CIDOC CRM-based modeling of archaeological catalogue data

Aline Deicke[1]

[1]Academy of Sciences and Literature | Mainz, Digital Academy, Mainz, Germany
Aline.Deicke@adwmainz.de

Over the last decades, the CIDOC Conceptual Reference Model has become the internationally recognized standard for modeling data from the field of cultural heritage. With the growing digitalization of the humanities, the scope of use cases for the CRM is being steadily extended to topics beyond museum knowledge. Among those is the digital publication of archaeological catalogues that focus on the description of sites, associated finds, objects, and the relations between them. This paper presents an exemplary modeling of a standard archaeological catalogue object and offers an outlook on the benefits such a model provides for researchers.

Keywords: CIDOC CRM · data modeling · digital humanities · digital archaeology

1 Introduction

Since its inception in 1996 by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM), the CIDOC Conceptual Rreference Model (CRM) has become the internationally recognized standard for the description of concepts from the field of cultural heritage. The implementation as an ISO-Standard (ISO 21127) in 2006 confirmed the status of the CRM as the essential ontology for all disciplines dealing with material culture. Through its strong semantics and its object-oriented approach it is independent of specific technologies and facilitates sharing and exchange of data [1], [3].

The overall scope of the CRM is defined as the »curated knowledge of museums«, but it is intended to include all forms of academic research and field work [3]. Implementations such as the Centre for Archaeology's CRM-EH have made use of its possibilities to create extensive models of their own domains [1,2].

Project-specific modelings such as these are especially important in cases where data is intended to be subsequently re-used by other projects and researchers, so that there can be a clear understanding of the modeled entities themselves and of any biases implicit in the database structure. One important use case in this context concerns the growing importance of Linked Open Data and the Semantic Web,

especially in archaeology. Projects like ARIADNE[1], Pelagios Commons[2], Perseus[3] or Arachne[4] have all contributed to the growing number of resources linked in the part of the Linked Data Cloud sometimes referred to as the Graph of Ancient World Data [1], [4], [6,7]. This potential for comprehensive analysis beyond the context of one specific project can only be unlocked if the underlying data is well structured according to a commonly used ontology.

1.1 Archaeological catalogues

Over the last decades, more and more of archaeological research and its publication have moved into the digital realm. While most publications still take the form of text, some projects have begun to make raw data and databases available online.

One of the most common forms of archaeological publication is the catalogue. It lists, describes and classifies finds from a certain context that can range from depositions like a single hoard to whole regions and countries. Depending on the objective of the author, the catalogue can be a detailed listing of objects and their properties, for example an excavation report, or it can give further typological, chronological or bibliographic information on its subjects, either as a stand-alone effort or as the basis of an independent research project. This data is especially interesting for subsequent use in projects utilizing digital research methods, e.g. network analysis, or as an addition to the Linked Data Cloud as described above.

The model presented in this article has been developed for a graph database. It serves as the foundation of the author's dissertation about the genesis of an elite identity in the last period of the Late Bronze Age as reflected by rich burials in the area of the so-called Urnfield Cultures. At the same time, it has been structured in a way that also takes into account a more generic understanding of catalogue editions as collections of archaeological finds, subsumed by context, of a certain period over an associated geographical area.

The benefits of a time-costly and complex modeling process [1], [6] such as this might not be evident at first, especially to independent researchers or small project teams with no ambition to further disseminate their raw research data. Yet the use of a well-known ontology and the subsequent publication of well-structured data facilitate communication not only about the data itself but also about research results, and opens the latter up to well-informed discussions. Also, modeling the data basis on which a study rests can lead to a deeper structural understanding of the study subject. Depending on the complexity of the source material, it can take on characteristics of an exploratory process that can result in new ideas and perspectives [5]. Furthermore, while data modeling – as all processes of categorization and interpretation – implies a

[1] http://ariadne-infrastructure.eu/ (Accessed 2016-12-01).
[2] http://commons.pelagios.org/ (Accessed 2016-11-07).
[2] http://commons.pelagios.org/ (Accessed 2016-11-07).
[3] http://www.perseus.tufts.edu/hopper/ (Accessed 2016-11-07).
[4] http://arachne.uni-koeln.de/drupal/ (Accessed 2016-11-07).

certain degree of generalization, ideally it does so in a systematic way that exposes these biases and makes them explicit and verifiable. This last point is especially important in a field like archaeology whose source material is inherently characterized by gaps that might not always be easily recognizable. Careful ontological modeling can help to identify these missing pieces and as such strengthens inferences drawn from the data [1].

