Internal Final Report

Archaeological Excavation Modelling Working Group

WP 4.4.12 excavation data

Version 1.0

24.11.2022

Grant Agreement number:	823914
Project acronym:	ARIADNEplus
Project title:	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe - plus
Funding Scheme:	H2020-INFRAIA-2018-1
Project coordinator name, Title and Organisation:	Prof. Franco Niccolucci, PIN Scrl - Polo Universitario "Città di Prato"
Tel:	+39 0574 602578
E-mail:	franco.niccolucci@pin.unifi.it
Project website address:	www.ariadne-infrastructure.eu

The research leading to these results has received funding from the European Community's Horizon 2020 Programme (H2020-INFRAIA-2018-1) under grant agreement n° 823914.

Editors: **Markos Katsianis, PP**

Denitsa Nenova, Takin.solutions

Contributing partners: **George Bruseker, Takin.solutions**

Paola Derudas, LU

Achille Felicetti, PIN

Gerald Hiebel, UIBK

Florian Hivert, UT

Markos Katsianis, PP

Olivier Marlet, CNRS

Denitsa Nenova, Takin.solutions

Rachel Opitz, UG

Julian Richards, ADS

Theo Roulet, UT

Giorgos Styliaras, PP

Christian-Emil Smith Ore, KHM-UO

Espen Uleberg, KHM-UO

Document History

- 4.10.2022 – Draft Version 0.1
- 15.11.2022 - Draft Version 0.5
- 24.11.2022 - Final version 1.0

Table of Contents

1. Executive Summary ***4***

2. Research by the group ***5***

2.1 Background and reasons for forming the group *5*

2.2 Group meetings and research progress *6*

2.3 Virtual Meeting organisation *11*

2.4 Presentations by the group *13*

3. Research results ***13***

4. Relevant work by individual partners ***18***

5. Conclusions – Next steps ***24***

6. References ***25***

Annexes ***26***

1. Executive Summary

The present report outlines and summarises the activity of the *Archaeological Excavation Modelling Working Group*, a sub-group that was formed within WP 4.4.12. The group has been active since June 2020. The group was formed to investigate the potential of developing an Application Profile for excavation data, explore the current state of excavation data modelling and propose a roadmap for further activities. During the ARIADNEplus project, group participants have convened eight times, organised a virtual workshop on excavation data modelling and prepared two presentations, one at an ARIADNEplus meeting and another targeting wider audiences (EAA2022). The entire work of the group is summarised in the present report and attached annexes.

2. Research by the group

The working party on CIDOC CRM mapping for excavation archives and intra-site research was formed in May 2020. The following people/partner institutions have participated or included in the group's communication: J. Birzescu (IAVP), G. Bruseker & D. Nenova (Takin.solutions), P. Derudas & N. Dell'Unto (UL), A. Felicetti (PIN), G. Hiebel (UIBK), F. Hivert, B. Markhoff & T. Roulet (UT), M. Katsianis & G. Styliaras (PP), D. Löwenborg (UU), O. Marlet & X. Rodier (CNRS), K. May (EH), Rachel Opitz (UG), C.E.S. Ore & E. Uleberg (KHM-UO), J. Richards (ADS), K. S. Rossenbach (INRAP)[1].

The group convened virtually eight (8) times between June 2020 and September 2022. The Meetings minutes were circulated among participants. Meetings 1-4 explored the suitability of an Application Profile (AP) for excavation data. Meetings 5-6 experimented with excavation data modelling. Meeting 7 was crucial in directing research away from the need for a new AP for excavation data and further towards planning the Virtual workshop. Finally, meeting 8 provided the opportunity to summarise, wrap up activities and set up future directions. In between meetings, several communication channels were used to coordinate the group and exchange modelling examples and ideas, such as shared documents, collaborative editing, basecamp and email threads.

2.1 Background and reasons for forming the group

As part of the data aggregation activities within the ARIADNEplus project, several partners working with excavation data had individual or groups of excavation datasets or were working on excavation data modelling to approach the item-level integration of their datasets within larger aggregation structures, like the ARIADNEplus portal. Meanwhile, as part of WP4, several sub-domains were working in the same direction through the study and development of a series of domain-specific Application Profiles (APs) that could link the AO-Cat model, designed to describe archaeological datasets at the collection level, with more specific descriptions of data at the sub-collection level. On several occasions (e.g. burial data), these activities focused on testing their applicability to provide further enhancements and extensions.

