



*This PDF is a simplified version of the original article published in Internet Archaeology. Enlarged images which support this publication can be found in the original version online. All links also go to the online version.*

*Please cite this as: Katsianis, M., Bruseker, G., Nenova, D., Marlet, O., Hivert, F., Hiebel, G., Ore, C.-E., Derudas, P., Opitz, R. and Uleberg, E. 2023 Semantic Modelling of Archaeological Excavation Data. A review of the current state of the art and a roadmap of activities, Internet Archaeology 64. <https://doi.org/10.11141/ia.64.12>*

# Semantic Modelling of Archaeological Excavation Data. A review of the current state of the art and a roadmap of activities

*Markos Katsianis, George Bruseker, Denitsa Nenova, Olivier Marlet, Florian Hivert, Gerald Hiebel, Christian-Emil Ore, Paola Derudas, Rachel Opitz and Espen Uleberg*

Archaeological data repositories usually manage excavation data collections as project-level entities with restricted capacities to facilitate search or aggregation of excavation data at the sub-collection level (trenches, finds, season reports or excavation diaries etc.). More granular access to excavation data collections would enable layered querying across their informational content. In the past decade, several attempts to adapt CIDOC CRM in order to provide more explicit descriptions of the excavation universe have resulted in the use of domain-specific model extensions (e.g. CRMarchaeo, CRMsci, CRMba). Each focuses on corresponding aspects of the excavation research process, while their combined usage has potential to support expressive data mappings at the sub-collection level. As part of the ARIADNEplus project, several CIDOC CRM developers and domain experts have collaborated to undertake conceptual mapping exercises, to address the practicalities of bringing excavation data descriptions together and to link these to our overall aspirations in terms of excavation data discoverability and reusability. In this contribution, we discuss the current state and future directions of the field of semantic representation of archaeological excavation data and consider several issues that constrain the applicability of existing solutions. We identify five key enabling technologies or research areas (Conceptual models and semantic data structures, Conceptual modelling patterns, Data mapping workflows and tools, Learning technologies and Semantic queries) and assign readiness levels to assess their level of technological maturity. Our research demonstrates that while the existing models and domain-specific extensions are deemed adequate, there is a need for more user-friendly methods and tools to structure data in meaningful and interoperable ways. The next steps involve consolidating relevant semantic structures, improving modelling implementation guidance, adhering to consistent workflows, developing engaging curricula, and documenting real-case examples to demonstrate the benefits and results of semantic data integration.



# 1. Introduction

Let's begin with a question. If the archaeological community succeeded in connecting its excavation data at a granular level following FAIR principles (Wilkinson [2016](#)), with links between individual entities (fields in tables or items in a triple) so that ceramics, lithics, phytoliths, and silty clay stratigraphic units could be aggregated, quantified, and interrogated across a region, what uses would be made of this information resource? The *Tiber Valley Restudy* (Patterson *et al.* [2020](#)), *Archaedyn* (Ferdrière [2023](#)), and the *English Landscapes and Identities Project* (Cooper and Green [2017](#)) provide examples of research applications exploiting the informational capacity of datasets aggregated at the level of distinct archaeological features or single objects. Aggregated data such as these have been used to revise date ranges assigned to ceramics, to assess shifts in regional settlement or land cover patterns, and to interrogate trade networks and links in material culture. These examples indicate two broad research application areas: 1) creating robust estimates for key items of information (e.g. the spatiotemporal distribution of a type of ceramic dish), which can then be used to refine or challenge existing interpretations (e.g. revise chronological time spans) or assess new material (e.g. the dates of newly excavated features), and 2) modelling patterns and trends in data across a region or network in response to needs identified by domain specialists or individual research interests (e.g. the orientation of burials in the European Neolithic or the use of water-sieving in excavation fieldwork).

For researchers, accountability to data and data exploration can also provide mechanisms for questioning and re-evaluating established narratives and assumptions about the character of the archaeological record. In development-led archaeology these data may be used as inputs to predictive models for assessing not simply the likely presence of a site, but the probable range of preservation levels. In heritage management they may contextualise the significance of individual features or landscape areas, while in broader land use planning, they can be used to better inform land use change decisions by providing information which allows a planner to understand whether a feature or find is common, uncommon, common but in an uncommon context, or part of a meaningful wider pattern.

Generalising from these examples, the projected usefulness of aggregated archaeological datasets depends on the premises that: 1) there are benefits to tying heritage management decisions or narratives constructed about the past closely and traceably with archaeological observations and data, and 2) these benefits outweigh the effort required to produce knowledge and take decisions with a high degree of accountability to the underpinning data.

The first premise in particular is, admittedly, debated, with some advocating for greater accountability and some seeing this model of knowledge production as a poor fit for the realities of archaeological interpretation and decision making. Various obstacles to holding interpretation accountable to data have been raised, from a lack of resources for migrating existing data into such a system, to the impossibility of getting agreement on the meanings of archaeological categories, but most salient is



the fact that most interpreters and decision makers simply don't use detailed data when making interpretations and decisions - not because of a lack of detailed data and tools to use them, but because detailed data are not part of this stage of the interpretive process. Furthermore, concerns have been raised about the kinds of knowledge left behind due to the '*emerging supremacy of structured digital data in archaeology*' (Hacıgüzeller *et al.* [2021](#)).

Making it easier to aggregate data and produce FAIR data is key to addressing critiques of the second premise, which emphasises that the effort required currently outweighs the benefits. In this article we focus on recent work within the ARIADNEplus community, assessing the technical and data management problems that hinder integration and add weight to the 'makes accountability to data too difficult' side of the costs-benefits balance sheet. We examine whether these issues, which encompass semantic, data modelling and ontological problems, remain an insurmountable obstacle, or if the real challenges now lie in successfully making the case for the benefits in various contexts to drive uptake.

To make a high-level assessment of this technology and data infrastructure landscape, data models provided by the CIDOC CRM and its extensions, as well as the AO-Cat model, were reviewed to check their coverage of the data elements commonly found in excavation data. Data cleaning, modelling, transformation, and query tools were then surveyed alongside storage and publication systems, together with case studies where these had been applied to assess their effectiveness and ease of use. Materials and toolkits for training were also reviewed to gain insight into their clarity and comprehensiveness. Finally, experimental modelling of real-world excavation data archives using various tools was undertaken as a cross-check. These experimental modelling exercises also provided critical insights into whether the data infrastructure and technologies of the CIDOC CRM were producing the kinds of knowledge desired by their users. The review presented here is preceded by a technology landscape synopsis that attempts to set our findings in the wider context of semantic modelling in cultural heritage.

## 2. The Current State of the Art

Excavation research has been at the forefront of archaeological data interoperability activities. This is understandable as it remains the principal process for scientific data collection and interpretive reasoning in the archaeological domain. Data recording systems and methodologies that aimed to standardise the excavation process appeared even before the discipline turned towards digital options. In British archaeology, General Pitt Rivers and Flinders Petrie, followed by Mortimer Wheeler, are usually credited with the advocacy of more systematic documentation methods (Lucas [2000](#), 18-51), while similar developments gradually emerged in several national contexts. With the advent of Processualism, a new impetus was given to concrete documentation procedures, which included devices such as the Harris Matrix (Harris [1979](#)), pro-forma recording sheets (see Chadwick [1997](#)) and relevant field drawing guides (e.g. Bettess [1984](#)) that gradually crystallised into context-based excavation manuals (e.g. Westman [1994](#)). Very early in the adoption of digital methods within the discipline, approaches such as Gardin's logicist school tried to formalise the archaeological argumentation process (Gardin [1980](#)), while similar



attempts showcased the benefits of data linkages within large-scale documentation projects (Chenhall [1971](#)). These were eventually followed by the development of data models (e.g. Andresen and Madsen [1992](#); Hadzilakos and Stoumbou [1996](#)) and several digital documentation systems that tried to identify the main concepts and information flows within archaeological excavation projects, and attempted to provide database structures that could accommodate different fieldwork conceptualisations or excavation methodologies, e.g. *ArchéoDATA* (Arroyo-Bishop and Lantada Zarzosa [1995](#)), *IADB* (Rains [1995](#)), *SYSAND* (Agresti *et al.* [1996](#)), *IDEA* (Madsen [1999](#)) and *ArchaeoInfo* (Madsen [2003](#)) or *SIDGEIPA* (Diez Castillo and Martinez Burgos [2001](#)). Meanwhile, the increasing use of digital media in the archaeological documentation process raised the problems of digital data preservation and interoperability (Richards [1997](#)).

Initial efforts were directed towards the integration and implementation of generic metadata standards, such as the [Dublin Core](#) (Wise and Miller [1997](#)). However, a consensus was quickly reached that the heterogeneity and diverse implementation of archaeological information required more complex knowledge representation specifications to align the underlying conceptual schemas of existing information structures within a common framework, rather than simple resource metadata. The [CIDOC CRM](#) was perceived as having particularly strong potential for providing an explicit high-level event-based formal ontology (Doerr [2003](#)), having already proven robust and flexible in the domain of museum documentation and other related cultural heritage subfields. Although, other proposals were put forward for the archaeological excavation domain (e.g. Kansa [2005](#); Zhang *et al.* [2002](#)), the broad uptake of the CIDOC CRM by the cultural heritage sector and its accreditation as an *ISO* (International Organization for Standardization) standard in 2006 eventually triggered its further development and penetration into domain-specific implementations or facets that extended its usability, with significant impact across important elements of documentation in the archaeological and cultural heritage domains, illustrated by the creation of *CRMdig*, *CRMsci*, *CRMinf*, *CRMarchaeo*, *CRMba* and *CRMgeo* (Bruseker *et al.* [2017](#)).

Despite these efforts to resolve conceptual and semantic problems, few examples of interoperable datasets exemplifying the utility of such work are published and accessible. The notorious diversity of archaeological methods, devices and data production practices is often cited as the leading cause for the continued lack of integration between datasets in terms of structure, granularity and terminology employed (see for example Attewell *et al.* [2004](#)). This variability in how excavation fieldwork is envisaged, tooled and performed is reflected in the archaeological data archives we produce, which employ distinct conceptual descriptions at variable granularities, target diverse information outputs (a report or a data archive), can often be unfinished or open-ended, and may be linked to all sorts of digital data types, each with its own complicated production workflow.

The [ARIADNE project](#) as a European and, aspirationally, global project provided a useful proving ground on which to address these challenges. The ARIADNE project aimed to consolidate and aggregate into its infrastructure, the *ARIADNE Content Cloud* (AC), a large number of archaeological datasets of different kinds (ranging from excavation and survey data, to national monuments records, to heritage



science outputs) (Aloia *et al.* 2017a; Niccolucci and Richards 2019; Richards 2023). The service supports searches based on spatial, temporal and thematic criteria for the retrieval of entire data collections with different levels of granularity, maintaining the conceptual links between different dataset levels (i.e., collection, find, trench, etc.) and enabling resource discovery (Aloia *et al.* 2017b). However, in contrast to contexts in which such efforts succeeded, such as the granular integration of coin datasets (Felicetti *et al.* 2015), the integration and retrieval of excavation data resources has remained limited and is mainly available in the ARIADNE portal at collection level and on the basis of subject keywords, rather than their item-level integration as individual objects. An example search using the terms 'neolithic ditch' illustrates the current modest level of integration and, implicitly, of semantic interoperability (Figure 1).

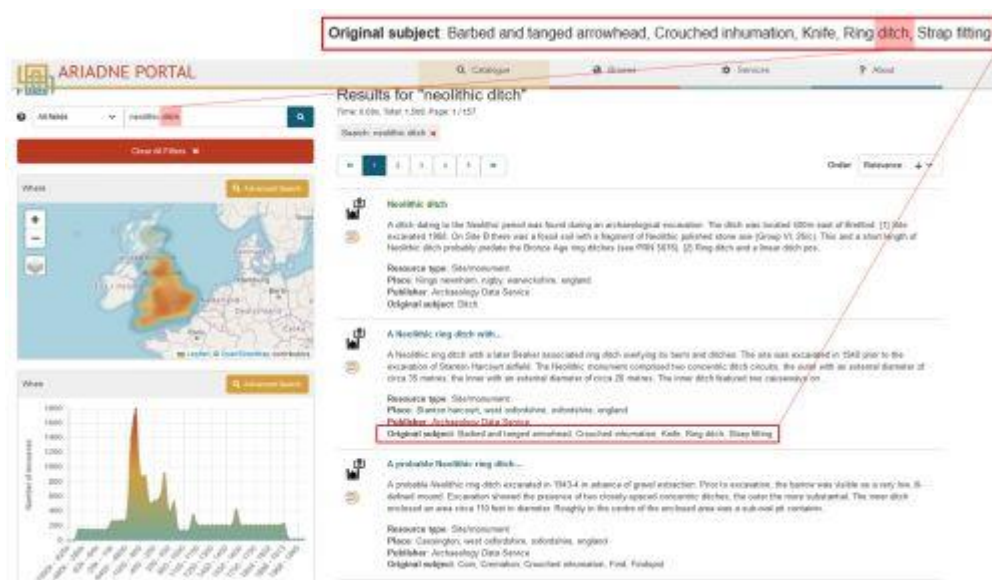


Figure 1: Searching for excavation data items may retrieve results only if the term (referring to a portion of that data collection) is used as a keyword or thematic subject for the entire data collection

While the example from the ARIADNE portal shows a query run at the scale of a pan-European or global meta-aggregator of archaeological datasets, it also reflects the current situation in many national archaeological repositories, where excavation data archives are still usually integrated as project-level data collections, with severely restricted capabilities to accommodate interoperability at the sub-collection or item level (e.g. by enabling connections between individual finds from multiple excavation archives or aggregations of geospatial data documenting archaeological features). To achieve the goal of interoperable excavation datasets at those levels, beyond aligning different conceptual understandings of the excavation domain, we need to understand the root causes of persistent modelling problems, assess the fitness-for-purpose of available modelling and data mapping tools, and use this information to improve the practical toolkits available to support data modelling workflows.





