Comparing and mapping archaeological excavation data from different recording systems for integration using ontologies

Anja MASUR¹ | Keith MAY² | Gerald HIEBEL³ | Edeltraud ASPÖCK¹

¹ Austrian Academy of Sciences, Institute for Oriental and European Archaeology, Vienna, Austria | ² English Heritage, Capacity Building Team, Heritage Protection Department, Portsmouth, England | ³ Center for Cultural Informatics, Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Greece

Abstract: Sharing archaeological data across national borders and between previously unconnected systems is a topic of increasing importance. Infrastructures such as ARIADNE aim to provide services that support sharing of archaeological research data. Ontologies such as the CIDOC CRM are an appropriate instrument to harmonize different data structures and thereby support data exchange. Before integrating data by mapping to ontologies it is crucial to establish where the shared meaning of the data lies and to understand the methodology used to record the data. As the largest proportion of archaeological data are derived from excavations or field investigations the initial focus falls on the documentation of these “raw data”. But documentation often varies depending on country-specific guidelines, different excavation methods and technologies, project management requirements, budget, etc. Therefore an analysis of the different recording forms should prove helpful to identify the common meanings of concepts and terms used in archaeological fieldwork.

This paper will show first results of research based on the collection of excavation report forms and manuals from different countries which cover a range of fieldwork methodologies (e.g. single context recording, palaeolithic excavations, etc.). The aim is to analyse and compare the different methodologies, the archaeological concepts involved and the data records, perhaps for the first time on an international level.

We want to discuss the challenges of integrating different concepts, terms and vocabularies, often in different languages, and whether problems with integrating such archaeological data could be addressed by additional archaeological extensions to the CIDOC CRM.

Keywords: archaeological recording systems, ontologies, CIDOC CRM, CRMarchaeo

Introduction

Sharing of data has become one of the mainstays of our digital world. To a certain degree this also applies to archaeological data. Online platforms such as Europeana, services like ADS – Archaeology Data Service [ADS 2014], Fasti Online [FASTI ONLINE 2013] and Open Context [OPEN CONTEXT 2013] provide information and access to international archaeological data and databases. The emerging infrastructure ARIADNE – Advanced Research Infrastructure for Archaeological Dataset Networking in Europe [ARIADNE 2014] hopes to deliver an integration of European online platforms. Mostly these services contain metadata about archaeological datasets, hence information about archaeological sites, research, objects etc.
However, there is a big gap concerning primary data from excavation. Excavation data are the cornerstone of any archaeological research. Excavations investigate the remains of past human activities. But they also destroy them or, as [LUCAS 2001] sees it, at least change these remains. Therefore focus has to lie on the documentation of all activities carried out as part of an archaeological excavation and on the preservation of this information.

There may be many reasons for the lack of online access of excavation data. Certainly one reason is that there exists a great variety of primary data resulting from different excavation methods and different documentation systems which can be unique for a site or specific to a country. Furthermore this problem is amplified by the use of different terminology, languages and meanings.

For putting excavation data from different sites together for joint analysis and comparison we see ontologies as an appropriate instrument. Ontologies provide a conceptual framework permitting a data exchange of heterogeneous sources. In this paper we would like to show how ontologies help to share semantically heterogeneous primary data from archaeological excavations without loss of meaning. We will focus on CIDOC CRM Conceptual Reference Model [CIDOC 2013] and its recently developed extension CRMarchaeo [DOERR et al. 2013]. The basis of our discussion will be excavation recording pro-forma sheets as representations of underlying recording systems and methodologies. We collected and compared published and unpublished excavation sheets from different countries representing different excavation methods, official guidelines and databases. We identified similarities and differences in both methods and vocabularies. Using some specific examples we aim to show how CIDOC CRM can support integrating them. The research took place as part of an ARIADNE working group on “Context, stratigraphic unit, excavated matter”.

**Understanding similarities and differences in excavation data**

As Roskams states, an excavation aims

> “to split the site into its component parts – its stratigraphic units, however defined – and then remove them in the reverse order to which they were deposited, recording their physical, spatial and stratigraphic properties in the process and collecting finds from them to agreed sampling policies as one proceeds” [ROSKAMS 2001, 110].