With respect to the excavation domain, the combined or collateral usage of the concepts included in CIDOC CRM and its domain-specific extensions have been employed to provide representations of the excavation process - an application domain notorious for its difficulty in achieving interoperable datasets. In this respect, the group was formed to investigate the possibility of creating an Application Profile for excavation data or explore the degree to which existing elements in the wider CIDOC CRM family of models would suffice to provide semantic descriptions that would allow the sub-collection or item-level integration of excavation dataset examples.

[1] All participants are included in alphabetical order (grouped by institute).

The group was formed after communication between interested partners in June 2020 and has since been steered by PP (Markos Katsianis).

2.2 Group meetings and research progress

A series of meetings were organised throughout the group's activities to facilitate discussion and group steering. This section provides a summary to provide a timeline of the group's research progress and detailed meetings.

- At the first online meeting (16/6/2020), P. Derudas, N. Dell'Unto (LU) and O. Marlet (CNRS) joined M. Katsianis (PP). Participants discussed their previous and current work on excavation data modelling. Through the discussion of particular examples, it was realised that there are different levels of research overlap. Katsianis were using a bottom-up approach to built-up the excavation archive from within a 3D GIS environment, Derudas and Dell' Unto employed a top-down approach to breaking down excavation reports to its 3D digital constituents and their complementary information, while Marlet had been working on linking primary and derived information through the mapping of the inferential process that employs the deconstruction of higher-level synthetic propositions through hypothetico-deductive reasoning into the primary evidence on which each proposition is based, thus allowing us to trace back to the field observation.

  The group proposed to experiment with existing datasets using two possible methods:

  a) Attempt to map the written configuration of an example site/season report using CIDOC CRM modelling elements and compare it to modelled examples of actual data blocks. The comparison of the two modelling views of the same data (top-down/bottom-up, written/structured, summarised/catalogued) could identify elements that are pronounced or common in each description and those hidden or truncated (shortcutted). These mappings could afterwards be compared to pinpoint the common features and elements that function as key nodes in the excavation knowledge-building process.
  b) Examine the available datasets and try using the family of CIDOC CRM-compatible models to describe the entities and their relations at the file or table level. By mapping existing data structures in their current form, comparing different solutions to data structuring could show which entities/properties are obscured and which are promoted. This exercise could help in the understanding of the implications of connecting data with different granularity and deposition requirements from a semantic perspective.

- The second meeting (15/12/2020) was attended by P. Derudas, N. Dell'Unto (LU), M. Katsianis (PP), O. Marlet, X. Rodier (CNRS), and C.E.S. Ore & E. Uleberg (KHM-UO). Preliminary testing of data mappings examples from the previous meeting was employed to provide the basic elements that could inform a potential AP for excavation data. A first draft was shared between group members and reviewed with respect to the targeting of a potential excavation domain

AP. Participants discussed their examples (a 3D publication platform by Derudas, the Paliambela Kolindros excavation archive by Katsianis, the ADED - Archaeological Digital Excavation Documentation project by Uleberg and Ore, Logicist-based analysis of excavation reports by Rodier). The group agreed on furthering our collective effort by testing the generic description of and overall applicability of the draft's main ideas with example data and revising/rebuilding the current text with comments and amendments.

- The third meeting (27/1/2021) was participated in by Iulian Birzescu (IAVP), P. Derudas, N. Dell'Unto (LU), F. Hivert (UT), M. Katsianis (PP), O. Marlet (CNRS), J. Richards (York, UK) and E. Uleberg (KHM-UO). The discussion revolved around the role of APs within the overall ARIADNEplus architecture and their potential usability concerning the development of relevant Virtual Research Environments (VREs). The group members shared their experience with data modelling, commenting on possible datasets that could be used to test the AP under development. The creation of an application profile should target the description of relations between sub-datasets, the connection of individual files (e.g. excavation images, shapefiles) to their actual documentation within database structures or individual data tables and the incorporation of elements of data lineage to differentiate between different versions of the same data. The group decided to proceed with editing the current draft accessible and include AP elements to describe individual test cases.

- The fourth meeting (15/4/2021) was joined by F. Hivert, B. Markhoff, T. Roulet (UT), M. Katsianis (PP), C.E.S. Ore, E. Uleberg (KHM-UO) and J. Richards (York, UK). In this meeting, the discussion focused on the necessary tools to create mappings and share these to produce comparisons. Again group members shared their experience using existing tools like the 3M, their functionality and limitations (e.g. multiple instantiations of concepts, foreign key handling, the inclusion of several models, like the CRMgeo and CRMinf in potential descriptions). Gradually, it was apparent that comparable modelling elements between individual mappings could be used to provide sets of modelling patterns. This approach was considered an interesting way to build up an application profile for excavation and post-excavation research that could be generic enough at top-level descriptions, but could open up to more domain-specific workflows, such as digital recording, data aggregation, sample processing, data analysis) using modelling elements from different CRM extensions. The problem of communicating modelling examples between the group was dealt with using simple diagram-based tools, like diagrams.net.