### 3. The Work of the ARIADNEplus Excavation Modelling Group

As part of the ARIADNEplus project, the *Archaeological Excavation Modelling Working Group* was established to bring together what has largely been a fragmented community of scholars working on related research, review diverse strategies for excavation data mapping, consolidate data modelling expertise and provide a roadmap for further developments.

The group set out to investigate whether some of the problems in the alignment of conceptualisations of the excavation process could be solved through the creation of a dedicated archaeological excavation Application Profile (i.e. a model that reuses *'the elements of existing ontologies and models for the description of similar entities in the new research context, limiting the development of new elements only for the description of peculiar and typical aspects of the specific discipline'*, Richards *et al.* [2022](#), 10) or whether existing elements in the wider CIDOC CRM family of models suffice to enable sub-collection or item-level integration of excavation datasets. The theoretical complexities, practical barriers and the diversity of technological solutions involved in excavation data modelling made a straightforward assessment impractical, leading to the conclusion that a survey was necessary to map the overall landscape of excavation data modelling including existing efforts, tools, methods and, often not considered, skills requirements.

On 15 June 2022, the group organised a virtual workshop on semantic mapping of excavation data. The presenters explored semantic modelling and the use of CIDOC CRM, as well as the tools developed to assist researchers with mapping their data. Five case studies on semantic mapping of excavation data were also presented. A report from the workshop provides summaries of the presentations and the ensuing dialogue, as well as an overall discussion on the current landscape of excavation data mapping methods, tools and case-studies (Katsianis *et al.* [2022a](#)). Video recordings of the presentations alongside an open-access version of the report have also been made available via the ARIADNEplus website through Zenodo, while a summary of the group's work (Katsianis *et al.* [2022b](#)) has been included in the open deliverables of the project (Richards *et al.* [2022](#)). Finally, a presentation at EAA2022 (Nenova *et al.* [2022](#)) discussed some of our initial findings (see also Bruseker [2022](#)) and provided the baseline for the present contribution.

The excavation data modelling landscape review, summarised here, revealed varied levels of technological maturity across different elements involved in archaeological excavation modelling. To structure the assessments, in finalising the review we assigned *Technology Readiness Levels* (TRLs) (EARTO [2014](#); EC [2014](#)) to five key enabling technologies or research areas (Table 1). Many elements have been assessed as at technology readiness levels 4-6. This places many of the technologies in the middle to latter part of the 'innovation' phase, which includes technologies that are mature but not yet ready for robust operational deployment. However, level 7 technologies are also present, as in the case of AO-Cat, which according to Richards *et al.* ([2022](#), 9) has reached a *'mature level of stability'* and is ready for *'wider usage outside the boundaries of the ARIADNEplus'* project (Table 1).



Based on this assessment, we present the current implementation environment and set out a roadmap for future directions, highlighting areas where particular effort is needed and proposing mechanisms to increase the maturity of these technologies, while growing an early user community.

Table 1: Technology Readiness Levels (TRLs) for key enabling technologies or research areas involved in archaeological excavation modelling

Key enabling technologies/research areas	TRL	Description	Example
Conceptual models and semantic data structures	7	System prototype demonstration in operational environment	<i>e.g. CIDOC CRM, AO-Cat, ARIADNEplus Knowledge base, GraphDB</i>
Conceptual modelling patterns	4	Technology validated in a lab	<i>e.g. Semantic Reference Data Models, Zelij Semantic Pattern Platform</i>
Data mapping workflows and tools	6	Technology demonstrated in a relevant environment	<i>e.g. Karma, Protégé-Ontop, SHACL, X3ML toolkit, Vocabulary Matching Tool, PeriodO</i>
Learning technologies	5	Technology validated in a relevant environment	<i>e.g. CIDOC CRM periodic table, OntoMatchGame</i>
Semantic queries	5	Technology validated in a relevant environment	<i>e.g. Openarcheo, Sparnatural</i>

## 4. Research Areas

In this section, the five key enabling technologies or research areas identified by the group as having potential to consolidate and coordinate ongoing work within the excavation data modelling community are discussed for their readiness levels through the retrospective consideration of their development, the presentation of diverse implementation examples and their integrative examination in the framework of ARIADNEplus.



## 4.1. Conceptual models

Early efforts like CRM-EH (Cripps *et al.* [2004](#)) tested the applicability of CIDOC CRM to model the archaeological excavation domain. Despite the overall suitability of the model, the identification of practical implementation issues and expressive gaps by the CIDOC CRM community (Ashley *et al.* [2011](#); May *et al.* [2011](#)) motivated the proposal and development of a domain-specific extension for excavation research. Based on the documentation practices of several countries and in combination with CRM-EH, *CRMarchaeo* was developed to describe the specifics of the archaeological excavation process (Doerr *et al.* [2016a](#)) with a focus on single-context excavation (Harris [1979](#)). The model makes reference to concepts that were already provided by the *CRMsci* extension, itself developed to integrate 'metadata about scientific observation, measurements and processed data in descriptive and empirical sciences' (Doerr *et al.* [2012](#), 4).

CRMarchaeo's launch was followed by implementation examples that employed different excavation traditions (e.g. Masur *et al.* [2014](#)), which revealed the necessity for additional concepts and led to further improvements (e.g. Hiebel *et al.* [2017](#)). Further harmonisation efforts targeted other domains that interface with excavation practices (e.g. for building archaeology see Ronzino [2017](#) for spatio-temporal reasoning see Hiebel *et al.* [2014](#) for archaeological argumentation see Marlet *et al.* [2019b](#)), while other examples included multiple concepts from several CIDOC CRM models (e.g. Hiebel *et al.* [2021](#); Christaki *et al.* [2017](#)). The model has recently been upgraded to version 2.0 (Felicetti *et al.* [2023](#))

At its heart, the effort to integrate different standardisation initiatives carried out at a conceptual and technical level into a common information layer will depend upon a conspicuous and clear meta-representation of processes and objects in a consistent set of conceptual models. The conceit of the evolution of the main CIDOC CRM (CRMbase) into a series of extensions allowing domain-specific definitions is that their combination should potentially be more successful in capturing domain-specific meaning using targeted domain-focused extensions, while still adhering to the core ontological pattern. This strategy, while theoretically powerful, at a practical level has increased implementation difficulties as modellers have to navigate within an ever-expanding universe of concepts and properties and inheritances, while multiple implementation paths with complementary concepts mean that there is more than one way to express certain complex, real-world facts using the same, formal conceptual system. Because archaeological excavations are multi-level processes, often taking place across multiple fields, using diverse methodologies and tools, and producing a range of documentation media (see e.g. Jones [2002](#)), modellers are regularly confronted with these difficulties when attempting to provide a coherent description.

One key issue is that during excavation the meaning of real-world entities or their documentation proxies can be mapped to different model concepts, depending on the research context or stage. What may be identified as a feature (E26) during excavation may be referred to as a stratigraphic interface (A3) that separates stratigraphic entities, as a rigid physical feature (S20) that can be examined for its properties, as a filled morphological building section (B3) in architectural studies, as





a belief (I2) to be a neolithic wall part in interpretative statements, as a digital data object (D9) that substitutes the original physical object in post-excavation research, and as a material entity that occupies a spatial location described through coordinate measurements (SP6). While such fine conceptual distinctions are the bread and butter of conceptual modelling, actually documenting, representing and relating them in practice is another matter. At a practical level, one needs to distinguish between different contexts or stages of the excavation process and establish common descriptions that facilitate a basic level of agreement (Figure 2).

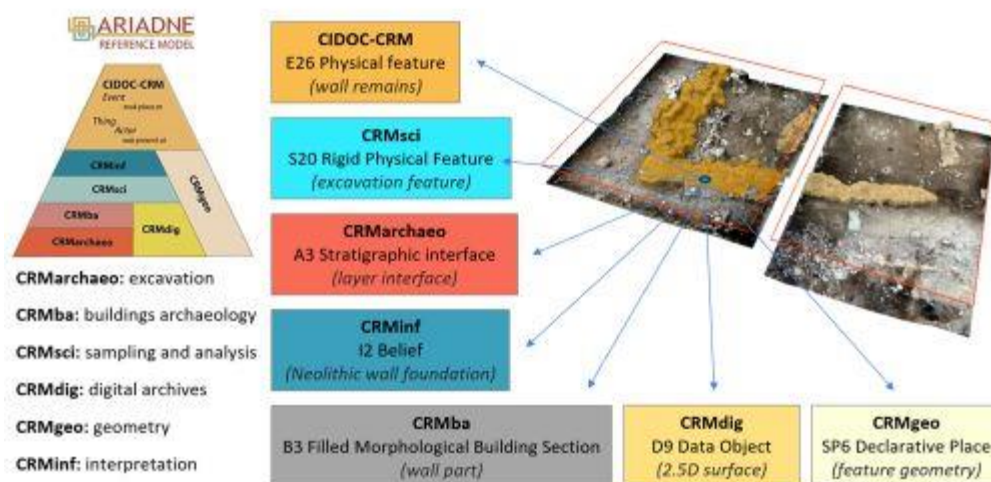


Figure 2: The ARIADNE reference model and the complexities of semantic modelling in the excavation domain with respect to retrieval, interpretation and representation of excavated entities

Another important issue is the multiplicity of potential paths for connecting entities and processes when modelling the same phenomenon. In the context of an excavation dataset, alternative semantic paths may be used to link entities and their underlying meanings, and these paths create strongly divergent results. This is important because the way an interpretation, for example that someone was buried intentionally, is linked with the archaeological evidence observed during excavation, reflects specific ideas we may want to highlight, such as past social activity, a specific action like digging, or the production of a burial place. Many possible routes can link the interpretation to the evidence. Some may be more direct or perhaps provide a better match with common archaeological data structures. While this directness or ease of modelling should be taken into consideration when choosing a path, it is also important to consider that the choice of route will affect the kinds of queries that can be easily run against the data (Figure 3).



Figure 3: An example of alternative routes to link entities that convey different meanings

Beyond these core conceptual difficulties, there is a practical barrier that regularly deters real-world implementation of CIDOC CRM models of excavation data. Changes to the CIDOC CRM concepts as a result of its evolution can require relatively frequent, labour-intensive updates that excavation projects do not have the resources to support. Although CIDOC CRM provides migration instructions for deprecated classes and properties (e.g. in CIDOC CRM v.7.1.1 - Bekiari *et al.* [2021](#) - the concept E38 Image was deprecated and replaced by the use of the more generic E36 Visual Item), the updates of descriptions required as a result may be more complicated, and resourcing this work will become more complicated over time, requiring curation provisions in the form of algorithmic techniques and tools for ensuring version control and backward compatibility (Tzitzikas *et al.* [2014](#)). Even if the mechanisms are there to ensure the validity of semantic descriptions, the idea that a completed conceptual mapping may require updating or even radical re-engineering may act as a deterrent to the modelling community (for a relevant example see 4.3.4).

In light of these issues, the practical activity of testing the application of semantics from the entire CIDOC CRM ecosystem to archaeological excavation data, focused on assessing their conceptual and syntactic correctness, while mitigating the negative effects of too much diversity in the modelling choices deployed. The aim then became to understand how to consistently create, document and apply common conceptual modelling expressions to the same real-world situations. That said, exceptions to this, which identified gaps requiring new modelling, did come up with regards to the question of how to model the relationship of archaeological survey information to excavation data, as well as some nuances of chronological dating. In the first instance, the work of the *Semafora project* complements the existing official CRM extensions by proposing new CRM classes to represent the survey process and link it to excavation research (*CRMSurv*, see Nenova *et*



al. 2023). In the second case, attempting to model the complexities of multiple temporalities (e.g. creation, deposition or recovery time of an artefact) that may also range between highly scientific and reliable (e.g. absolute dating) to more subjective and relative judgements (e.g. based on artefact typologies), led to the development of CRM classes and properties as a part of the [CRMaaa model](#), an informal extension of the CIDOC CRM designed for application in the area of art and architectural historical research.