It is assumed that every excavation follows these principles. Excavation sheets are used to document this process. There exists a variety of ways of documenting excavations and excavation activities. Integrating them using semantic technologies and perhaps publishing them as linked data poses challenges. Some countries have standardized documentation sheets and in many countries particular recording methodologies have become a de-facto standard, e.g. the Museum of London “single context recording system” [MoLAS 1994]. In recent years, guidelines and standards for archaeological practice have been

---

1 Excavation in spits, which means removing soil in clearly defined steps e.g. removing an arbitrary “layer” 20cm thick is seldom used nowadays (BIEL & KلونK 1994). Exceptions may be specific situations, perhaps where a stratigraphic unit is of a thickness above-average and the excavator wants to divide this into more manageable sub-units.
published in many European countries. For example, in Austria guidelines for archaeological excavations were released in 2012 BDA Richtlinien für archäologische Massnahmen [BDA 2012]. Hence, excavation documentation sheets usually vary according to country. But there are also recording sheets which were adapted to particular site requirements or types of sites. Excavation sheets for e.g. underwater sites differ from those used to document Palaeolithic sites such as caves.

In our survey we analysed the commonalities (and differences) of excavation sheets from Great Britain, Austria, Germany and Israel (fig. 1 and 2). A series of recording forms from several other countries were screened too. We asked the following questions: (1) Are there any commonalities regarding entry fields of recording sheets? (2) If yes, what are the main terms used and how do they differ?

**Commonalities and differences: place, stratigraphic unit and finds**

At first glance most of the recording sheets seem to be different (not only in layout). But, we could recognize three conceptual elements that show up on every sheet: spatial information, information about the stratigraphic units and about the associated finds.

Every excavation is located at a certain place, or investigates a certain site, and thus every sheet deals with information about the examined geo-locational or spatial area. Some sheets only specify the name of the site and as a rule a project code that identifies the site unambiguously. Others provide detailed information about the location of the place as e.g. municipality, district, federal state. But there is also different information about the site itself depending on the kind of excavation. As research can rarely cover the whole site often a specific section is chosen for excavation. Furthermore these are divided into artificial segments that are represented on the sheets as fields like grid number, number of the square meter (e.g. at Palaeolithic excavations) or trench, part of a feature (e.g. a big burial mound) etc. So we have to handle different types of spatial information with different levels of detail.

The same problem occurs when looking on the ways of how stratigraphic units are described. Stratigraphic units are any part of an excavation that is “formed by a process of stratified deposition and removal” [MOLAS 1994, p.7] and is delimited from other elements by discrete boundaries. According to this definition a stratigraphic unit can have numerous forms and it is of no surprise that a huge number of possibilities exist for characterizing it. First and foremost it depends on whether a positive feature like a hearth, the filling of a pit, some debris, or a negative feature like a post-hole, a pit etc. is recorded. Negative stratigraphic units or interfaces are often described by their shape in plan or profile, the shape of their corners, base or sides, positive ones, such as deposits or structures, additionally by colour, texture, soil type, composition and inclusions. Often this is complemented by measurements like length, width, depth etc. Although the different sheets have many entry fields in common, they differ depending on the kind of excavation and the methods used. Many recording sheets try to be as general as possible to keep all options open regarding the characterizing of a feature. This particularly applies to recording forms that refer to country-specific
guidelines or other standards. By that institutions want to make sure that all kinds of features can be documented and can be compared in future analyses.

According to the law of superposition, derived from natural history and geology, *layers are deposited in a time sequence, with the oldest on the bottom and the youngest on the top* [HARRIS 1989]. This law states the relative sequence of layers and thereby enables archaeologists to apply a relative chronology to the associated layers. Thus a layer, or more precisely, a stratigraphic unit (using our preferred term), is a fundamental concept used in archaeological excavation. As a logical consequence statements about the different relationships between Stratigraphic Units are part of the recording and thus are documented on the excavation sheets. These relationships might include if e.g. one stratigraphic unit cuts or fills another one or if one is above or below another one and thus give the temporal sequence of their formation. Fields describing the grouping of stratigraphic units, e.g. post-holes belonging to one house, are often part of the excavation sheets. But such relationships may only be recognised, or fully understood, later in the post-excavation analysis of all the data from a site (e.g. after an analysis of samples, or other specialist reports).

Another common concept we identify by examining the sheets are *finds*. Finds are items deposited in the past by someone or some process. Small or bulk finds may be recorded. Finds often are not recorded by a single sheet but on an extra list or else are included in the descriptions of the contents of features. In the case of Palaeolithic excavations documenting finds separately plays a huge role, e.g. position, orientation, condition etc. are recorded.