- The fifth meeting (28/6/2021) included further members that joined the group as affiliated partners of the ARIADNEplus, namely G. Hiebel (UIBK) and G. Bruseker, D. Nenova (Takin.solutions), M. Katsianis (PP), T. Roulet (UT), O. Marlet (CNRS), and E. Uleberg (KHM-UO) welcomed and briefed the new members about the work until that point.

  The scope of the draft document regarding the application profile proposal was reviewed and recognised as a very initial top-level generic description in need of being specified at a middle level, i.e. the actual target application profile description. During the discussion, several important points were raised, such as the need to break down the draft's generic description into more expanded and specified modelling parts that would allow the detection of the minimum units of information considered important in the description of excavation datasets and workflows. Another point was raised concerning the prioritisation between excavation data modelling and excavation dataset mapping with respect to actual data aggregation within the ARIADNEplus portal. The problem of sharing data models was also highlighted, especially in the absence of existing RDFs that could be used as examples, and as a solution, it was deemed more plausible to start by sharing a simple google spreadsheet document with the basic fields that need to be included in order to create semantic descriptions at the middle level. Smith-Ore's document on the INTRASIS alignment with CIDOC CRM contained an implementation that could inform the shared template for adding data mappings (Fig.1).

Meta-Id	Relation name	CDOC-CRM domain	Intrasis parent	CIDOC CRM properties	CIDOC CRM range	INTRASIS child
495	Has/is taken from	A7 Embedding	000015 Find unit	O4i sampled at (was sampling location of) : S2 Sample Taking. O5 removed (was removed by)	S13 Sample	000016 Sample, 010857 Earlier Sample
500	Shows/is shown at	E36 Visual Item	000021 Drawing	P67 refers to (is referred to by)	A10 Excavation Interface	000022 Profile
503	cuts/is cut by	A8 Stratigraphic Unit	000011 Archaeo-logical Object	AP7 produced (was produced by): A4 Stratigraphic Genesis. AP28 occurs before (occurs after): A4 Stratigraphic Genesis. AP7 produced (was produced by) and or only P121 overlaps with	A8 Stratigraphic Unit	000011 Archaeo-logical Object

Figure 1. Example of a form used by Ore during data mapping from INTRASIS documentation.

The group agreed to share models, descriptions and documents that describe, e.g. the internal logic of each dataset, deposition parameters, dataset particularities, initial modelling attempts etc., so that each group member can start gaining ideas on the available datasets. The template could act as a common space for group members to start populating it with models. Additional diagrams or documents could enhance the understanding of individual data modelling examples. As far as proposing a semantic mapping VRE for translating mappings into graphic visualisations (and vice versa) or setting up a training webinar for using the 3M or additional modelling tools, we decided to see how the simple spreadsheet works and discuss it again in our next meeting.

- The sixth meeting (11/10/2021) was joined by G. Bruseker, D. Nenova (Takin.solutions), A. Felicetti (PIN, ITA), M. Katsianis (PP), D. Löwenborg (UU), R. Opitz (UG) and Smith-Ore (KHM-UO). In the meeting, the observations that arose from populating the shared spreadsheet template (prepared by D. Nenova and G. Bruseker) were discussed (Fig.2). The main things noted were the requirement of long descriptions with empty intermediate classes, as well as the possibility of alternative descriptions (using different or combined CRM extension sets) depending on the expressive focus of the description/relationship. It was realised that mapping available datasets/examples and cross-checking for potential patterning was quite a complex task. However, the exercise focused on the elements that currently are difficult to map using existing solutions.

Origin Table/ Field	Root Node	Target Table/ Field	CRM Path to Target Node	rdf	Reference Resource	CRM Long Path	Notes
ExcUnit	A2	ExcUnit:ID	->P1->E42	crm:P1_is_identified_by<https://ariadne/example/conceptual_object> a crm: E42_Identifier		->P1_is_identified_by->E42_Identifier	
ExcUnit	A2	ExcUnit:Name	->P1->E33_E41	crm:P1_is_identified_by<https://ariadne/example/conceptual_object> a crm: E33_E41_Linguistic_Appellation			
ExcUnit	A2	ExcUnit:Physical Relation to Other Units	->AP11->A2	crmarchaeo:AP11_has_physical_relation<https://ariadne/example/physical_object> a crmarchaeo: A2_Stratigraphic_Volume_Unit			
ExcUnit	A2	ExcUnit:Artifacts	->AP21->E18	crmarchaeo:AP21_contains<https://ariadne/example/physical_object> a crmarchaeo: E18_Physical_Thing	Artifacts		

Figure 2. The spreadsheet template for adding data mappings by group participants.