Finally, several application profiles, such as *CRMhs* (for encoding information derived from Heritage Science research), *CRMtex* (for the description of textual entities), the *Mortuary Data AP* (for modelling different types of burial data) or the *aDNA AP* (for capturing contextual information about the processing of respective samples) have been developed as part of the ARIADNEplus ontology implementation (Theodoridou 2022; Aspöck *et al.* 2023) and have potential to be linked with processes that form part of the archaeological excavation domain.

## 4.2. Conceptual modelling patterns

In considering the application of ontologies and semantics to archaeological datasets a gap exists between the possibilities enabled by such high-level representations and their practical implementation. In practice, archaeologists asked to model their information semantically face two challenges: understanding the ontology and its application and arriving at modelling compatible with that of others with whom they wish to share their data.

Application profiles are offered as a solution, consisting, in essence, of a subset of classes and properties from one or more ontologies, which are deemed appropriate and adequate to the modelling situation. [OntoMe](#) exemplifies an initiative to provide the tools to carry out this action (Beretta 2021). However, it does not enable a deep understanding of the ontology that its users are asked to apply, nor does it give specific instructions on how to string together the classes and properties in a consistent way to generate a consistent data representation.

Creating *Semantic Reference Data Models* (SRDMs) for early calls seealso Tudhope and Binding 2013) is another proposed approach, which seems particularly useful for describing the archaeological excavation domain. SRDMs provide a guide for how to implement a chosen ontology or ontologies within a certain context for a collection of typically documented items. A set of SRDMs covers a range of typical entities documented and their interrelations. Within the context of archaeological excavation this results in a meta analysis of the kind of things that show up in documentation and then creating a set of patterns for representing their typical attributes recorded, using the chosen ontology. It provides users with *semantic patterns* or *recipes* that can be reproduced. In doing so, explanations of what the model documents and how it is to be used can be described in the context of scholarly discourse rather than in the abstractions of an ontology. This approach can foster a less challenging familiarisation with semantic modelling processes for domain experts, increasing the possibilities for a greater number of compatible implementations. It can also provide critical studies of model definitions in an applied form and identify the problematic areas that require further development with respect to their ontological integrity or



their compatibility with different archaeological excavation methods and interpretative procedures.

In our effort, we brought together data-mapping examples to explore similarities and differences, and compare their scope and meaning. The analysis of existing modelling examples and implementations has the potential to identify such modelling scenarios and provide a closer understanding of the evolution of the respective CIDOC CRM extensions (see for example similarities and differences in Cripps *et al.* 2004; Cripps and May 2010; Christaki *et al.* 2017; Giagkoudi *et al.* 2018). Such a strategy can obviously start from core or generic entities involved in the excavation process and subsequently be extended to cover more specific meanings. In carrying out this task, we also relied on existing work in the cultural heritage semantic data-modelling community, particularly borrowing from the patterns of the [Linked.Art](#) and [SARI](#) modelling work. Each of these communities have worked to develop fundamental patterns for the documentation of basic attributes of objects in the cultural heritage domain, from which we were able to produce suitable 'field' and 'collection' level modelling patterns, extending them and supporting the creation of a unique set of SRDMs for excavation data (Nenova 2022).

The common entities for all or at least most archaeological projects' records have to do with the process of excavation and the discoveries resulting from that process. These discoveries or the lack thereof further invoke the documentation of interpretative and analytical data and relationships. In total we have outlined 22 SRDMs covering important entities or notions of the excavation universe: *archaeological project, archaeological excavation, archaeological survey, archaeological site, stratigraphic unit, context, area, trench, place, finds collection, artefact, feature, sample, image, digital object, biological object, person, institution, period, analysis, textual work, and archival unit*. Although some of the SRDMs are still under development (e.g. analysis), we believe that the entire set allows the modelling of some of the most typical phenomena recorded in the course of archaeological work, providing reusable recipes for their documentation (Figure 4).

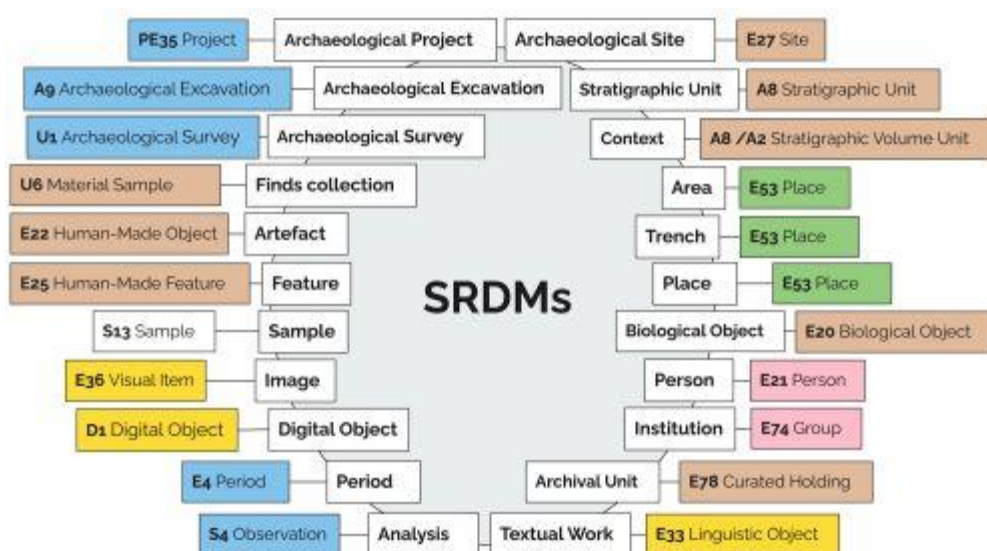


Figure 4: Recognition and assignment of base classes and models





*Archaeological Project* (PE35) represents the overarching collection of activities serving a particular archaeological research goal. The core of the model records its timespan, participants and any descriptive information defining the purpose of the project or other relevant information. An archaeological project may include both excavations and surveys as well as different kinds of analyses and research. Since archaeological excavations and archaeological surveys have distinct natures in terms of methodology, techniques and field practices, they have been separated in two different SRDMs represented by two different CRM classes.

*Stratigraphic Unit* and *Context* SRDMs cover the main stratigraphic entities in archaeology, allowing us to determine and interpret synchronous events and their relationships with other events. Besides such relationships, these models allow information to be recorded about the dimensions, soil characteristics, inclusions, artefact associations, image, spatial information, as well as anything related to the process of excavation expressed by the CRM class of *Archaeological Process Unit* (A1) capturing the excavation activity. Under this node one can record the timespan, participants, techniques and tools used in the process and any other descriptive information that may relate.

*Area*, *Trench* and *Place* are three semi-overlapping concepts in the archaeological excavation documentation context. Both area and trench are modelled as *place* (E53) but serve somewhat different purposes depending on different project's excavation methodology. Such models capture spatial data, dimensions, related contexts and stratigraphic units. This SRDM consists of fields capturing data related to the description of a place as a geographic unit and associating various dimensions, geographic definitions and coordinate files. Each place or entity associated with a place (e.g. a feature) comes with the necessity of expressing spatial information, such as location coordinates and their coordinate system, which can be captured by a reusable pattern employing classes and properties from CRMgeo (Figure 5)



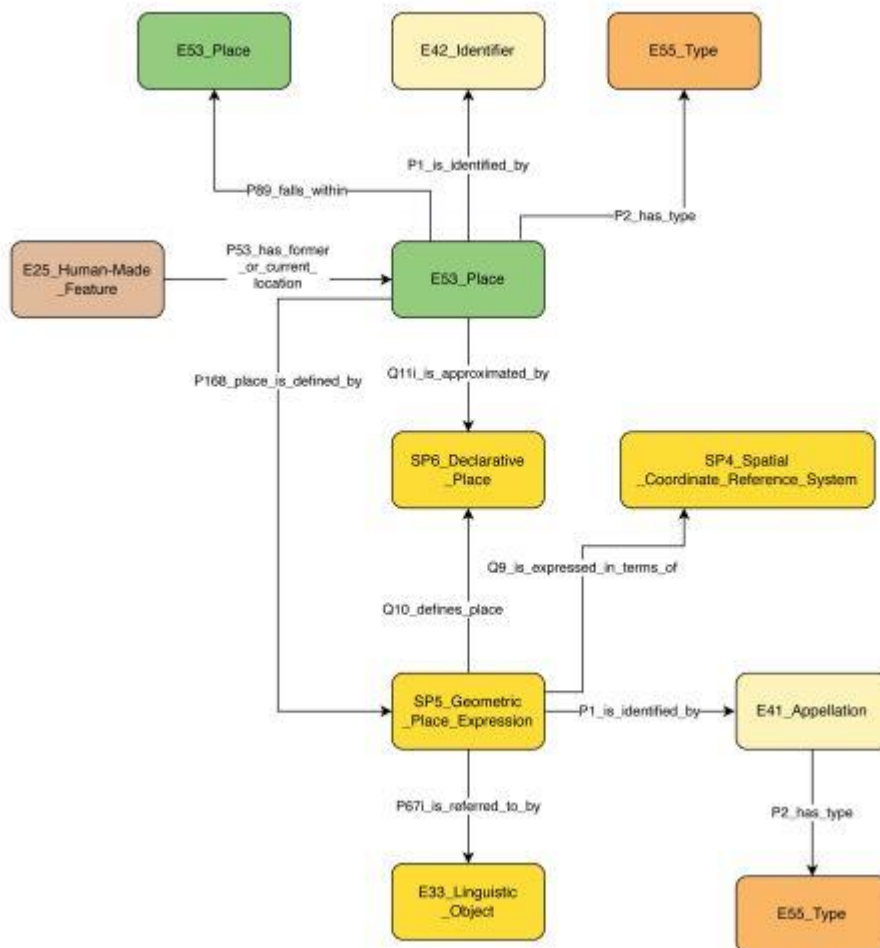


Figure 5: Modelling of geospatial data such as coordinates, coordinate systems and projections

The *Finds Collection* model helps to document the cumulative collection of finds from a context or a survey unit. Its root node, *U6 Material Sample*, is borrowed from CRMSurv and functions as a subclass of *S13 Sample* and *E18 Physical Thing*. The finds collections are often referred to as 'bags' of finds and may contain human-made as well as any other physical or biological objects. Often these 'bags' are given an identifier in the field and later processed with a reference to this identifier along with other information related to the collection process. This model documents data around this collection process as well as definitions associated with various general types of objects collected, dimensions such as weight or count and other descriptive, spatial or documentation-related data.

The *Artefact* model (E22) handles information about any human-made object, inventoried separately or as a part of an associated group of objects. The artefacts may or may not be a part of the finds collections described above and may also require isolated spatial reference. Dimensions, typological descriptions, typological comparisons, chronological assignment as well as documentation reference are other types of data points included in this model. The contents and the associations of the *Biological Object* SRDM, created to document human and animal bones but including shell and other biological non-modified-by-humans objects, are similar.



Information about human and animal bones discovered during the excavation are expressed as Biological Objects in a separate SRDM *Bones* (E20).

The *Sample* SRDM (S13) documents any portions of material extracted from a context or an object to be a subject of scientific analysis. Examples are radiocarbon, micromorphological, petrographic, etc. The model is set up to record data related to the sampling process as well as the physical nature of the sample and the type of analysis it is intended to serve. This model is closely related to the *Analysis* SRDM (E7), which requires further work to be able to apprehend all of the specifics of different analyses. The class *Sample* (S13) is employed to cover portions of an entity sampled to be analysed scientifically, involving analyses and techniques such as micromorphology, radiocarbon, OLS, etc. The model covers the process of sampling, the analysis involved and the interpretation of the results, the material nature of the sample with its dimensions and storage location, as well as any actors involved.

The *Feature* model (E25) captures any standing structures as human-made objects such as walls, floors, pits, hearths, etc. The relevant fields are similar to the artefact model without modelling the interpretative level of such structures at this point. Both *Human-Made Feature* (E25) and *Human-Made Object* (E22) classes stand to document the features/structures and artefacts respectively within Feature and Artefact SRDMs. The patterns cover data associated with their discovery, association with contexts and other features, levels of interpretation such as assigning types based on comparative studies, dating, current location, etc. Tasks are modelled to record activities such as drawing, photographic, sampling of the artefacts including types of tasks, assignees, timespans.

Most of these models have associated *Image* (E36) and/or *Digital Object* reference (D1), pointing at photographs, drawings, spatial files and any other available digital resources. *Textual Work* (E33) model stands for mostly bibliographic references or any 'notebook' data entries and reports. References to actual archives are recorded in an *Archival Unit* (E78) curated holding model.

Two models that exist as a reference in almost all of the other SRDMs are *Person* (E21) and *Group* (E74), which cover information about people and most commonly institutions involved in the projects and different activities associated with them. The SRDMs that refer to people and groups also contain fields where one can record in what capacity those actors were involved (Figure 6).