2 Exemplary data model for archaeological catalogue data

The provision and presentation of catalogue data has a different focus than the documentation of museum information or the processing of excavation results. The emphasis on *events* as evident in the CRM [3] shifts towards *contexts* as established by the discovery and the academic analysis of finds. The presented model shows an ongoing effort to capture these contexts and the semantic structure of an archaeological catalogue listing detailed information about elite graves of the Late Urnfield Culture. As such, figure 1 concentrates on the core aspect of a closed find and its associations and presents several related concepts such as the detailed description of specific timespans or the activities surrounding an excavation in an abridged version.

The central object of the model is the closed context of the find itself, be it a grave, a hoard or a single find. It is assigned the class *E19 Physical Object*, which includes «all aggregates of objects made for functional purposes of whatever kind» [3]. While it itself can be part of a larger context (*E27 Site*), it can also contain several types of objects, namely *E20 Biological Object* such as animal bones or remains of organic materials, *E21 Person* comprised of *E20* such as a skeleton or cremated bones, *E25 Man-Made Feature* detailing e.g. grave architecture, and *E22 Man-Made Object* which stands for the single objects composing the find.

Several of the classes connected to *E22* can also be used to describe properties of *E20*, *E21* or *E25*, like *E57 Material* or *E3 Condition State*. Specific mostly to *E22* is the construct of typological classification built from several instances of *E55 Type* connected through *P127 has broader type*, and *E83 Type Creation*.

This last class is one of the components of the model that refers to the scientific analysis of the material and its embedment into the broader archaeological literature. Another example of this is presented in the structure around *E4 period* that aims less to connect the object to a specific timespan, than to set it in relation to the intricacies that can make up archaeological relative dating systems. Some cases that are included in figure 1 are periods that are specific to regional groups (*E4 > P7 took place at > E53 Place*) or periods that are known under different names in different regions (*E4 > P114 is equal in time to > E4*). Each *E4* can be identified by its relative term *E49 Time Appellation* (e.g. Ha B3) as well as absolute dates *E50 Date*. Given the equal standing of absolute and relative chronology in archaeology, the at times tenuous connection between the two and the problems absolute dating methods such as radiocarbon dating can have in determining the exact age of objects, *E50 as* a

subclass of *E49 Time Appellation* is used instead of *E52 Time-Span*[5] to emphasize the subjectivity inherent in assigning both relative as well as absolute dates.

Both appellations, *E49* as well as *E50*, should be used only in conjunction with the pattern *P70 documents < E31* Doc*ument*, which places the appellations in the context of the corresponding literature. Indeed, due to the nature of the database as a tool for and documentation of a research process, this construct can and should be used for almost every class represented in the model. This ensures that concepts which have been defined by more than one author, for example typo-chronological assignations, can be recognized in the intended meaning, and as such guarantees the study's scientific viability. Other generally applicable classes that give further details and context are *P3 has note > E62 String*, and *E42 Identifier*.

Several other areas of the model deal with the construction of geographic information or the context of discorvery, e.g. excavation, as can be grasped from figure 1.

3 Conclusions and Outlook

The increasing digitalization of archaeology, especially the growing importance of linked data and semantic web technologies, demands data that is well structured, documented and easy to share. At the same time, data modeling as a process can lead to a deeper structural understanding of the study subject, expose gaps and biases in the data and results in new ideas and perspectives.

The CIDOC CRM provides the common ontology that enables research projects as well as independent researchers to engage in such a process to meet the mentioned criteria. Many efforts have already been undertaken [2], [4], yet further discussions about modeling processes are necessary to cover the diverse and varied areas of archaeological research.

The paper presents an ongoing approach at modeling one of these areas, namely objects described in a standard archaeological catalogue, to facilitate discussion about and sharing of research data. It proposes a way of modeling several key concepts common to archaeological research, chief among them the bibliographical documentation of implicit assumptions, and is intended to serve as a starting point for researchers dealing with archaeological catalogue data to create models for their respective use cases.

Further efforts to refine the model will include the positioning of objects (*E19*, *E21*, *E20*, *E22, E25*) in relation to each other, a closer look at more complicated patterns of deposition, the possibility of assigning cultural affiliations to entities and that of documenting the reasoning behind decisions about chronological or typological classifications.

[5] This class is used elsewhere in the model in conjunction with *E7 activity*, which describes a recent and therefore dated event, e.g. an excavation.

