

An Ontology for Legacy Data on Ancient Ceramics of the Plain of Catania*

Rodolfo Brancato¹, Marianna Nicolosi-Asmundo², Grazia Pagano², Daniele Francesco Santamaria², and Salvatore Uchino²

¹ Department of Human Sciences, University of Catania
email: rodolfobrancato@gmail.com

² Department of Mathematics and Computer Science, University of Catania
email: {nicolosi,santamaria}@dmi.unict.it,
{grazia.pagano89,turirg}@gmail.com

Abstract. Digital representation and organization of legacy data plays a crucial role in the diffusion, use, and understanding of data stored in old publications, archives, and museums. An interesting case study comes from data of potteries discovered in ancient rural territories of Eastern Sicily, as the majority of legacy data for this research area exists in the form of old maps and paper catalogues: to make these datasets available at a global level, innovative digital technologies are needed.

The Semantic Web offers well established methodologies and tools to semantically model application domains and to integrate data, making them global entities available on the Web.

In this contribution, we present *OntoCeramic 2.0*, an OWL 2 (Web Ontology Language 2) ontology storing archaeological data from the plain of Catania regarding ancient potteries, and whose taxonomy refines and extends *OntoCeramic 1.0*, an ontology for the classification of ancient ceramics defined in a previous work by some of the authors. *OntoCeramic 2.0*, constructed according to the standard CIDOC Conceptual Reference Model (CRM), represents and integrates new survey and legacy data on ancient pottery stored in the archives of *Heritage Superintendence* of Syracuse and Catania, in the *Regional Technical Office* of Sicily, and in the *State Archives* of Palermo and Catania.

1 Introduction

Archaeological studies carried out in Sicily for over a century report on forgotten cities, necropolises, monuments, artefact scatters, and other landscape features. Resulting data are still limited in quantity and variable in quality: this problem is particularly prevalent in the countryside. Because of that, legacy data nowadays available for the plain of Catania (Sicily) are of basilar importance for archaeologists. This holds, in particular, if such data are represented and organized in a digital way, globally accessible on-line, and easily verifiable.

* We gratefully acknowledge support by “Università degli Studi di Catania, Piano della ricerca 2016/2018 Linea di intervento 2”.

Currently, a MySQL database containing legacy data on potteries is being developed. A first version of it is described in [5] and is part of the Ru.N.S. project (Rural Networks in Sicily). Analysis of potteries is fundamental for scholars since it helps in providing a clear image of rural population trends in ancient times, and in reconstructing the organization of the agrarian territory in the Hellenistic and Roman ages [18]. The study area considered in this contribution is located in the western portion of the plain of Catania, the area between the Simeto river to the north and the Margi river to the south. With an extension of 540 km², the area forms a perfect case study due to the number of excavations and survey projects carried out by the *Heritage Superintendence* of Catania and the *Chair of Ancient Topography* (University of Catania) over the last few decades. An overview on excavations and survey projects in the area can be found in [4, 21].

Relational databases, however, even though well-assessed tools for organizing and querying information, do not support global data and flexible integration mechanisms with other sources. Moreover, they suffer from limited modelling and reasoning capabilities.

Semantic Web [3] offers methodologies, languages, and tools for knowledge representation systems in which data are published, accessed, and integrated with information from other sources at a global level, thus allowing coherence and dissemination of knowledge. Moreover, the presence of dedicated automated reasoning systems permits to verify the consistency of the model, to query the dataset, and to infer implicit knowledge present both in the taxonomy and in the data. The definition of a specific domain is widely called an ontology.

Recently, capabilities of ontologies have been understood and appreciated by archaeologists [16, 17]. Some projects have been started regarding specific kinds of archaeological finds such as ancient manuscripts [12] and epigraphs [13]. In [9] we presented **OntoCeramic 1.0**,³ an OWL (Web Ontology Language) ontology for cataloguing and classifying ceramics, originated by a synergic effort between computer scientists and archaeologists to address the problem of efficiently automatize the task of correctly cataloguing ceramics and to make such knowledge easily accessible and usable by researchers of the field.