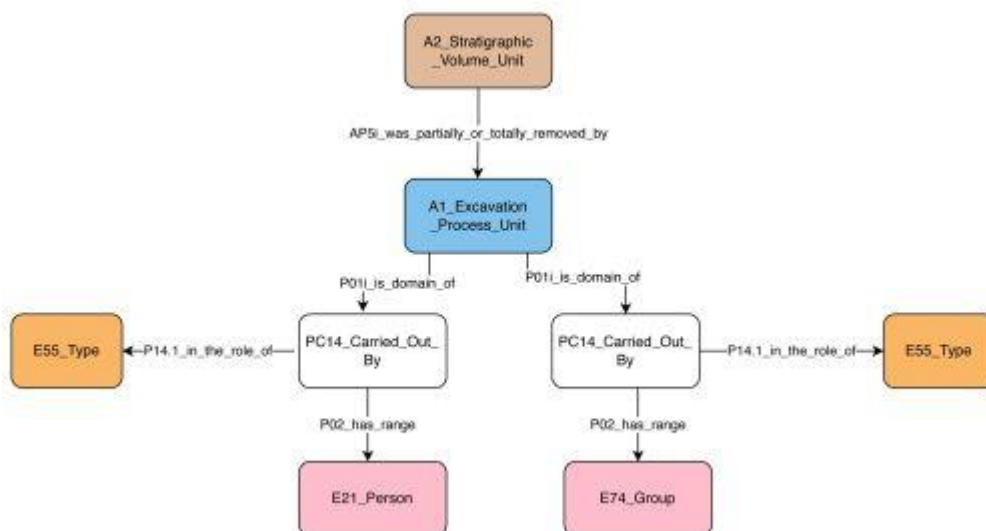


Figure 6: Modelling links from excavation activities to the people or groups involved in their performance

The *Period* SRDM (E4) replicates information about historical and archaeological periods known by certain names in different parts of the world. The period's timespan as commonly assigned by scholars, as well as its corresponding online reference uri, are also a part of the model (Figure 7).

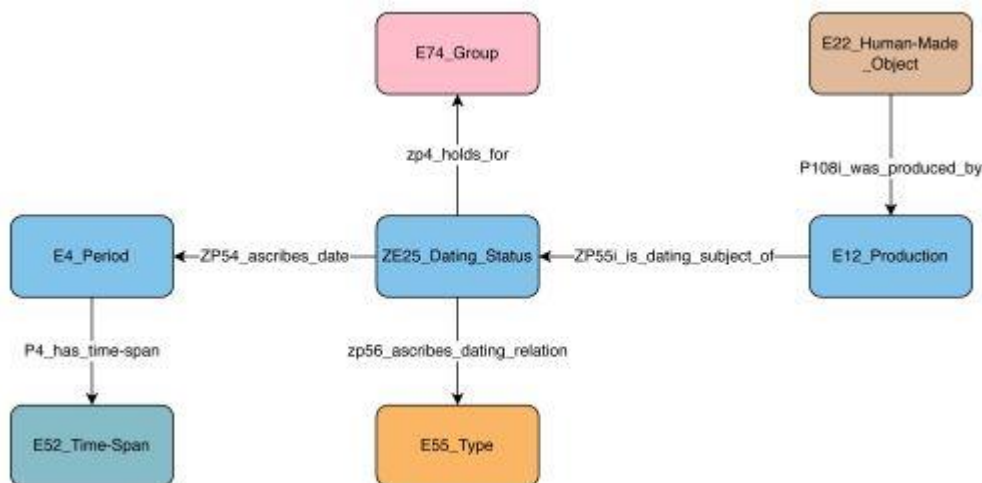


Figure 7: Modelling the assignment of chronological statements to archaeological materials

The resulting work, derived from looking at different excavation data and excavation data modelling patterns, is a relatively comprehensive set of interlinking models that provide recipes for the implementation of consistent modelling patterns for common scenarios in the archaeological excavation data modelling context using CIDOC CRM and its extensions. The present state of the full documentation can be found in Katsianis *et al.* [2022b](#) (annex B). This is a version 1 working draft that is targeted for further development: 1) to include concepts from unofficial CRM models such as CRMaaa or new/other APs, 2) to form a modular data description building process or a **modelling cookbook**, 3) to be linked to implementation examples and modelling guidelines, and 4) to provide a research avenue for the potential usability of *Machine*



*Learning* (ML) procedures that could transform domain-specific semantic modelling into a marketplace for assisted data model creation or alignment (for an assessment of ML techniques for leveraging CIDOC CRM see Tzitzikas *et al.* [2022](#)).

## 4.3. Workflows and tools

Counter-intuitively, the standardisation of the representation of archaeological excavation data through semantics leads not to a consolidation of set tools for the creation, representation and reproduction of archaeological data. Instead, it has opened the data to be taken up by a plethora of different software tools designed to handle different aspects of the data lifecycle. As a research area it is characterised by rapid development and little consolidation. What is interesting to observe **in different efforts** are the possible successful and unsuccessful combination of different tools to different ends.

In this section, we document several such methods in different use cases in order to illustrate some of the main strategies and tools presently used, and how they were put together into functional workflows for the generation and maintenance of semantic archaeological excavation data. Particularly, we have identified several different types of functionality that archaeological excavation semantic data is put towards: integration of completed research datasets, intra-organisational data homogenisation, the creation of archaeological excavation data ecosystems, the supply of meta repositories, and their applications in publishing. The use cases that follow illustrate the progress in these areas and how different tools are combined to create and make use of semantic archaeological excavation data.

### 4.3.1. Finalised research datasets

An approach that manifests its usefulness for aligning finalised or closed datasets, attempts the direct creation of RDFs from spreadsheets or databases. An example of such a data-mapping pipeline employs the excavation data from the project *Prehistoric Copper Production in the Eastern and Central Alps*, which documents prehistoric mining activities in the eastern Alps of Austria following the documentation guidelines of the Austrian Federal Monuments Office (BDA – Bundesdenkmalamt).

An Open Research Data Pilot project led by the Archaeological Department of the University of Innsbruck, analysed the respective datasets and rearranged them into sets of spreadsheets or data tables that correspond with a combination of semantic standards, including the CRMbase with its extension CRMsci (modelling of physical things) and CRMarchaeo (modelling of archaeological documentation). Further concepts specific to Mining Archaeology research were mapped onto the [DARIAH Back Bone Thesaurus](#), a *metathesaurus*, under which different vocabularies and terminologies in use in the domain of the Arts and Humanities can be aligned, while [SKOS](#) (Simple Knowledge Organization System) was used to organise specific vocabulary terms. Individual tables were then integrated within a PostgreSQL database that allowed further data structuring (e.g. URI addition). Afterwards, semantic tools, such as [Karma](#) or [OntoText-Refine](#) were used for RDF creation to be ingested into a [triple store](#) (Figure 8).







structures. However, it involves intensive effort in data structuring and cleaning, requiring a balance between the mapping depth and the available resources. In many cases, where legacy datasets with different data structures are involved, such an approach may require decisions regarding the selection of the essential information worthy of being semantically aligned in each case.

### 4.3.2. Data homogenisation within organisations

In many cases, the adoption of digital documentation procedures may enforce a certain degree of conceptual conformity across an organisation or even entire archaeological sectors. The case of Norway is telling. Notwithstanding the standardisation of artefact databases since the 1990s, the introduction of digital documentation in excavations resulted in diverse solutions. In 2011, the Swedish [INTRASIS](#) was chosen as the default excavation documentation system across the archaeological sector. An INTRASIS database consists of two distinct parts: a metadata template and the actual data. The templates are object-orientated, enabling users to define class-subclass-attribute hierarchies and relationships between the (sub)classes. The flexibility and the ease of changing the templates on the fly may make data integration across individual databases challenging (e.g. Løwenborg *et al.* [2021](#)).

In Norway the institutions responsible for archaeological excavations agreed to a common template, despite the possibility of adding restrictions to excavation documentation implementations. In the last 12 years this template has been revised on an annual basis. In addition, there have been ad hoc minor changes during excavation fieldwork, usually at the subclass and attribute levels. Consequently, individual metadata schemes may be similar but not identical, causing problems for data aggregation. The [ADED](#) (Archaeological Digital Excavation Documentation 2018-2021) project targeted the integration of hundreds of excavation datasets into a single searchable information system based on PostgreSQL/PostGIS, as well as data exporting provisions to a format conforming to the CIDOC CRM family of models, for compatibility with ARIADNEplus requirements. The analysis of the excavation databases in ADED contributed important insights that have been considered and used in the development of the current version 2.0 of CRMarchaeo.

This task was approached by mapping all individual templates to a single common template. First, an empty INTRASIS scheme with all tables extended with a column for data provenance, that is, the name of the individual INTRASIS database, was created. All individual databases were imported into this common database and analysed for their structural differences. In an object-orientated structure the 'class, subclass(es), attribute' hierarchy can also be transcribed into a set path from the root(s) to the attributes (leaves) (Figure 9).

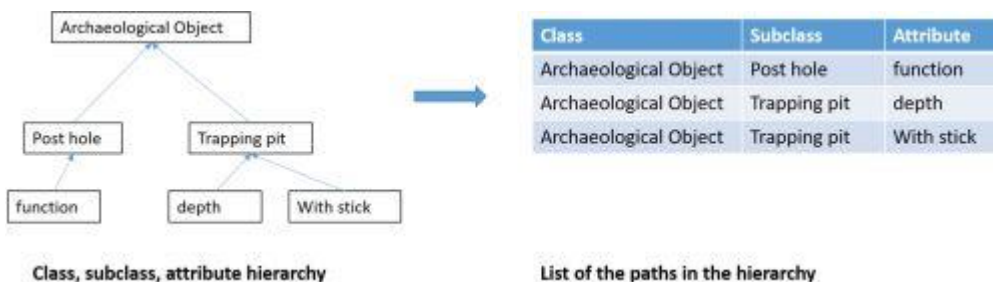


Figure 9: A hierarchy transformed into a list of paths

For a set of 597 INTRASIS datasets the path table consisted of 5100 lines. The normalisation process started with the classes, then the subclasses, their attributes and finally all relationships, requiring less than a month of manual processing. The results of this analysis contributed to an assessment of the degree of field usage that was used to inform the scientific board responsible for developing the national template. Afterwards, work proceeded with mapping the template(s) onto CIDOC CRM, including CRMInf, CRMsci and CRMarchaeo. INTRASIS classes were mapped to CRM classes while, surprisingly enough, almost all subclasses were mapped to types. Attributes were mapped to types or typed strings/values (Figure 10). The mapping of the relationships further revealed unexpected inconsistencies and singularities in the templates (Ore 2022).

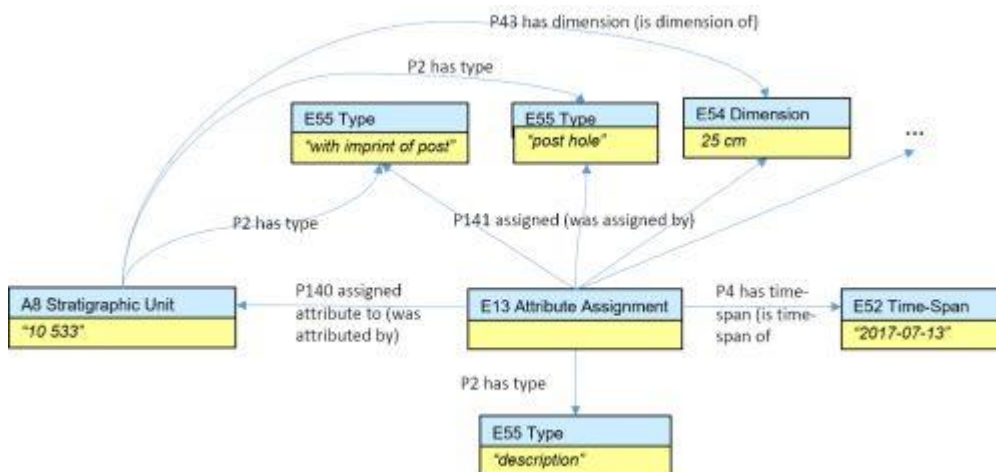


Figure 10: INTRASIS: An object (id:10533) of the subclass 'post hole' of the class 'Archaeological Object' with the attribute 'post function' with the value 'roof supporting', with the attribute 'described at' with the value '2017-07-13' and attribute 'depth' with the value '25 cm'

The normalised path tables have been used in subsequent work involving data cleaning, term normalisation and alignment with reference thesauri (e.g. AAT or PeriodO). This workflow demonstrates the effort involved in aligning large numbers of similar datasets together, showcasing the benefits of following standardised denominations for database elements and controlled vocabularies, when implementing a data schema for archaeological excavation research.

The excavation data from the INTRASIS databases are published in the open [ADED](#) interface. In addition this interface gives access to the published



excavation reports, photo documentation and artefacts. The interface for photos and artefacts builds on the [unimus API](#), which is also accessible via the [unimus portal](#). Links to the open version of the National Heritage and Environment Register, [Askeladden](#), are also included. In this way, linking information at the site level is achieved, while further work should be directed towards facilitating linkages at the level of single structures within excavations.