**Discussion**

As described above excavation recording sheets share some common conceptual elements. But we found many possibilities for how these are described. Problematic for comparing data are the differences of meanings (semantics) represented by key terms. This becomes apparent for example when we look at all the different possibilities for expressing the idea of a 'stratigraphic unit'. On the excavation sheets the following terms appeared: layer, context, locus and feature. More possibilities can be found elsewhere [PAVEL 2010]. If data described by these different terms were to be linked, one would have to know that these terms have the same meaning and actually refer to the same thing [FREGE 1892]. Here another problem turns up. Some of these terms are ambiguous. In one respect they may all represent a single unit of spatio-temporal activity and can be recognized consistently and distinguished from other units by their conform structure or borders (e.g. a cut). Still some of them have slightly different meanings depending on the methodology employed to record them. The terms stratum and layer can be seen as equivalents but sometimes one term is used as a subcategory of the other. This phenomenon typically appears on Palaeolithic excavations where archaeologists have to deal with fine, thin layers that often are hardly distinguishably from each other. Sometimes it is also used to differentiate geological from archaeological formation processes. In both examples a stratum can consist of many layers but also vice versa. For this reason a stratum or layer is not exclusively used in the sense of a single archaeological recording unit. The same holds true for the term feature. It may be used for describing a stratigraphic unit but more frequently expresses a combination of stratigraphic units, e.g. a pit with a series of fillings. The same can
apply for a hearth where the walls, the bottom and the filling of the hearth can be described as one feature. In Austria, the term “Objekt” (English: object) is used for this phenomenon (fig. 1).

But we also run into problems if we have to deal with terms that represent a defined and accepted concept of some excavation system but do not exist in another system although their methods might be the same. The locus-basket system can serve as an example regarding this matter. This system is often used in Near Eastern archaeology, e.g. in Israel, and was introduced by Wheeler [CALLAWAY 1979]. Whereas the locus corresponds to the definition of the stratigraphic unit [CLINE 2009], the situation is different with the term basket. Baskets probably were used first at the Tell en-Nasbeh excavations in the 1920s or 1930s [PAVEL 2011]. They are used to collect all items found on an excavation and are assigned to one locus [TEL GEZER 2006]. A new basket is usually started on a daily basis (so may have a temporal notion) and this methodology makes the recording of ‘basket numbers’ different to other systems. This provides implicitly the opportunity for documenting the archaeologist’s activities at excavation and is a circumstance that rarely occurs at any European excavation. Among the analysed excavation sheets we can only name one comparable example coming from the excavation guidelines of the German state Rhineland-Palatinate. The so called “Stellenkarte” (fig. 2) lists all actions conducted by the archaeologists on a single stratigraphic unit resp. feature associated with an ID, a date, and a short description [LVR-AMT 2011].
Fig.1 – Excavation sheet for documenting the stratigraphic unit (Copyright: BDA http://www.bundesdenkmalamt.at/document/621701608.pdf)
The locus-basket system and the “Stellenkarte” are maybe more exceptional, but they have the advantage that they document the archaeological excavation process too. Information on the “Stellenkarte” records the whole working process according to date and stratigraphic unit. It also contains information such as setting of additional nails for measuring, measuring and drawing processes, observations on sediments etc. This allows subsequent researchers to understand the way how archaeologists worked. Such information is otherwise recorded in site diaries, which do not easily fit the requirements for comparing data using computers.

Generally, one of the main problems of archaeological recording is distinguishing between description and interpretation. Excavation pro-forma sheets make archaeologists believe that they can record their work as objectively as possible. Excavation guidelines, selection fields within the sheets that restrict free text and therefore personal opinions, manuals for e.g. determining soil texture by feel or soil colour further support this notion. They seem to separate on-site recording from the post-excavation interpretation, which can be seen as one way to distinguish data and interpretation [ROSKAMS 2001, 245]. But every archaeologist is an individual influenced by their own experiences, know-how and skills that cannot simply be set aside. Because of this it is important to know how certain knowledge was derived and “to record the processes by which the data were produced” [BERGGREN & HODDER 2003].
Conceptual Reference Modelling

In order to share primary excavation data on an international level several problems such as different terms and the semantics of different vocabularies have to be overcome [Tudhope 2011]. On the one hand this can be terms with identical spellings that cannot be compared due to their different levels of meaning (homonyms). On the other hand we had to deal with similar concepts that are represented by different terms (synonyms). In the course of our analysis we also had to face the problem that these concepts are rarely clearly defined by archaeologists and that excavation methods as well as documentation and interpretation processes are not always recorded explicitly.