In this contribution, we present **OntoCeramic 2.0**,⁴ an ontology storing data of ancient ceramics discovered on the western side of the plain of Catania in Sicily and collected by the Ru.N.S. project. **OntoCeramic 2.0** models principal features of potteries such as ceramic class, shape, type, dough, and chronological periods of production of the finds. The ontology is completely mapped in the CIDOC Conceptual Reference Model (CRM) [11], the international standard of exchange of cultural heritage knowledge. This makes it flexible and fully embeddable with ontologies conceived for different application domains.

³ <https://github.com/dfsantamaria/OntoCeramic-1.0/blob/master/OntoCeramic1.owl>

⁴ <https://github.com/dfsantamaria/OntoCeramic-2.0/blob/master/OntoCeramic2.owl>

2 Preliminaries

2.1 Semantic Web and Web Ontologies

Semantic Web is a vision of the World Wide Web in which machine-readable data enable software agents to access, extract, integrate, manipulate, and query information on behalf of human agents, and thus to gain a deeper knowledge of the domain. To achieve such goals at global level, information must carry an explicit meaning and must be modelled by appropriate languages endowed with formally defined semantics supporting automated reasoning procedures. For this purpose, the World Wide Web Consortium (W3C) identifies the Web Ontology Language (OWL), a family of knowledge representation languages relying on *Description Logics* (DLs) [2], as the standard for representing ontologies. We recall that an ontology [15, 19] is a formal description of the domain of interest carried out by combining three basic syntactic categories: entities, expressions, and axioms, which constitute the logical part of ontologies, namely what ontologies can express and the type of inferences that can be drawn. Entities represent primitive terms of an ontology and are identified in a unique way. They are individuals (actors), object- and data-properties (actions), and classes (sets of actors with common features).

OWL⁵, currently in version 2.1, is based on the idea of triples, which are ways to connect two individuals or an individual and a data-value. In order to provide a formal description of the domain, OWL 2 triples can be organized in two main categories: axioms and expressions. Axioms are constructed by applying OWL 2 primitives to entities, thus forming complex descriptions, whereas expressions describe what is true in the domain. For example, one can combine an axiom for equivalent classes with an expression of class union such as

$$\textit{EquivalentClasses}(\textit{Late Antiquity}, \textit{UnionOf}(\textit{Ostrogothic Age}, \textit{Late Imperial Age}))$$

to define the class *Late Antiquity* as the period corresponding to the union of the periods represented by the classes *Ostrogothic Age* and *Late Imperial Age*.

OWL 2 admits three main types of expressions: object-property expressions, data-property expressions, and class expressions. Object-property expressions represent binary relationships among individuals, whereas data-property expressions represent binary relationships among individuals and data type values. Class expressions represent sets of individuals sharing common characteristics. Such individuals are said to be instances of the respective class expressions. Class expressions are constructed recursively by using classes, properties, and class expressions, and by applying restrictions on property expressions. For a detailed explanation of axioms and expressions introduced in OWL 2, the reader is referred to [1, 14].

⁵ <https://www.w3.org/TR/owl2-overview/>.

2.2 OntoCeramic 1.0

OntoCeramic 1.0 is an OWL 2 ontology presented in [9] for cataloguing and classifying ancient potteries, designed with the purpose of efficiently addressing significant problems concerning knowledge management about potteries such as the classification by shape, type, and class, and the analysis of finds by their components and discovery places. The ontology has been designed on ICCD⁶ (Istituto Centrale per il Catalogo e la Documentazione) data sheets taking into account relevant papers in the field [10, 16]. It contains more than 90 classes, 33 object-properties, 20 data-properties, and 13 SWRL rules permitting several reasoning tasks on the knowledge base in a short time. The expressive power of the language underlying Ontoceramic 1.0 has been studied in [6, 22]. In particular, in [22], an OWL 2 profile representing OntoCeramic 1.0 has been constructed from a decidable fragment of set theory, and it has been proved that the computational complexity of the consistency problem for its knowledge bases is NP-complete.