### 4.3.3. Archaeological data ecosystems

In the framework of the [MASA consortium](#) of the French *TGIR Huma-Num* a workflow covering the archaeological excavation data life-cycle has been gradually developed and refined over the years using a set of operational open-source tools and infrastructures (Figure 11). These include a first step of dataset cleansing with [OpenRefine](#) and vocabulary normalisation with [Opentheso](#), followed by ontology structuration and vocabulary alignment into interoperable data clusters that are subsequently mapped with [Protégé-Ontop](#) and validated with [SHACL](#) as an RDF TripleStore, to be finally implemented using [GraphDB](#) within [OpenArcheo](#), a federated semantic web platform. Using this generic backbone, an archaeological data ecosystem is formulated, linking data ingestion processes within a high-level interoperable data pool (Marlet [2022](#)).



Figure 11: Workflow used by the MASA Huma-Num consortium to support the publication of archaeological datasets on the semantic Web



The conceptual glue is provided by a generic model that employs a selection of CRMbase, CRMsci, CRMarchaeo and CRMba entities, as well as relevant properties that are considered necessary and sufficient to represent the core of excavation data (such as archaeological site, artefact, documentation, etc.). Several gazetteers and standard vocabularies (*PACTOLS*, *GeoNames*, *VIAF*, *PeriodO*) are also employed to homogenise the terminology used across the various datasets. In this respect, the Openarcheo model provides an additional overarching conceptual overlay for heterogeneous datasets that identifies their main commonalities and allows their investigation.

The core functionalities of OpenArcheo rely on [Sparnatural](#), a Javascript-based system allowing the user to query an RDF graph with a graphic interface and without having to write directly in SPARQL. Instead, a visual query interface allows the user to query the RDF graphs without a deep knowledge of the CIDOC CRM ontology, facilitating the interpretation and translation of the OpenArcheo generic model from Entities or Triplets into a sentence construction process. To allow for a more intuitive concept understanding, an 'editorial ontology' replaces the original denomination of the CIDOC CRM concepts and is supplemented by a system of icons for the main components of archaeological data. For example, E22\_Man-Made\_Object is communicated as «Artefact» in the interface with its own distinct icon. The result of a query provides a list of URIs that redirects to the source publication of the datasets (Marlet *et al.* [2019a](#); Hivert [2022](#) see also section 4.5 for an example).

In short, Openarcheo provides a data graph querying service in a simplified visual-based manner that makes use of a generic CIDOC CRM description that does not impose extensive semantic descriptions onto heterogeneous excavation datasets. In this respect, it provides one of the few existing examples in maintaining the original data structures, while allowing operational semantic interoperability.

#### 4.3.4. Meta-repositories and schema translation tools

As part of the data integration within the ARIADNEplus knowledge base, data mapping workflows for aligning completely different datasets employ the *X3ML toolkit*, which comprises a set of small, open-source software components for information integration. These include the [X3ML Mapping Definition Language](#), the [3M Mapping Memory Manager](#), the [X3ML Engine](#) and the [RDF Visualiser](#). The combined use of these tools facilitates a complete workflow for transforming XML exports of datasets/databases into CIDOC CRM compatible RDFs (Theodoridou and Kritsotakis [2022](#)). This process can be combined with tools for achieving vocabulary homogenisation such as the [Vocabulary Matching Tool](#) (VMT) by the University of Wales (accessed also from within the ARIADNEplus VRE services to map concepts to Getty's AAT).

An example of such an approach was provided by the mapping of the Paliambela Kolindros excavation dataset, the product of one of the first 3D GIS documentation workflows that included conceptual modelling using CIDOC CRM v.4.3 (Crofts *et al.* [2008](#)) to describe data components (Katsianis [2012](#); Katsianis *et al.* [2021](#)). The preparation of a partial and incomplete 3D geospatial data archive for deposition almost a decade later has made numerous challenges apparent, some of which included the necessity to update the previous semantic description using concepts









interactive visualisation and editorialisation of content, in an attempt to standardise digital-based archaeological reporting processes. The system is a hybrid solution that integrates the 3D visualisation and interaction components of [3DHOP](#) into the Content Management System [Omeka S](#). Using the excavation project of Västra Vång, Sweden, as a case study, researchers from DARKLab at Lund University and the National Institute of Art History in Paris (INHA) collaborated to provide a data structure for the system that covers different aspects of excavation documentation and data building, and is aligned with standard documentation procedures based on the widespread use of the INTRASIS system in Swedish Archaeology.

To ensure an easy future alignment with larger and international knowledge bases, such as ARIADNE, the data modelling included elements from the CRMbase and its extensions, CRMarchaeo, CRMdig, CRMSci to cover the basic entities of the model and their linkages. In addition, *Dublin Core* was employed to describe core metadata shared among the basic entities of the model, while *Schema.org* was employed to simplify the geographical allocations. Two further custom ontologies, *air* and *tdhop*, were developed to handle information about platform-specific entities and describe 3D visualisation parameters, respectively. Currently, controlled vocabularies are being aligned to common reference thesauri, such as the *AAT*, *Geonames* and *Pleiades* (Derudas *et al.* [2023](#); Derudas and Nurra [2022](#)).

However, mapping the Västra Vång model to the CRM was challenging, especially because of the use of a non event-based data model consisting of numerous entities (e.g. in the case of an 'archaeological context', seven tables describe in detail the different context typology metadata). To provide archaeologists with a comprehensive tool for use in fieldwork documentation and also for report writing, the entire heterogeneous dataset had to be encompassed, enabling access to and linkages between all of its components. In this respect, to connect entities that could not directly be linked in CIDOC CRM, custom shortcut properties were defined to be eventually aligned through unpacking the short-cutted connections (see Figure 13).

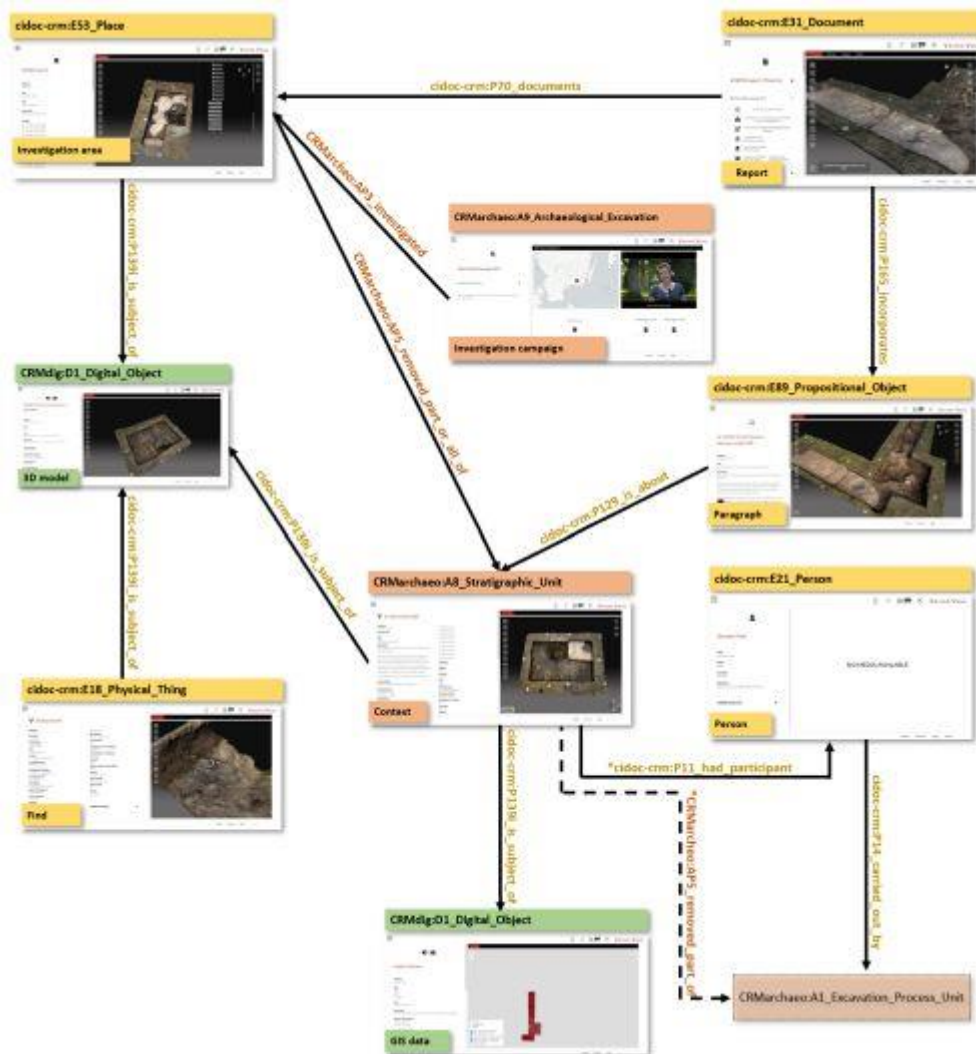


Figure 13: AIR conceptual data model: a selection of the entities and related semantic classes interconnected through properties. The different models used are indicated by colour. The dotted line indicates the connection with a 'bridging' class that adheres to CIDOC CRM (where in AIR a shortcut was used)

Although the model has not been implemented as an RDF Triplestore with a SPARQL endpoint, it showcases an extensive modelling depth that also takes into account issues of web page indexation as well as visualisation specific semantics (e.g. 3DHOP Scene), hence the need to include additional or custom-made reference ontologies. In many ways, it provides a useful example of the difficulties posed in application development and usability that stem from the fractalisation of data structures imposed by structured semantic descriptions (i.e. shortcutting long linkages) or the additional documentation requirements for different parameters of complex digital objects (related, for example, to the creation or the representation of GIS data and 3D models).

#### 4.3.6. Summary

The cases presented demonstrate that the current ecosystem of data modelling and knowledge organisation tools is advanced, but also far from standardised.



Depending on the structure and form of a given dataset (e.g. updatable content vs closed datasets) or the suite of digital tools that are employed in an implementation, some workflows may be more suitable than others. Work remains to be done in creating a clearer understanding and documentation of the kinds of use towards which semantic data is put and towards critically testing the tools and combination of tools that can be applied in different situations. For the more general uptake of such tooling, a more clear and generic elaboration of the workflows and documentation of the tools and methods that connect to them would be important areas of research. The purposeful further development of selected tools by the archaeological community and the explanation of the benefits or overlaps of specific tool sets with actual examples should also be pursued.

## 4.4. Learning and training

The appropriation and utility of semantic data, which depends on the user understanding the meaning of the data shown to them, ultimately depends on the adoption and apprehension of semantics by the intended target audience, in this case, field archaeologists. There is a fundamental challenge to meet here, given that the work of field archaeology and that of semantic data modelling are of highly different natures. To overcome this barrier, the provision of educational material, digital facilities and teaching opportunities for semantic modelling is key in order to enable an interested researcher to get up to speed. The production of such materials falls within the digital pedagogy, raising questions of digital literacy and its embeddedness in general archaeological curricula. The main goal here is to understand how to make the semantic strategy and its benefits accessible, reusable and ultimately attractive for reuse by its target audience in order to enable and further their archaeological research proper through the creation and integration of semantic datasets.

The main access point for those who want to approach the documentation and relevant material concerning the development of the CRM family of models, along with a history documenting the gradual solving of conceptual issues, can be found at the [CIDOC CRM](#) website. However, orienting oneself using the definition documents of each model and their documentation can be a difficult task for the uninitiated. The documentation is chiefly aimed at a semantic modelling audience, who already have a clear understanding of the concepts presented, rather than providing easy on-boarding documentation for those coming to CRM for the first time.

In order to address some of these shortcomings, several attempts have been made towards creating a more interactive engagement with the models for the learner and potential user during their familiarisation with the model concepts and properties. A very useful effort is outlined by the [CIDOC CRM periodic table](#), available in GitHub, which presents an interactive point of access to the CIDOC CRM base facilitating learning through experimentation. Further relevant resources concerning tool, services, workflows and training materials for the entire data life cycle of archaeological data can also be discovered in the [SSH Open Marketplace](#), a portal built by DARIAH to bring together, contextualise and make available a wide set of solutions and research practices as part of the *Social Sciences and Humanities Open Cloud project* (SSHOC).



Another very useful idea for increasing the awareness of entry-level practitioners is the [CIDOC CRM Game](#) created by Anaïs Guillem and George Bruseker, either as a tabletop or an online edition, [OntoMatchGame](#). The former is available in different versions (e.g. preventive archaeology or excavation editions) allowing for the relevant material (board and cards) to be downloaded, printed and cut to facilitate in person seminars and class tuition. The player can deploy the cards widely and brainstorm about building a complete model. The latter is a digital edition that can be played online to explore semantic modelling concepts and linkages (Figure 14). Its design is orientated towards guiding the user to discover what an ontology is, how CIDOC CRM works and how to make mappings. In this respect, the two versions of the game complement each other. Both editions provide an entry point for potential domain experts who want to get a first playful interaction with the mechanics and benefits of semantic structures (Guillem *et al.* [2018](#); Marlet [2022](#)).

