology, 2004.

[26] A. Costopoulos, Digital archeology is here (and has Been for a While), *Frontiers in Digital Humanities* **3** (2016).

[27] E.C. Kansa and S.W. Kansa, We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication, *Journal of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1) (2013), 88–97.

[28] I. Faniel, E. Kansa and S.W. Kansa, The challenges of digging data: a study of context in archaeological data reuse, in: *Proceedings of 13th ACM/IEEE-CS joint conference on digital libraries, Indianapolis, IN*, New York: ACM, 22–25 July 2013, pp. 295–304.

[29] E. Kansa and S.W. Kansa, Digital Data and Data Literacy in Archaeology Now and in the New Decade, *Advances in Archaeological Practice* **9**(1) (2021), 81–85..

[30] C. Binding, D. Tudhope and A. Vlachidis, A study of semantic integration across archaeological data and reports in different languages, *Journal of Information Science* **45**(3) (2019), 364–386.

[31] M.A. López, R. Tringham and C. Perlingieri, Last House on the Hill: Digitally Remediating Data and Media for Preservation and Access, *Journal on Computing and Cultural Heritage (JOCCH)* **4** (2011), 109–116. doi:10.1145/2050096.2050098.

[32] D. Myers, A. Dalgity and I. Avramides, The Arches heritage inventory and management system: a platform for the heritage field, *Journal of Cultural Heritage Management and Sustainable Development* **6**(2) (2016), 213–224.

[33] L. Pouchard, Revisiting the Data Lifecycle with Big Data Curation, *International Journal of Digital Curation* **10** (2015). doi:10.2218/ijdc.v10i2.342.

[34] S. Higgins, The DCC curation lifecycle model, 2008, p. 453. doi:10.1145/1378889.1378998.

[35] M. Patel, S. Coles, D. Giaretta, S. Rankin and B. McIlwrath, The role of OAIS representation information in the digital curation of crystallography data, 2009, IEEE eScience 2009 ; Conference date: 09-12-2009 Through 11-12-2009. doi:10.1109/e-Science.2009.27.

[36] F. Niccolucci and A. Felicetti, A CIDOC CRM-based Model for the Documentation of Heritage Sciences, in: *Proceedings of the 3rd Digital Heritage International Congress (Digital Heritage) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, San Francisco, USA, 2018, pp. 1–6.

[37] L.N. Stutz, A Future for Archaeology: In Defense of an Intellectually Engaged, Collaborative and Confident Archaeology, *Norwegian Archaeological Review* **51:1-2** (2018), 48–56. doi:10.1080/00293652.2018.1544168.

[38] E.C. Harris, *Principles of Archaeological Stratigraphy*, Academic Press, London, 1989.

[39] M.C. Suárez-Figueroa, A. Gómez-Pérez and M. Fernández-López, *The NeOn Methodology for Ontology Engineering*, in: *Ontology Engineering in a Networked World*, M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta and A. Gangemi, eds, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 9–34. ISBN 978-3-642-24794-1. doi:10.1007/978-3-642-24794-1_2.

[40] V.A. Carriero, A. Gangemi, M.L. Mancinelli, L. Marinucci, A.G. Nuzzolese, V. Presutti and C. Veninata, ArCo ontology network and LOD on Italian Cultural Heritage, in: *ODOCH@CAiSE*, 2019.

[41] Y. Tadanao, *Dictionary of Japanese Archaeological terms*, Tokyo Bijutsu Publishing, Tokyo, 2001.

[42] P.S. Quinn, *Ceramic petrography: The interpretation of archaeological pottery*, Archaeopress, London, 2013.