2.3 CIDOC CRM

The CIDOC Conceptual Reference Model (CRM) is the international standard for the controlled exchange of cultural heritage information since 2006. It provides a general specification which can be adopted in any cultural heritage context to construct a Semantic Web-based information system, to serve as a guide for good practices of conceptual modelling, and to improve information sharing. There are several institutions that successfully implement CIDOC such as galleries, libraries, museums, archives, as well as any other cultural environment based on cultural heritage data that publishes and shares its information in the Semantic Web formats. The CIDOC core covers several general aspects of cultural information, such as material and immaterial entities, events, space, and time. Such general concepts can be specialized, contextualized, and integrated in order to address practical aspects of cultural heritage issues. It models several notions, such as participation, appellation, parthood and structure, material and immaterial stuffs, location, assessment and identification, motivation, and so on.

3 The Ontology OntoCeramic 2.0

In this section, we illustrate the ontology OntoCeramic 2.0 which refines and extends OntoCeramic 1.0 to model and reason on the survey and legacy data of the plain of Catania collected in the ambit of the Ru.N.S. project. Specifically, in OntoCeramic 2.0 we refined the definition and enriched the classification of (a) fabric and pottery types (also called ceramic class and type, respectively), which help in determining the production site of archaeological finds, (b) the shape of the find, which helps in determining the pottery type, and (c) the pottery sizes, usually determined by measuring the external diameter of the rim in millimeters.

⁶ <http://www.iccd.beniculturali.it>

In *OntoCeramic 2.0* we also consider further features: (d) the pottery chronological context with respect to the Sicilian historical periods, and (e) populating the ontology with data collected from the Ru.N.S. *Finds catalog* dataset. The latter is an Excel file consisting of 4384 rows, each containing basilar information on the archaeological finds discovered during the recognition phase of the considered territories (western edges of the plain of Catania), namely identification code, class, shape, type, conservation state, dimensions, and free-standing text descriptions. The task of populating the ontology has been performed exploiting the Protégé plug-in Cellfie,⁷ which allows one to parse Excel files and to map Excel entries to OWL triples.

OntoCeramic 2.0 has been defined according to the standard CIDOC CRM and it uses the *LinkedGeoData*⁸ ontology for describing locations and for identifying the discovery place of finds. It consists of more than 220 classes, 40 object-properties, 20 data-properties, and 9000 individuals, excluding entities imported by CIDOC CRM and *LinkedGeoData*.

The rest of this section is devoted to the description of the ontology *OntoCeramic 2.0*. We first list the main classes of the ontology and their characteristics, and then describe the general structure of the taxonomy.

- *Archaeological_Find*: this class collects individuals representing archaeological finds. It is defined as subclass of the CIDOC class *E22 Man-Made Object*.
- *Ceramic_Class*: is the root of a class hierarchy describing ceramic classes to which a find may belong to.
- *Facies*: is a subclass of *Ceramic_Class* of particular interest because it models, together with its subclasses, all the ceramic classes in the Sicilian context.
- *Shape*: is defined as a subclass of the CIDOC class *E26 Physical Feature* and describes the shape of finds. One of its notable subclasses is the class *Undistinguished_Shape*, introduced to model ambiguous shapes.
- *Archaeological_Type*: describes the type of finds by specifying the shape and, when available, the class. Among its subclasses, a relevant one is the class *Undistinguished_Type* modelling ambiguous types.
- *Decoration*: is defined as a subclass of the CIDOC class *E26 Physical-Features* and describes the decoration of archaeological finds.
- *Description*: contains a free-text description concerning finds and is a subclass of the CIDOC class *E73 Information Object*.
- *Dimension*: defines the size of finds and is a subclass of the CIDOC class *E54 Dimension*.
- *Dough*: describes the elements used to compose the dough of finds and is a subclass of the CIDOC class *E26 Physical-Features*.
- *Conservation_State*: reports on the physical conditions of finds at their discovery time and is defined as a subclass of the CIDOC class *E14 Condition_State*.
- *Sicilian_Period*: is the root of a class hierarchy that models Sicilian historical periods and is defined as a subclass of the CIDOC class *E4 Period*.