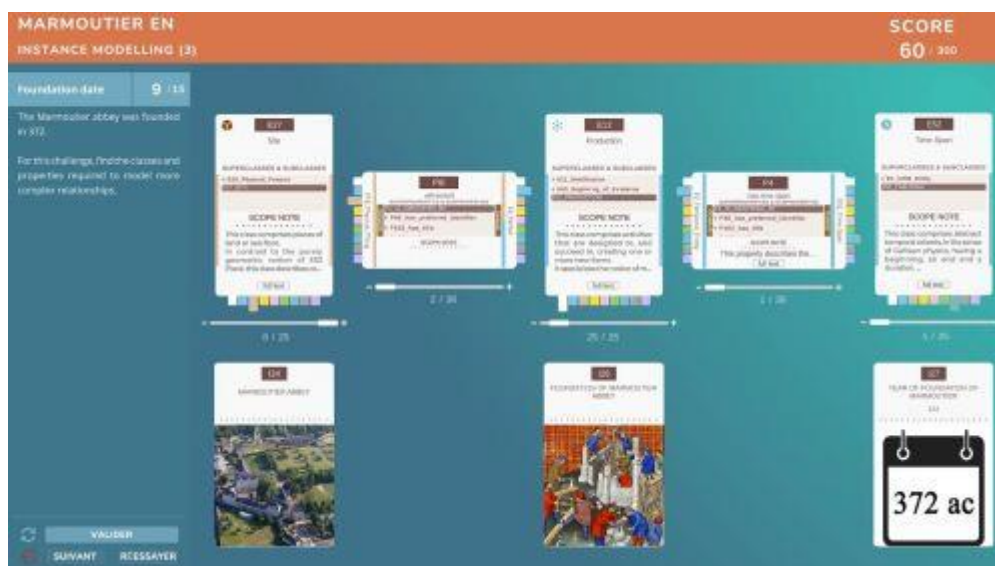


Figure 14: OntoMatchGame: a serious online game to learn ontologies and mapping, shown here featuring a game about the archaeological excavations of Marmoutier Abbey (France)

The emphasis on the identification of excavation data modelling patterns by the present team (see Section 4.2) could provide a further step in semantic pedagogy through their compilation into the form of a *cookbook* for intermediate-level data modellers, targeting the gradual or eventual creation of an archaeological marketplace, including data modelling examples and workflows, complemented by structured tutorials, educational videos and targeted workshops, such as the formal training events in several tools and methodologies (including introduction to data mapping) offered periodically by the [MASA consortium](#). Such intermediate and advanced training meetings can expand the range of modelling examples, bringing together diverse or complementary views of the excavation universe and informing both archaeological theory and data management methods.

Overall, there is still no systematic way for an archaeologist to develop a knowledge of the ontologies presented, but they must be learned through a variety of tools. The development of easier tools in order to understand the ontological standards is a good first step. The customisation of the material to address archaeological audiences is of prime importance in our context. The potential of moving forward on



the basis of the SRDM patterns presented in Section 4.2 to create rich, complex teaching material is one way in which future research could look at addressing the problem of embedding this sort of digital literacy into archaeological practice.

## 4.5. Semantic queries

The end goal of the adoption of semantic data strategies for archaeological information is to support archaeological research, the basis of which is the testing of hypotheses against empirical data. Thus, at the other end of the various efforts to develop models for representing archaeological excavation data, generating semantic datasets and learning how to create them, lies the ability to employ these datasets to respond to different kinds of archaeological queries. The task of thinking through what might form such competency questions, however, is not trivial. Moving from a local to a generic data structure with the possibilities of asking questions across excavation datasets entails thinking again about what we want to know from our data.

Would it be meaningful for example to query different excavations for Stratigraphic Units with a specific Munsell colour value? Probably not. The degree of meaningfulness of combining detailed data across different archaeological excavation datasets has been associated with the spatiotemporal extent of similarities identified in the archaeological record (Ore [2018](#)), but there can be many more potential associations with variable usability. Certainly, some may be very useful from a research perspective (e.g. Roman coins from different excavations), while others may be more useful from the perspective of data management in the context of commercial archaeology or national/regional heritage agencies (e.g. cemeteries/tombs excavated between 1950-1970). However, one cannot exclude the possibility of associations that may prove meaningful after addressing the result of a particular query.

To understand the potential of acquiring information from aggregated excavation datasets we need to think of and try out questions that researchers and heritage professionals would formulate in order to retrieve existing reusable data. This exercise can outline the depth of the data description that is required to explore possible associations and, from a practical side, can reveal what each question involves in terms of the respective semantic syntax. Here we include an example of what a complex question that returns item-level entries would look like in terms of the syntax involved and showcase its practical implementation using the visual query interface of Openarchaeo (Figure 15).

Question: I want to find pottery that has been retrieved from burials dating to the Mediaeval period.

Response: all records of pottery artefacts of pottery type that are associated with Mediaeval burials

Semantic syntax: E25\_Human\_Made\_Feature:

∨ P2\_has\_type → "burial"

∨ AP21\_contains → E22\_Human-Made\_Object → P101\_had\_as\_general\_use → "pottery"





∖ P94i\_was\_created\_by → E65\_Creation → P4\_has\_time-span → E52\_Time-Span ("Mediaeval"):  
∖ P82a\_begin\_of\_the\_begin → E61\_Time\_Primitive ("1050-01-01")  
∖ P82b\_end\_of\_the\_end → E61\_Time\_Primitive ("1536-12-31")

This example also showcases the way chronological periods are aligned with start-end numerical dates to facilitate queries across timelines. In essence, chronological terms are aligned with time-spans that can be further regulated for their spatio-temporal overlap through facilities, such as PeriodO (see Section 2).

The screenshot shows the OpenArcheo Explorer interface. At the top, it says "OpenArcheo Explorer" and "Explore". Below that, it lists "Sources queried: AERBA, ArSoL, Chronique des fouilles en ligne, Arkeogis - Hydromed, Arkeogis - Verduron, Conservatoire 3D, EpiCherchell, Base ICERAMM, Kiticon-Pervolia, OUTAGR, SOLIDAR".

The main part of the interface is a query builder with three rows of conditions:

- Row 1: Burial (icon) → composed of → Artifact (icon)
- Row 2: Where (icon) → And (icon) → Artifact (icon) → function is → Type (icon) → pottery (icon)
- Row 3: Burial (icon) → dated from → Time (icon) → De 1050 à 1536 (icon)

Below the query builder, there is an "EXECUTE QUERY" button and a dropdown menu set to "Table".

Below the query builder, it says "Showing 1 to 5 of 5 entries". There is a search bar and a dropdown menu set to "50" entries.

this	thisLabel
1 https://iceramm.huma-num.fr/bdresultats.php?notice=620/contexte	Dépôts funéraires de la phase 2
2 https://iceramm.huma-num.fr/bdresultats.php?notice=626/contexte	Ensemble 1
3 https://iceramm.huma-num.fr/bdresultats.php?notice=627/contexte	Ensemble 2
4 https://iceramm.huma-num.fr/bdresultats.php?notice=628/contexte	En cours
5 https://iceramm.huma-num.fr/bdresultats.php?notice=629/contexte	En cours

Figure 15: An example of a cross-dataset query executed in SPARNATURAL using the underlying semantics of Openarchaeo

Similar semantic queries testing the response performance of complex queries across datasets at the item level using SPARQL have also been performed in ARIADNEplus using the integrated data from [THANADOS](#), an online application of published archaeological and anthropological data of early medieval cemeteries in Austria (see Aspöck *et al.* 2023). This is rendered possible by accessing the ARIADNE knowledge base using the relevant GraphDB SPARQL endpoint functionality from the [ARIADNEplus Lab](#), developed to integrate different toolsets as accessible virtual services for setting up archaeological data processing, curation and retrieval workflows (Assante *et al.* 2022).



In any case, by posing questions and analysing their syntax we may succeed in identifying a minimum baseline for useful excavation data semantic descriptions. This exercise can help test the minimal and economical descriptions that can be standardised and reused in new data modelling pursuits such as those detailed in Sections 4.2 and 4.3. In addition, it can help test what an inclusive excavation domain description may mean in terms of basic concepts and linkages and how these can then relate to more exhaustive descriptions of the excavation domain or other related sub-domains. Finally, providing a set of predetermined question patterns that respond to typical archaeological questions and that can be used both within a specific dataset and across datasets would be a way of both making the semantic datasets generated more accessible and opening up new thinking on what it means to query across projects to reveal new, wider patterns, than those typically investigated based on single datasets.

## 5. Conclusions: Roadmap of Activities and Raising Awareness

It is clear from the discussion above that the overall field of semantic representation of archaeological excavation data is a lively and developing field, but we are clearly still some way from the seamless reusability of digital archaeological excavation datasets. The centrality of the excavation process as the most significant data generator of archaeological practice requires renewed attention in terms of its ontological significance and interfacing with sub-domain specific practices (like digital fieldwork documentation, laboratory studies and interpretative processes). The research undertaken by this group found that the present CIDOC CRM and official extensions alongside some proposed external extensions or application profiles are adequate for the raw task of semantic representation of data.

However, as shown here, the analytical work in the description of the excavation universe now needs to be complemented by synthetic attempts that can foster a more stable and user-friendly ecosystem of methods and tools for non-specialists to structure their data in meaningful and interoperable ways. To this end, we assessed conceptual models, modelling patterns, workflow and tools, learning and training and reusable queries to identify specific areas where further work in the standardisation of practice could lead to a more dynamic convergence of datasets and generate benefits for archaeological research.

Several concrete areas emerged as priorities for future research and standardisation. The ever-expanding universe of concepts, linkages and inheritances or possible deprecation of models may lead to increased difficulties in practical data modelling. Creating and sharing descriptions that cover how to generate typical data in different excavation stages or fields in an archaeologist-orientated way is an important development in semantic modelling patterns, which requires further testing and feedback from the community to assess its understandability and reusability. A more systematic analysis of the ways that data are interacted with and created in the semantic data context to identify the relations and dependencies between workflows and tools would be beneficial to many data users. This would enable consistent practices using particular tools to be built and reused by other practitioners. This



does not imply a fixed set of tools or workflows, but rather improved guidance on the main tasks involved and which tools are adequate to each task. Good guidance would be an important step toward the standardisation of data integration practices. To this end, further work targeting the *Synergy Reference Model* (Doerr *et al.* [2016b](#)) may bring the procedural requirements for data provision and aggregation into line.

Pedagogy around the use of semantic data in general is limited, and archaeology is particularly poorly provisioned. A patchwork of resources and tools offer a way to assemble an ad hoc knowledge of the semantic data process. Further work in this field could look to identify elements necessary to support a basic curriculum or curricula to support the archaeologist who wants to create or reuse semantic data. Many of the resources identified in this article could be more systematically deployed to that end. Finally, to provide convincing evidence that might persuade sceptics of semantic data, the elaboration of semantic data queries that both validate the representation and demonstrate how the integration of data opens new possibilities for research should be a priority. Again, further research is needed to achieve the necessary systematisation.

To return to the question posed at the beginning of this article, the jury is likely still out on the utility of the adoption of semantic data in archaeological excavation practice. This is not a result of the conceptual inadequacy of the programme or the proposed conceptual modelling constructs, which is the most technologically mature part of the ecosystem. Rather, the readiness for real-world use is held back by the ongoing fragmentation of efforts at integration of the wider conceptual and digital toolset into a real programme of action and study. Semantic technology and data proposals for archaeology find themselves at a moment of ferment, with many developments building off the platform provided by stable, standardised conceptual models for data representation. What is required now, as illustrated in the case studies presented here, is to consolidate this creative burgeoning toward a convergence of praxis in describing models, adhering to consistent workflows, generating engaging curricula to pass this knowledge into functional practice, and demonstrating the benefits and results through the creation, documentation and dissemination of suggestive queries.

## Acknowledgements

We would like to acknowledge the editors for the invitation to participate in this special issue. This work was implemented within the framework of ARIANDEplus, a Horizon 2020 project funded by the European Commission under Grant Agreement n. 823914. The views and opinions expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. An earlier version of this article was presented at the 28th EAA Annual Meeting in Budapest, Hungary 31 August 3 September 2022, in the session entitled 'FAIRly Front-loading the Archive: Moving beyond Findable, Accessible and Interoperable to Reuse of Archaeological Data'.



# Bibliography

Agresti, A., Maggiolo-Schettini, A., Saccoccio, R., Pierobon, M. and Pierobon-Benoit, R. 1996 'Handling excavation maps in SYSAND' in H. Kamermans and K. Fennema (eds) *Interfacing the Past. Computer Applications and Quantitative Methods in Archaeology 1995 (CAA 95)*, Leiden: University of Leiden. 31-36.

Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrilis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., Vlachidis, A. and Wright, H. 2017a 'Enabling European archaeological research: the ARIADNE E-Infrastructure', *Internet Archaeology* **43**. <https://doi.org/10.11141/ia.43.11>

Aloia, N., Debole, F., Felicetti, A., Galluccio, I. and Theodoridou, M. 2017b. 'Mapping the ARIADNE catalog data model to CIDOC CRM: bridging resource discovery and item-level access', *SCIRES-IT : SCientific RESearch and Information Technology* **7**, 1-8. <http://dx.doi.org/10.2423/i22394303v7n1p1>

Andresen, J. and Madsen, T. 1992 'Data structures for excavation recording: a case of complex information management' in C.U. Larsen (ed) *Sites and Monuments: National Archaeological Records*, Copenhagen: National Museum of Denmark. 49–67.

Arroyo-Bishop, D. and Lantada Zarzosa, M.T. 1995 'To be or not to be: will an object-space-time GIS/AIS become a scientific reality or end up an archaeological entity?' in G.R. Lock and Z. Stancic (eds) *Archaeology and Geographic Information Systems: A European Perspective*, London: Taylor & Francis. 43-53.

Ashley, M., Tringham, R. and Perlingieri, C. 2011 'Last house on the hill: digitally remediating data and media for preservation and access', *Journal on Computing and Cultural Heritage* **4**(4). <https://doi.org/10.1145/2050096.2050098>

Aspöck, E., Eichert, S., Theodoridou, M., Richards, N. and Felicetti, A. 2023 'Integrating Data on Early Medieval Graves: Mapping the THANADOS database to the ARIADNE infrastructure with the Mortuary Data Application Profile', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.11>

Assante, M., Cirillo, R., Dell'Amico, A. and Pagano, P. 2022 *D13.4 – VREs Operation Final Activity Report JRA2* (with contributions by Candela, L., Frosini, L., Lelii, L., Mangiacrapa, F., Panichi, G., Piccioli, T. and Sinibaldi, F.), ARIADNEplus, Zenodo. <https://doi.org/10.5281/zenodo.7509956>

Attewell, B., Cripps, P., Crosby, V., Cromwell, T., Graham, K., Heathcote, J., Jones, C., Lyons, E., May, K., May, S., Payne, A., Reilly, S., Robinson, D., Schuster, J., Stonell Walker, K. and Walkden, M. 2004 *Revelation: Phase 1 Assessment*, Centre for Archaeology Report **78/2004**. English Heritage. [https://historicensland.org.uk/research/results/reports/5382/Revelation\\_Phase1Assessment](https://historicensland.org.uk/research/results/reports/5382/Revelation_Phase1Assessment) [Last accessed: 20 March 2022].



Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S. and Velios, A. 2021 *Definition of the CIDOC Conceptual Reference Model*, Version 7.1.1. <https://doi.org/10.26225/FDZH-X261>

Beretta, F. 2021 'A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)', *Semantic Web* 12(2), 279-94. <https://doi.org/10.3233/SW-200416>

Bettess, F. 1984 *Surveying for Archaeologists*, University of Durham.

Bruseker, G., Carboni, N. and Guillem, A. 2017 'Cultural heritage data management: the role of formal ontology and CIDOC CRM' in M. Vincent, V. López-Menchero Bendicho, M. Ioannides and T. Levy (eds) *Heritage and Archaeology in the Digital Age. Quantitative Methods in the Humanities and Social Sciences*, Springer, Cham. [https://doi.org/10.1007/978-3-319-65370-9\\_6](https://doi.org/10.1007/978-3-319-65370-9_6)

Chadwick, A. 1997. 'Archaeology at the edge of chaos: further towards reflexive excavation methodologies', *assemblage* 3. <https://archaeologydataservice.ac.uk/archives/view/assemblage/html/3/3chad.html> [Last accessed: 20 March 2022].

Chenhall, R.G. 1971 'The archaeological data bank: a progress report', *Computers and the Humanities* 5(3), 159-69. <http://www.jstor.org/stable/30199402> [Last accessed: 20 March 2022].

Christaki, E., Doerr, M., Bekiari, C. and Bruseker, G. 2017 'Μεθοδολογία μοντελοποίησης ανασκαφικής διαδικασίας με σωζόμενα αρχιτεκτονικά λείψανα' in G. Vavouranakis, M. Katsianis, Y. Papadatos, M. Mouliou and P. Petridis (eds) *Digital Pasts for the Present. Proceedings of the 2nd conference on Computer Applications and Quantitative Methods in Archaeology Greek Chapter (CAA-GR). Athens, 22-23/12/2016*, Athens: Department of History and Archaeology. National and Kapodistrian University of Athens. 26-35.

Cooper, A. and Green, C. 2017 'Big questions for large, complex datasets: approaching time and space using composite object assemblages', *Internet Archaeology* 45. <https://doi.org/10.11141/ia.45.1>

Cripps, P. and May, K. 2010 'To OO or not to OO? Revelations from ontological modelling of an archaeological information system' in F. Niccolucci and S. Hermon (eds) *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004, Prato 13–17 April 2004*, Budapest: Archaeolingua Foundation. 59-63. [https://proceedings.caaconference.org/files/2004/08\\_Cripps\\_May\\_CAA\\_2004.pdf](https://proceedings.caaconference.org/files/2004/08_Cripps_May_CAA_2004.pdf) [Last accessed: 20 March 2022].

Cripps, P., Greenhalgh, A., Fellows, D., May, K. and Robinson, D. 2004 *Ontological Modelling of the Work of the Centre for Archaeology*, CIDOC CRM Technical Paper. [https://CIDOC-CRM.org/sites/default/files/cidoc\\_crm\\_version\\_4.3\\_Nov08.pdf](https://CIDOC-CRM.org/sites/default/files/cidoc_crm_version_4.3_Nov08.pdf) [Last accessed: 20 March 2022].





Crofts, N., Doerr, M., Gill, T., Stead, S. and Stiff, M. 2008 *Definition of the CIDOC Conceptual Reference Model*, Version 4.3. [http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_4.3\\_Nov08.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.3_Nov08.pdf) [Last accessed: 20 March 2022].

Derudas, P., Nurra, F. and Svensson, A. 2023 'New AIR for the archaeological process? The use of 3D web semantic for publishing archaeological reports', *Journal on Computing and Cultural Heritage* **16**(3), 1-23. <https://doi.org/10.1145/3594722>

Diez Castillo, A. and Martinez Burgos, C. 2001 'SIDGEIPA: an archaeological information system' in W. Börner and L. Dollhofer (eds) *Archäologie und Computer*, Workshop 6, Vienna, 5-6/11/2001. <http://hdl.handle.net/10550/57059>.

Doerr, M. 2003 'The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata', *AI Magazine* **24**, 75–92.

Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., Theodoridou, M. *et al.* 2012 *CRMsci: the Scientific Observation Model An Extension of CIDOC-CRM to Support Scientific Observation*, Version 1.2. <https://cidoc-crm.org/crmsci/sites/default/files/CRMsci1.2.pdf> [Last accessed: 20 March 2022].

Doerr, M., Achille F., Sorin H., Hiebel, G., Kritsotaki, A., Masur, A., May, K., Ronzino, P., Schmidle, W., Theodoridou, M., Tsiadaki, D. *et al.* 2016a *Definition of the CRMarchaeo: An Extension of CIDOC CRM to support the archaeological excavation process. Technical Report 1.4*, Prato, Italy: PIN S.c.R.L. [https://cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo\\_v1.4.1.pdf](https://cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.1.pdf) [Last accessed: 20 March 2022].

Doerr, M., Achille F., de Jong, G., Konsolaki, K., Norton, B., Oldman, D., Theodoridou, M. and Wikman, T. 2016b *The Synergy Reference Model of Data Provision and Aggregation*, Draft 1.5. [https://cidoc-crm.org/sites/default/files/SRM\\_v1.5.pdf](https://cidoc-crm.org/sites/default/files/SRM_v1.5.pdf) [Last accessed: 20 March 2022].

EARTO 2014 *The TRL Scale as a Research & Innovation Policy Tool*, EARTO Recommendations. [https://www.earto.eu/wp-content/uploads/The\\_TRL\\_Scale\\_as\\_a\\_R\\_I\\_Policy\\_Tool\\_-\\_EARTO\\_Recommendations\\_-\\_Final.pdf](https://www.earto.eu/wp-content/uploads/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf) [Last accessed: 20 March 2022].

European Commission (EC) 2014 'Annex G. Technology readiness levels (TRL)', HORIZON 2020 – WORK PROGRAMME 2014-2015 General Annexes. [https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014\\_2015/annexes/h2020-wp1415-annex-g-trl\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf) [Last accessed: 20 March 2022].

Felicetti, A., Gerth, P., Meghini, C. and Theodoridou, M. 2015 'Integrating heterogeneous coin datasets in the context of archaeological research' in P. Ronzino and F. Niccolucci (eds) *Extending, Mapping and Focusing the CIDOC CRM (CRMEX 2015) Workshop, 19th International Conference on Theory and Practice of Digital Libraries (TPDL 2015)*, Poznan, Poland, September 17, 2015. <https://ceur-ws.org/Vol-1656/paper2.pdf>



Felicetti A., Hiebel G., Katsianis M., Ore C.E. (eds) Christaki E., Doerr M., Hermon S., Kritsotaki A., Masur A., May K., Ronzino P., Schmidle W., Theodoridou M., Tsiafaki D., et al. (contr.) 2023 *Definition of the CRMarchaeo: An Extension of CIDOC CRM to support the archaeological excavation process*, Version 2.0, Prato, Italy: PIN S.c.R.L. <https://cidoc-crm.org/crmarchaeo/ModelVersion/crmarchaeo-version-2.0> [Last accessed: 14 October 2023]

Ferdière, A. 2023 'Gauthier E., Georges-Leroy M., Poirier N. et Weller O. (dir.) - *Archæodyn. Dynamique spatiale des territoires de la Préhistoire au Moyen Âge*, vol. 1', *Revue archéologique du Centre de la France* **62**. <http://journals.openedition.org/raef/5849>

Gardin, J.-C. 1980 *Archaeological Constructs: An aspect of theoretical archaeology*, Cambridge; New York: Cambridge University Press.

Giagkoudi, E., Tsiafaki, D. and Papatheodorou, C. 2018 'Describing and revealing the semantics of excavation notebooks', *CIDOC 2018 Heraklion, Crete, Greece*. [https://users.ionio.gr/~papatheodor/papers/CIDOC2018\\_paper\\_71.pdf](https://users.ionio.gr/~papatheodor/papers/CIDOC2018_paper_71.pdf) [Last accessed: 20 March 2022].

Guillem, A., Bruseker, G. and Lercari, N. 2018 'Learning through doing: an online game for appropriating ontological modelling methodology', Paper presented at 46th Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2018), Tübingen, Mar. 19-23, 2018.

Hacıgüzeller, P., Taylor, J.S. and Perry, S. 2021 'On the emerging supremacy of structured digital data in archaeology: a preliminary assessment of information, knowledge and wisdom left behind', *Open Archaeology* **7**(1), 1709-30. <https://doi.org/10.1515/opar-2020-0220>.

Hadzilakos, T. and Stoumbou, P.M. 1996 'Conceptual data modelling for prehistoric excavation documentation' in H. Kamermans and K. Fennema (eds) *Interfacing the Past: Computer Applications and Quantitative Methods in Archaeology 1995* (CAA 95), Leiden: University of Leiden. 21-30.

Harris, E.C. 1979 *Principles of Archaeological Stratigraphy*, London and New York: Academic Press.

Hiebel, G., Doerr, M., Hanke, K. and Masur, A. 2014 'How to put archaeological geometric data into context? Representing mining history research with CIDOC CRM and extensions', *International Journal of Heritage in the Digital Era* **3**(3), 557–77. <https://doi.org/10.1260/2047-4970.3.3.557>

Hiebel, G., Aspöck, E. and Kucera, M. 2017 'Excavation interface in CRMarchaeo?', *Presentation at the 38th joint meeting of the CIDOC CRM SIG and ISO/TC46/SC4/WG9 and the 31st FRBR - CIDOC CRM Harmonization meeting*. [https://www.cidoc-crm.org/sites/default/files/Excavation\\_Inference4CRMarchaeo.pdf](https://www.cidoc-crm.org/sites/default/files/Excavation_Inference4CRMarchaeo.pdf) [Last accessed: 20 March 2022].



Hiebel, G., Aspöck, E. and Kopetzky, K. 2021 'Ontological modeling for excavation documentation and virtual reconstruction of an Ancient Egyptian site', *Journal on Computing and Cultural Heritage* **14**(3). <https://doi.org/10.1145/3439735>

Hiebel, G., Danthine, B., Peralta Friedburg, M. and Scherer-Windisch, M. 2023 'Prehistoric Mining Data: How to create Open Data from archaeological research for the ARIADNE community and beyond', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.8>

Jones, A. 2002 *Archaeological Theory and Scientific Practice*, Cambridge: Cambridge University Press.