Ontologies in general, but also CIDOC CRM as an ontology for the field of cultural heritage in particular, allow data integration and data migration by “describing common conceptualization behind multiple schemata” [DOERR 2009]. As such CIDOC CRM is suitable to integrate heterogeneous data dealing with collections, but also sites and monuments from fields such as history, fine arts, archaeology etc. [CROFTS et al. 2011]. It enables the modelling of past and present activities, through the specification of classes for Physical Things, Events, Actors, Places and Time Spans with their relevant relations. An essential characteristic of this ontology is the event centric modelling. This means that Physical Things are not directly related to Actors, Places or Time Spans but through Events, and that it is well suited for modelling archaeological activities [CRIPPS 2004]. The advantage of the model becomes obvious when thinking about the layers (stratigraphic units) in an archaeological excavation. They have been created, maybe modified and finally destroyed through events like deposition, maybe disturbance and excavation. These events are related to time spans. Traditional entity-relationship models often directly assign a layer to a time span and loose the possibility to represent the different stages in the lifetime of an object which are nevertheless relevant for scientific analysis.

CRMarchaeo represents an extension of CIDOC CRM which explicitly supports integrating archaeological excavation data [DOERR et al. 2013]. It is developed in the ARIADNE Framework [NICCOLOCCI & RICHARDS 2013] within a team of archaeologists and ontology engineers and is still in an early stage of development. New classes introduced by the CRMarchaeo extension include the “Stratigraphic Unit” and the “Excavation Process Unit” on which we will focus here. The “Stratigraphic Unit” refers to the undisturbed stratigraphy which exists pre-excavation. It represents an entity that was produced by a “Stratigraphic Genesis Event”. This class is defined in order to model the formation of stratigraphic units by processes resulting in the displacement of a defined amount of matter which has settled into a relatively stable form. As Stratigraphic Units and the events of their formation are differentiated, the CRMarchaeo model enables the assignment of possible “Time-Spans” to the formation events.

By assigning temporal relationships based on the law of superposition to instances of the class “Stratigraphic Unit” we can model stratigraphic sequences, hence the stratigraphic matrix of a site which is recorded on the excavation sheets. Furthermore all the different terms found on the examined excavation sheets (context, layer, locus, etc.) can be assigned to the class “Stratigraphic Unit”. Positive and negative stratigraphic units are distinguished by further classes. A Stratigraphic Unit is not the same entity that archaeologists excavate, at least in an ontological sense. To record (explicitly) the event of the excavation of a stratigraphic unit by an archaeologist we have to define an “Excavation Process Unit” carried out by an archaeologist who tries to
excavate approximately the extent of the stratigraphic unit. This apparent complexity recognizes that in the process of excavating, archaeologists can never meet the exact (atomic level) boundaries of a stratigraphic unit. As a useful addition, we are also able to model arbitrary units like spits by also defining them as “Excavation Process Units” that contain one or several “Stratigraphic Units”. Because the definition of the “Excavation Process Unit” allows us to relate this activity to a specific “Time-Span”, e.g. a working day, we can thereby also incorporate the way the archaeologist’s action is documented in the “Stellenkarte” or by the concept of basket. Furthermore, as ontologies enable us to relate one class to another and “Excavation Process Unit” describes an activity, we can connect excavating a stratigraphic unit to the finding of objects.

Consequently events in the past and activities of the archaeologists in the present are differentiated and explicitly available by the concepts of “Stratigraphic Genesis Event” and “Excavation Process Unit”. Work processes carried out by the archaeologists can become traceable in the records if documented and modelled this way. By naming the actions carried out at an excavation we are able to detect the methods employed more easily. Additionally we can describe them in our conceptual reference modelling more clearly and precisely.

CIDOC CRM and its extension CRMarchaeo seem well suited for enabling interoperability between archaeological excavation data. But still, mapping excavation data or other existing information systems to the ontology requires special effort and technical solutions, although additional work has been done regarding this matter [MAY, BINDING & TUDHOPE 2011, SCHOLZ & GOERZ 2012]. With further extensions of the CIDOC CRM to subfields of cultural heritage, and emerging technical developments enabling easier implementation of the CIDOC CRM [TUDHOPE, BINDING, MAY & CHARNO 2013] the likelihood of greater interoperability between currently unconnected archaeological, and other, data is envisaged in the near future.

Conclusion
The mapping of data and systems to CIDOC CRM or its extensions is not a trivial exercise. But a major advantage of using an ontology to map the data and relationships in your archaeological recording system is that you only need to do so once in order to then enable every project recorded using that system to be interoperable with any other project data mapped to that ontology.

With the CRMarchaeo extension we have got a powerful means for representing primary excavation data and the potential to cross-reference and analyse multiple datasets in a way that was not possible before. The fact that we can distinguish between a stratigraphic unit and the activities archaeologists carry out to excavate it does not enable us to automatically differentiate between excavation and interpretation processes [BERGGREN & HODDER 2003]. But as this extension allows us to document excavation activities we are one major step closer to facilitating the reporting of separate interpretative processes at a more detailed level.
Acknowledgement
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 313193.

References