The difficulty in creating and comparing different mappings made us rethink the suitability of an AP for excavation research. The presentation of the CRMhs AP by Felicetti, its logic and its functioning manifested the difficulties in creating an AP for excavation research given the facts that: a) excavation work is central to the archaeological process and interfaces different aspects of archaeological research and b) existing extensions targeting archaeological fieldwork already exist (CRMarchaeo).

In a sense, it was realised that the focus could perhaps be directed towards defining where current CRM family models perform well and where they don't (e.g. multiple instantiations of concepts, alternative description paths or practical issues like the creation of long descriptions with empty nodes). In addition, the initial focus of the group in the description of layered and finalised datasets that distinguish between data elements based on different levels of the excavation activity (i.e. fieldwork/post-ex analysis/post-ex synthesis/finalised) could perhaps require a different direction. To feed this conversation, Katsianis would start working on a shared document elaborating on the initial AP draft describing more thoroughly the minimal set of entities that could perhaps support the above-mentioned AP targeting. The idea was to see whether the provision of concrete conceptualizations of the excavation universe from the data collection and processing point of view would accommodate dataset integration in a more flexible manner (i.e. hide long descriptions/focus on specific mapping aspects).

- The seventh meeting (10/2/2022) had the most participants, including P. Derudas (LU), G. Hiebel (UIBK), F. Hivert (UT), M. Katsianis (PP), D. Löwenborg (UU), G. Bruseker, D. Nenova (Takin.solutions), R. Opitz (UG) and C.E. Smith-Ore (KHM-UO). The meeting continued the discussion of observations arising from populating the shared excel spreadsheet and the points raised during the mapping exercises.

  In this meeting, doubts about the necessity for an excavation AP were raised since the existence of CRMarchaeo and the difficulties encountered already in our examples led to the idea of a dedicated excavation AP as redundant. Instead of targeting a new AP, it was understood as much more sensible to focus on potential problems with existing models and offer solutions in the form of explicit modelling descriptions. This idea of directing our effort into identifying semantic patterns through actual examples could generate different sets of modelling "recipes" that could eventually be brought together into some kind of a modelling "cookbook" or even a helpdesk blog for users to ask questions and receive potential answers based on the available semantic patterning samples.

The main strategy could be to explore data mappings focusing on using CIDOC CRM and CRMarchaeo, complemented by other CIDOC CRM extensions, to enhance data interoperability at the most fundamental or basic level. In this respect, earlier work for defining an AP could be employed to identify links or similarities between CIDOC CRM entities and respective domain-specific model extensions and APs.

Potential problems in describing excavation data elements or procedures that could not be addressed, e.g. potential semantic inconsistencies, could be forwarded to the respective CIDOC CRM SIGs. Towards that end it was deemed necessary to reflect on the intended usability of aggregated data from excavations through the exemplification of potential questions that could be asked by users to retrieve item-level results across relevant excavation projects.

Further observations put forward the idea of organising a training event where individual experiences could be shared among group members, while informing a larger audience involved in excavation data modelling. This opportunity could further consolidate the overall experience of individual members of the group and wider interested parties in terms of available modelling tools or data mapping procedures.

Based on the points raised, a series of activities that have since been crystalised in finalised or working documents were launched. These documents are attached in their present form as appendices to this report and include:

a) A collection of relevant references to map the evolution and current progress of excavation semantic modelling, educational facilities and other tools. (*Annex A*)
b) An inventory of modelling patterns concerning different aspects of modelling the excavation domain. (*Annex B*)
c) A document where potential semantic problems identified in building up semantic patterns could be listed, discussed within the group and communicated to the respective CIDOC CRM SIG groups. (*Annex C*)
d) A document where potential questions for item-level inter-excavation results are formulated and discussed with respect to their semantic syntax or at least semantic requirements. (*Annex D*)

- The eighth meeting (28/9/2022) followed the virtual workshop and the EAA presentation. G. Hiebel (UIBK), F. Hivert (UT), M. Katsianis (PP) and D. Nenova (Takin.solutions) convened to reflect on the results of the respective activities and discuss ARIADNEplus reporting preparation and the post-ARIADNEplus future of the group.