Kansa, E. 2005 'A community approach to data integration: authorship and building meaningful links across diverse archaeological data sets', *Geosphere* **1**(2), 97-109. <https://doi.org/10.1130/GES00013.1>.

Katsianis, M. 2012 'Conceptual and practical issues in the use of GIS for archaeological excavations' in A. Chrysanthi, M.P. Flores and C. Papadopoulos (eds) *Thinking Beyond the Tool: Archaeological Computing and the Interpretive Process*, BAR International Series **2344**, Oxford: Archaeopress. 51-71.

Katsianis, M., Kotsakis, K. and Stefanou, F. 2021 'Reconfiguring the 3D excavation archive. Technological shift and data remix in the archaeological project of Paliambela Kolindros, Greece', *Journal of Archaeological Science: Reports* **36**, 102857. <https://doi.org/10.1016/j.jasrep.2021.102857>.

Katsianis, M., Styliaras, G., Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Kritsotakis, V., Martlet, O., Nenova, D., Nurra F., Ore, C-E. and Theodoridou, M. 2022a *ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological Excavation Data*, ARIADNE Project, Zenodo. <https://doi.org/10.5281/zenodo.7112918> [Last accessed: 20 March 2022].

Katsianis, M., Nenova, D. Bruseker, G., Derudas, P., Felicetti, A., Hiebel, G., Hivert, F., Martlet, O., Opitz, R., Richards, J., Roulet, T., Styliaras, G., Ore, C-E. and Uleberg, E. 2022b *Archaeological Excavation Modelling Working Group: WP 4.4.12 excavation data*, ARIADNE Project, Zenodo. <https://doi.org/10.5281/zenodo.7377910> [Last accessed: 20 March 2022].

Löwenborg, D., Jonsson, M., Larsson, A. and Nordinge, J. 2021 'A turn towards the digital. An overview of Swedish heritage information management today', *Internet Archaeology* **58**. <https://doi.org/10.11141/ia.58.19>.

Lucas, G. 2000 *Critical Approaches to Fieldwork: Contemporary and Historical Archaeological Practice*, 1st edition, Routledge. <https://doi.org/10.4324/9780203132258>.

Madsen, T. 1999 'Coping with complexity. Towards a formalised methodology of contextual archaeology', *Archaeologia e Calcolatori* **10**, 125-44.



Madsen, T. 2003 'Archaeoinfo: an object-oriented information system for archaeological excavations' in Magistrat der Stadt Wien - Referat Kulturelles Erbe - Stadtarchäologie Wien (ed) [*Enter the Past*]. *The E-way into the four Dimensions of Cultural Heritage, Computer Applications and Quantitative Methods in Archaeology (CAA 2003)*, Wien. <https://www.archaeoinfo.dk/PDF/CAA2003.pdf> [Last accessed: 20 March 2022].

Marlet, O., Francart, T., Markhoff, B. and Rodier, X. 2019a 'OpenArchaeo for usable semantic interoperability' in A. Poggi (ed) *Proceedings of First International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH) co-located with the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019)*, Rome, Italy, June 3, 2019, Rome: Sapienza University of Rome. <http://ceur-ws.org/Vol-2375/paper1.pdf> [Last accessed: 20 March 2022].

Marlet, O., Zadora-Rio, E., Buard, P.-Y., Markhoff, B. and Rodier, X. 2019b 'The archaeological excavation report of Rigny: an example of an interoperable logicist publication', *Heritage* 2(1), 761-73. <https://doi.org/10.3390/heritage2010049>.

Masur, A., Keith M., Hiebel G. and Aspöck, E. 2014 'Comparing and mapping archaeological excavation data from different recording systems for integration using ontologies' in *18th International Conference on Cultural Heritage and New Technologies (CHNT18)*, Vienna, November 11-13, 2013. [https://archiv.chnt.at/wp-content/uploads/Masur\\_etal\\_2014.pdf](https://archiv.chnt.at/wp-content/uploads/Masur_etal_2014.pdf). [Last accessed: 20 March 2022].

May, K., Binding, C. and Tudhope, D. 2011 'A STAR is born: some emerging semantic technologies for archaeological resources' in E. Jerem, F. Redö and V. Szeverényi (eds) *On the Road to Reconstructing the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 36th International Conference. Budapest, April 2-6, 2008*, Budapest: Archaeolingua Foundation. 111-16. [https://proceedings.caaconference.org/files/2008/CD53\\_May\\_et\\_al\\_CAA2008.pdf](https://proceedings.caaconference.org/files/2008/CD53_May_et_al_CAA2008.pdf) [Last accessed: 20 March 2022].

Nawaf, M., Drap, P., Ben-Ellefi, M., Nocerino, E., Chemisky, B., Chassaing, T., Colpani, A., Noumossie, V., Hyttinen, K., Wood, J., Gambin, T. and Sourisseau, J.C. 2021 'Using virtual or augmented reality for the time-based study of complex underwater archaeological excavations', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VIII-M-1-2021, 117-24. <https://doi.org/10.5194/isprs-annals-VIII-M-1-2021-117-2021>

Nenova, D., Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Katsianis, M., Marlet, O., Opitz, R., Ore, C.-E. and Uleberg, E. 2022 'Bringing excavation data together. Are we there yet and where is that?', paper presented at *28th EAA Annual Meeting (EAA 2022)*, Budapest. <https://doi.org/10.5281/zenodo.7117049>. [Last accessed: 20 March 2022].

Nenova, D., Bruseker, G., de Haas, T. and van Leusen, M. 2023 *CRMsurv Ontology Specification*, Version 1, Zenodo. <https://doi.org/10.5281/zenodo.7649986>



Niccolucci, F. and Richards, J. 2019 'ARIADNE and ARIADNEplus' in J. Richards and F. Niccolucci (eds) *The ARIADNE Impact*, Budapest: Archaeolingua Foundation. <https://zenodo.org/record/4319058>

Ore, C.-E. 2018 'Oceans of data: creating a safe haven for information' in M. Matsumoto and E. Uleberg (eds) *Oceans of Data. Proceedings of the 44th Annual Conference on Computer Applications and Quantitative Methods in Archaeology 2018*, 3-9. <https://www.duo.uio.no/handle/10852/66862> [Last accessed: 20 March 2022].

Patterson, H., Witcher, R.E. and Di Giuseppe, H. 2020 *The Changing Landscapes of Rome's Northern Hinterland: the British School at Rome's Tiber Valley Project*, Archaeopress Roman Archaeology 70, Oxford: Archaeopress. <https://www.archaeopress.com/Archaeopress/download/9781789696158> [Last accessed: 20 March 2022].

Rains, M. 1995 'Towards a computerised desktop: the integrated archaeological database system' in J. Huggett and N. Ryan (eds) *Computer Applications and Quantitative Methods in Archaeology 1994*, Oxford: Tempus Reparatum. 207-10.

Richards, J.D. 1997 'Preservation and re-use of digital data: the role of the Archaeology Data Service', *Antiquity* **71**, 1057-59. <https://doi.org/10.1017/S0003598X00086014>

Richards, J.D. 2023 'Joined up Thinking: Aggregating archaeological datasets at an international scale', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.3>

Richards, J., Felicetti, A., Meghini, C. and Theodoridou, M. 2022 *D4.4 – Final Report on Ontology Implementation* (with contributions by Aspöck, E., Novak, D., Jensen, P., Katsianis, M., Rossenbach, K.-S., Santos Silva, A., van Leusen, M., Buckland, P. and Stular, B.), ARIADNEplus, Zenodo. <https://doi.org/10.5281/zenodo.7636720>

Ronzino, P. 2017 'Harmonizing the CRMba and CRMarchaeo models', *International Journal of Digital Libraries* **18**, 253-261. <https://doi.org/10.1007/s00799-016-0193-3>.

Theodoridou, M. 2022 *D14.2 – Final Report on the ARIADNEplus Knowledge Management System* (with contributions by Kritsotakis, V., Felicetti, A., Tudhope, D., Binding, C. and Aspöck, E.), ARIADNEplus, Zenodo. <https://doi.org/10.5281/zenodo.7651281>

Tudhope, D. and Binding, C. 2013 'CRMEX position paper: mapping patterns for CIDOC CRM' in V. Alexiev, V. Ivanov and M. Grinberg (eds) *Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) Workshop, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), 26 September 2013, Valletta, Malta*. <https://ceur-ws.org/Vol-1117/paper10.pdf> [Last accessed: 20 March 2022].

Tzitzikas, Y., Kampouraki, M. and Analyti, A. 2014 'Curating the specificity of ontological descriptions under ontology evolution', *Journal on Data Semantics* **3**, 75-106. <https://doi.org/10.1007/s13740-013-0027-z>





Tzitzikas, Y., Mountantonakis, M., Fafalios, P. and Marketakis, Y. 2022 'CIDOC-CRM and machine learning: a survey and future research', *Heritage* **5**, 1612-36. <https://doi.org/10.3390/heritage5030084>

Wand, Y. and Weber, R. 1993 'On the ontological expressiveness of information systems analysis and design grammars', *Information Systems Journal* **3**, 217-37. <https://doi.org/10.1111/j.1365-2575.1993.tb00127.x>

Westman A. (ed) 1994 *Site Manual*, London: MoLAS. <https://behindthetrowel.files.wordpress.com/2018/04/molasmanual94.pdf>

Wise, A. and Miller, P. 1997 'Why metadata matters in archaeology', *Internet Archaeology* **2**. <https://doi.org/10.11141/ia.2.5> [Last accessed: 20 March 2022].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva, Bonino, Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenberg, P., Wolstencroft, K., Zhao, J. and Mons, B. 2016 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data* **3**, 160018. <https://doi.org/10.1038/sdata.2016.18>

Zhang, C., Cao, C., Gu, F. and Si, J. 2002 'A domain-specific formal ontology for archaeological knowledge sharing and reusing' in D. Karagiannis and U. Reimer (eds) *Practical Aspects of Knowledge Management*, PAKM 2002, Lecture Notes in Computer Science **2569**, Berlin/Heidelberg: Springer. [https://doi.org/10.1007/3-540-36277-0\\_20](https://doi.org/10.1007/3-540-36277-0_20)

## ***ARIADNEplus Virtual Meeting on excavation modelling presentations***

Bruseker, G. 2022 'Semantic data modelling and archaeological research data - why, how and where we are now', *ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological Excavation Data*, 15 June 2022, Zenodo. <https://doi.org/10.5281/zenodo.7108479> (Presentation video accessible at <https://vimeo.com/753115015>) [Last accessed: 20 March 2022].

Derudas, P. and Nurra, F. 2022 'Archaeological interactive report: a trait d'union between data management and semantic publication', *ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological Excavation Data*, 15 June 2022, Zenodo. <https://doi.org/10.5281/zenodo.7108723> (Presentation video accessible at <https://vimeo.com/753115206>) [Last accessed: 20 March 2022].

Hiebel, G. 2022 'An approach to model archaeological data and create RDF from spreadsheets', *ARIADNEplus Virtual Workshop on Semantic Mapping of*



*Archaeological Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108589> (Presentation video accessible  
at <https://vimeo.com/753115377>) [Last accessed: 20 March 2022].

Hivert, F. 2022 'OpenArchaeo: a semantic Web platform for archaeological data',  
in *ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological  
Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108606> (Presentation video accessible  
at <https://vimeo.com/753115552>) [Last accessed: 20 March 2022].

Katsianis, M. and Styliaras, G. 2022 'Reworking aged excavation mappings with new  
models and tools', *ARIADNEplus Virtual Workshop on Semantic Mapping of  
Archaeological Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108671> (Presentation video accessible  
at <https://vimeo.com/753115671>) [Last accessed: 20 March 2022].

Marlet, O. 2022 'From modelling to mappings: how to appropriate the CIDOC  
CRM', *ARIADNEplus Virtual Workshop on Semantic Mapping of Archaeological  
Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108493> (Presentation video accessible  
at <https://vimeo.com/753115841>) [Last accessed: 20 March 2022].

Nenova, D. 2022 'Modelling archaeological excavations. Theoretical patterns and  
practical recipes', *ARIADNEplus Virtual Workshop on Semantic Mapping of  
Archaeological Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108645> (Presentation video accessible  
at <https://vimeo.com/753115940>) [Last accessed: 20 March 2022].

Ore, C.-E. 2022. '597 Norwegian excavation databases and CIDOC CRM - a  
practical exercise', *ARIADNEplus Virtual Workshop on Semantic Mapping of  
Archaeological Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108712> (Presentation video accessible  
at <https://vimeo.com/753116099>) [Last accessed: 20 March 2022].

Theodoridou, M. and Kritsotakis, V. 2022 'The X3ML toolkit: how to map excavation  
data to CIDOC CRM', *ARIADNEplus Virtual Workshop on Semantic Mapping of  
Archaeological Excavation Data*, 15 June 2022,  
Zenodo. <https://doi.org/10.5281/zenodo.7108577> (Presentation video accessible  
at <https://vimeo.com/753116253>) [Last accessed: 20 March 2022].