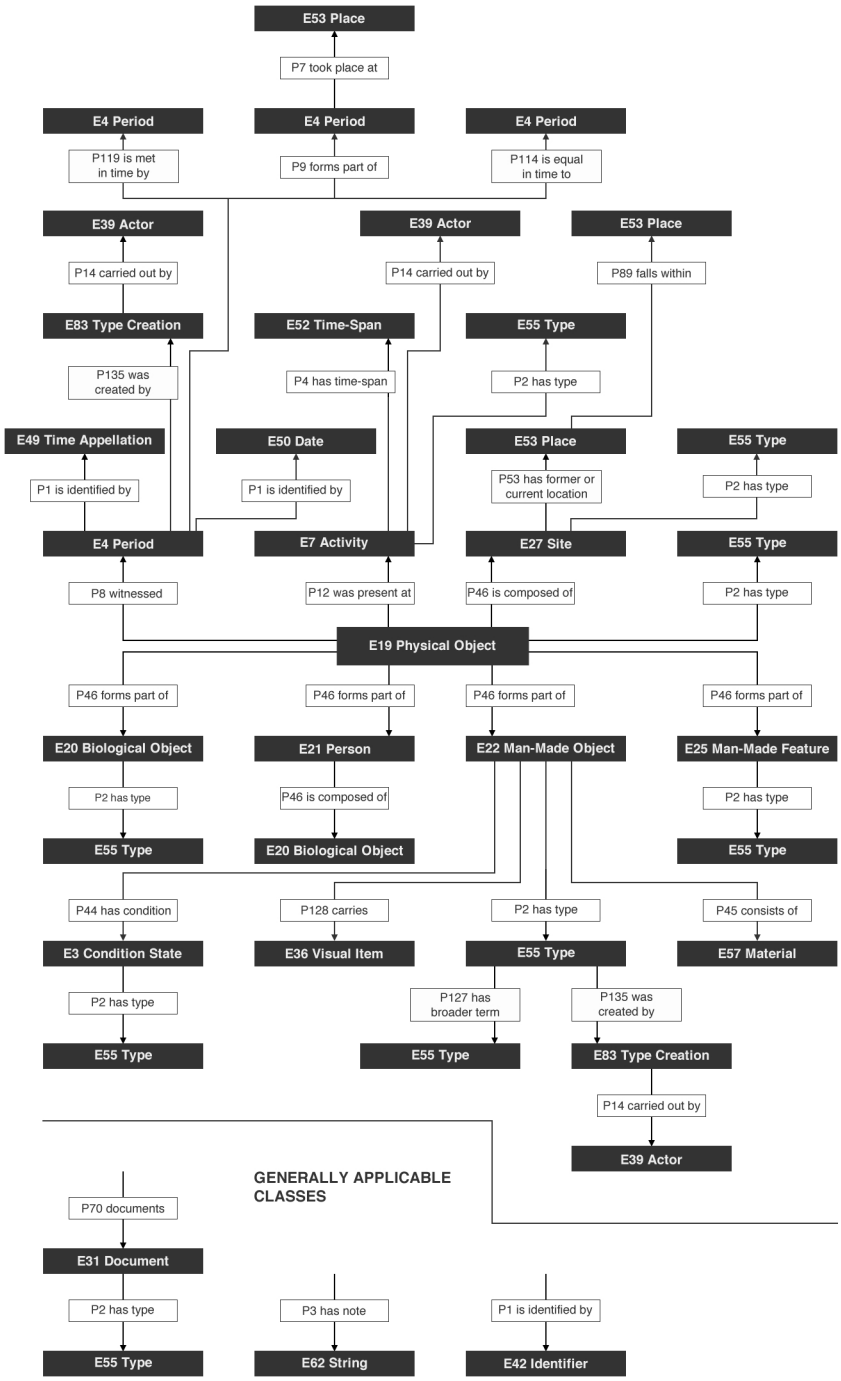

Fig. 1: CIDOC CRM-based data model of a catalogue object

4 References

1. Cripps, P., Greenhalgh, A., Fellows, D., May, K., Robinson, D.: Ontological Modelling of the Work of the Centre for Archaeology. Centre for Archaeology, English Heritage (2004)
2. crmeh. For people interested in Conceptual Reference Modelling and Ontologies for the broader Cultural Heritage domain, `https://crmeh.wordpress.com/`
3. Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (eds.): Definition of the CIDOC Conceptual Reference Model. CIDOC (2011)
4. Erlangen CRM / OWL, `http://erlangen-crm.org/`
5. Flanders, J., Jannidis, F., Knowledge Organization and Data Modeling in the Humanities (2015)
6. Isaksen, L., Simon, R., Barker, E. T. E., de Soto Canamares, P.: Pelagios and the emerging graph of ancient world data. In: WebSci '14. Proceedings of the 2014 ACM conference on Web science, pp. 197–201. ACM, New York (2014)
7. Robineau, R.: Graph of Ancient World Data (2012)