⁷ <https://github.com/protegeproject/cellfie-plugin>

⁸ <http://linkedgeo.org/>

The classes *UndistinguishedShape* and *UndistinguishedType* have been introduced to represent finds that have not clear shapes and types, respectively. Fig. 1 partially illustrates the class hierarchy with root *UndistinguishedShape* and shows how to model the shape of an archaeological find when one is uncertain on the fact that it has the shape of a mortar or of a basin. The *Mortar-Basin* class contains individuals that may belong either to the class *Mortar* or to the class *Basin*. The object-property *identifiedAs* relates such “hybrid” individuals with instances of the class *Mortar* or with instances of the class *Basin*, which clearly define different shape classes.

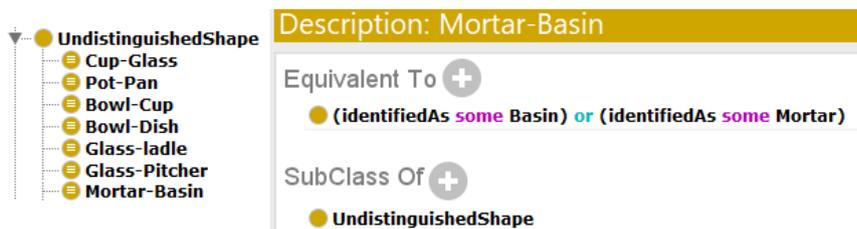


Fig. 1. Modelling of an uncertain shape in OntoCeramic 2.0.

The core of OntoCeramic 2.0 is depicted in Fig. 2. In bold we introduce classes (resp., properties) specifically defined for OntoCeramic 2.0, immediately below them we report the corresponding superclasses (resp., superproperties) from CIDOC CRM.

Instances of the class *Archaeological Finds* are linked to their types, shapes, and classes, by means of the object-properties *hasArchaeologicalType* (subproperty of *P41i was classified by*), *hasShape* (subproperty of *P56 bears features of*), and *hasClass*, respectively.

The class *ArchaeologicalType* is associated to the classes *Ceramic-Class* and *Shape* by means of the object-properties *specifiedByClass* and *specifiedByShape* (subproperty of *P41 classified*), respectively.

We have separated the notion of shape from the notion of functionality of finds, i.e., the usage finds have been originally intended for. For example, an archaeological find may have the shape of a basin and the functionality of an holy water font. The notion of functionality is defined by exploiting the class *Functionality*.

Archaeological finds are related with their functionality by means of the object-property *hasFunctionality*, with their conservation state by means of the object-property *hasConservationState* (subproperty of *P34i was assessed by*), and with related free-text descriptions by means of the object-property *hasDescription* (subproperty of the CIDOC relation *P128 carries*).

Moreover, finds are related with their dimensions, modelled as instances of the class *Dimension*, by means of the object-property *has_dimension* (subproperty of *P43 has dimension*). Since dimensions of finds can be irregular and mea-

surement errors may occur, we introduce two subclasses of *Dimension*, the classes *Max.Dimension* and *Min.Dimension*. The object-property *has.value* (subproperty of the CIDOC relation *P90 has value*) relates each dimension with its value, represented by a double. Finally, finds are related with fragments composing them by means of the object-property *formed by* (subproperty of the CIDOC property *P46 is composed of*).

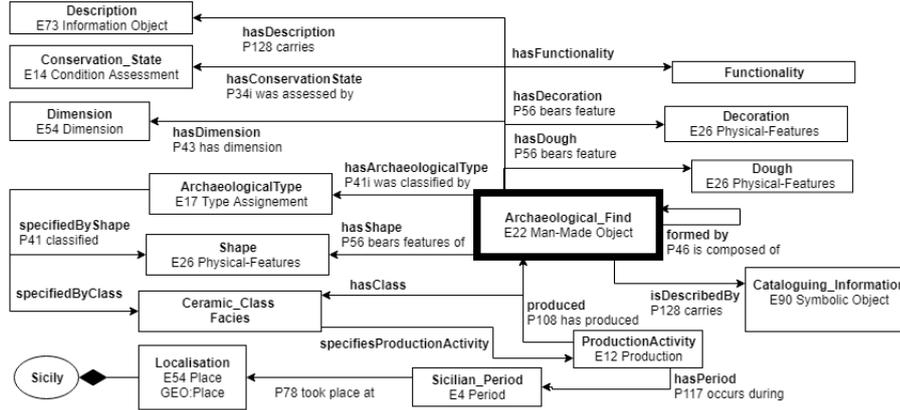


Fig. 2. The main structure of OntoCeramic 2.0.

As mentioned above, OntoCeramic 2.0 is endowed with an accurate chronological modelling of the historical periods concerning the production activity of archaeological finds in the Sicilian territories. Principal historical periods are represented by means of a hierarchy of classes having as root the class *Sicilian_Period* and whose instances are related with the individual *Sicily*, (instance of the class *Localisation*) by means of the CIDOC property *P78 took place at*. The class *Localisation* is defined as a subclass of the *LinkedGeoData* class *Place* and of the CIDOC class *E54 Place*. The data-properties *start_date* and *end_date* link each period with its start and end dates, respectively. Each period is described by means of an OWL expression representing the time interval between its start and end dates. For instance, the period *Sicilian_Iron_Age* is defined as the Sicilian period ranging from the year -900 to the year -476 (in absolute value) and contains as instances the sub-periods *sicilian_iron_age_1*, starting in -900 and ending in -734, and *sicilian_iron_age_2*, starting in -733 and ending in -476 (see Fig. 3). Such definitions force DL reasoners to place individuals representing specific sub-periods in the correct subclass of *Sicilian_Period*. This is useful when one wishes to relate historical periods of different regions of the world. Historical periods, indeed, vary from a region to another, since social-cultural and environmental phenomena arise in different moments. For example, the Late Bronze Age in Malta, starting in -700 and ending in -500, occurs dur-

ing the Iron Age in Sicily. As Fig. 3 shows, this fact is correctly deduced by the Pellet DL reasoner that places the individual *malta_late_bronze_age*, modelling the Late Bronze Age in Malta, in the class *Sicilian_Iron_Age*.

Description: Sicilian_Iron_Age		Description: malta_late_bronze_age	
Equivalent To +	Instances +	Data property assertions +	
<ul style="list-style-type: none"> ● Period and (end_date some xsd:int[<="^-476"^^xsd:int]) and (start_date some xsd:int[>="^-900"^^xsd:int]) 	<ul style="list-style-type: none"> ◆ malta_late_bronze_age ◆ sicilian_iron_age_1 ◆ sicilian_iron_age_2 	<ul style="list-style-type: none"> ■ start_date "-700"^^xsd:int ■ end_date "-500"^^xsd:int 	

Fig. 3. Definition of Iron Age.

Determining the ceramic class of finds not only helps in correctly dating them, but also in reconstructing the chronological information of the archaeological context [20]. For instance, the shapes of the rim and of the body of Greek black burnished wares are good chronological markers of the production activity. Hence, the task of reasoning on the relationships among ceramic classes, archaeological finds, and historical periods turns out to be crucial to recognize the production activity and to collocate finds in the correct chronological context. The production activity is modelled in *OntoCeramic 2.0* by means of the class *ProductionActivity* (subclass of the CIDOC class *E12 Production*). Ceramic classes and facies are related with production activities by means of the object-property *specifiesProductionActivity*. Finally, finds are related with instances of the class *ProductionActivity* by means of the object-property *produced* (subproperty of the CIDOC property *P108 has produced*).

3.1 Conclusions

We presented *OntoCeramic 2.0*, an OWL 2 ontology storing new survey and legacy data on ancient pottery stored in the archives of *Heritage Superintendence* of Syracuse and Catania, in the *Regional Technical Office* of Sicily, and in the *State Archives* of Palermo and Catania, and collected within the Ru.N.S. project.

We integrated *OntoCeramic 2.0* in the standard CIDOC CRM and defined important features of ceramics such as class, shape, type, dough, and chronological periods of archaeological finds. We plan to extend *OntoCeramic 2.0* in such a way as to support stratigraphic excavations, production factories, topographical information, and bibliographic references management. In addition, we consider to integrate *OntoCeramic 2.0* with data from the Eastern side of the Sicily and with ontologies for other types archaeological finds.

Finally, we intend to define a set-theoretic representation of *OntoCeramic 2.0* in the flavour of [6]. However, since *OntoCeramic 2.0* contains existential restrictions, we also need to modify the underlying set-theoretic fragment in such a way as to allow a restricted form of the composition operator. The related reasoning procedure will then be adapted to the new set-theoretic fragments exploiting the techniques introduced in [7, 8] in the area of relational dual tableaux.

References

1. D. Allemang and Hendler J. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. *Elsevier*, 2011.
2. F. Baader, I. Horrocks, C. Lutz, and U. Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.
3. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
4. E. Bonacini. *Il Territorio Calatino nella Sicilia Imperiale e Tardoromana*. British Archaeological Reports British Series, Oxford, 2007.
5. R. Brancato. Profilo Topografico dei Paesaggi Rurali della Piana di Catania. In *forthcoming, Ph.D. Thesis, University of Catania*, 2019.
6. D. Cantone, C. Longo, M. Nicolosi-Asmundo, and D. F. Santamaria. Web Ontology Representation and Reasoning via Fragments of Set Theory. In *Cate, B. and Mileo, A. (eds) Web Reasoning and Rule Systems. LNCS, vol. 9209. Springer*, 2015.
7. D. Cantone, M. Nicolosi-Asmundo, and E. Orłowska. Dual tableau-based decision procedures for some relational logics. In *Proceedings of the 25th Italian Conference on Computational Logic, CEUR-WS Vol. 598, Rende, Italy, July 7-9, 2010*, 2010.
8. D. Cantone, M. Nicolosi-Asmundo, and E. Orłowska. Dual tableau-based decision procedures for relational logics with restricted composition operator. *Journal of Applied Non-Classical Logics*, 21(2):177–200, 2011.
9. D. Cantone, M. Nicolosi-Asmundo, D. F. Santamaria, and F. Trapani. Ontoceramic: an OWL Ontology for Ceramics Classification. In *Proceedings of CILC 2015, CEUR-WS, vol. 1459, pp. 122–127, Genova, July 1-3, 2015*.
10. L. Corti. *I Beni Culturali E La Loro Catalogazione*. Mondadori, Milano, 2003.
11. M. Doerr. The CIDOC CRM An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine, Vol.24, N. 3, 75-92*, 2003.
12. A. Felicetti and F. Murano. Scripta Manent: a CIDOC CRM Semiotic Reading of Ancient Texts. *International Journal on Digital Libraries*, 18(4):263–270, 2017.
13. A. Felicetti, F. Murano, P. Ronzino, and F. Niccolucci. CIDOC CRM and Epigraphy: a Hermeneutic Challenge. In *Proc. of the Workshop on Extending, Mapping and Focusing the CRM, Poznań, Poland, September 17*, pages 55–68, 2015.
14. A. Grigoris, P. Groth, F. van Harmelen, and R. Hoekstra. *A Semantic Web Primer, Third Edition*. The MIT Press, 2012.
15. T. Hofweber. *Logic and Ontology*. Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), 2018.
16. A. La Fragola. *L’Atlante delle Forme Ceramiche dell’Enciclopedia dell’Arte Antica: Ipotesi di Progetto per una Gestione e Fruizione Elettronica dei Dati in Ambiente XML*. Bollettino d’Informazioni, Centro di Ricerche per i Beni Culturali, XII, n. 1, Pisa, pages 113–119, 2002.
17. R. Letricot and A.V. Szabados. L’ontologie CIDOC CRM appliquée aux objets du patrimoine antique. In *Archeologia e Calcolatori, sup. 5, pages 257–272*, 2014.
18. M. Mazza. *L’Economia Siciliana tra Impero e Tardo-Impero*. in *Contributi per una Storia economica della Sicilia, Regione Siciliana*, pages 15–62, 1987.
19. D. Oberle, N. Guarino, and S. Staab. *What is an ontology? Handbook on Ontologies*. Springer, 2009.
20. C. Orton, P. Tyers, and A. Vince. *Pottery in Archaeology*. Cambridge, 1993.
21. F. Privitera. *Dall’Alcantara agli Iblei: la Ricerca Archeologica in Provincia di Catania*. Spigo, U. (eds), Regione Siciliana, 2005.
22. D. F. Santamaria. *A Set-Theoretical Representation for OWL 2 Profiles*. LAP Lambert Academic Publishing, ISBN 978-3-659-68797-6, 2015.