  Regarding the 28th EAA Annual Meeting participation, Nenova (presenter) got the idea that our contribution may have been one of the most anticipated. Yet, the presentation itself and the entire session did not quite manage to spark discussion among participants. With respect to the virtual workshop and the resulting material, we all felt that in terms of output, the envisioned set-up of the report, alongside the presentations in both pdf and video formats provides a complete picture of the current state of things when it comes to excavation modelling.

Concerning the group's future, several options were discussed, with the most realistic being to maintain this group as an informal "discussion" or "special interest" group, and probably in relation to similar initiatives like "SHeRD". To encourage participation, relevant sessions in forthcoming conferences (like CAA or CHNT) should be scheduled. Katsianis has agreed to steer the initiative even after the end of 2022, but an open invitation for additional/alternative steering members has been initiated.

2.3 Virtual Meeting organisation

On 15 June 2022 the "*Virtual Workshop on Semantic mapping of excavation data*" took place. The event was organised as an open forum to illustrate aspects of the work carried out by the *Archaeological Excavation Modelling Working Group*, a sub-group within WP 4.4.12. The presenters, both Partners and Associate Partners of the ARIADNEplus consortium, explored semantic modelling and the use of CIDOC CRM, as well as the tools developed to assist researchers with mapping their data under four presentations.

- G. Bruseker, Takin.solutions: *Semantic Data Modelling and Archaeological Research Data - Why, How and Where We are Now*.
- O. Marlet, CNRS: *From modelling to mappings: how to appropriate the CIDOC CRM*
- M. Theodoridou & V. Kritsotakis, FORTH: *The X3ML toolkit: How to map excavation data to CIDOC CRM*
- G. Hiebel, UIBK: *An approach to model archaeological data and create RDF from spreadsheets*

Five case studies on semantic mapping of excavation data were also presented.

- F. Hivert, CNRS: *OpenArchaeo: a semantic Web platform for archaeological data*
- D. Nenova, Takin.solutions: *Modelling Archaeological Excavations. Theoretical Patterns and Practical Recipes*
- M. Katsianis & G. Styliaras, PP: *Reworking aged excavation mappings with new models and tools*
- C.E.S. Ore, KHM-UO: *597 Norwegian excavation databases and CIDOC CRM - a practical exercise*
- P. Derudas, LU, and F. Nurra, INHA: *Archaeological Interactive Report: a trait d'union between data management and semantic publication*

Each presentation was followed by a Q&A, while a discussion at the end of each session allowed participants to engage in conversation and contribute their experiences and ideas to make excavation data FAIR (Findable, Accessible, Interoperable and Reusable).

The virtual workshop was hosted by PP and executed online via Zoom video conferencing services. One hundred and four (104) people registered for the event with the following geographical distribution. Simultaneous participation peaked at 62 people (Fig.3).

Country	Initials	Participants
Argentina	AR	2
Austria	AT	13
Bulgaria	BG	2
Brazil	BR	1
Cyprus	CY	3
Czech republic	CZ	2
Germany	DE	4
Ethiopia	ET	1
Finland	FI	1
France	FR	13
Great Britain	GB	4
Greece	GR	18
Hungary	HU	4
Ireland	IE	1
Israel	IL	1
Italy	IT	12
Nederlands	NL	2
Norway	NO	3
Pakistan	PK	1
Portugal	PT	1
Sweden	SE	7
Slovenia	SI	5
Turkey	TR	3
		104

Figure 3. List of participants at the virtual workshop by country.

All presentations and the report have been made available through Zenodo (https://doi.org/10.5281/zenodo.7112917) and the respective page at the ARIADNEplus website (https://ariadne-infrastructure.eu/semantic-mapping-of-excavation-data/).

2.4 Presentations by the group

Apart from the virtual workshop, the group has delivered two more presentations.

The first was in the framework of the March 22 2022, ARIADNEplus Steering Committee meeting, where Katsianis delivered a presentation entitled "*Item level excavation data integration at the ARIADNEplus. A roadmap for activities*" concerning the overall theme of the group's activities and progress. The presentation summarised the main problems of modelling the archaeological excavation domain, introduced the group to the ARIADNEplus community, discussed working ideas and presented the group's ongoing and scheduled activities.

The second took place at the 28th EAA Annual Meeting in Budapest, Hungary, 31 August - 3 September 2022, in session 273 entitled "*FAIRly Front-loading the Archive: Moving beyond Findable, Accessible and Interoperable to Reuse of Archaeological Data*'' organised by E. Aspöck - Austrian Academy of Sciences, K. May - Historic England and H. Wright - ADS/York University. The presentation entitled "*Bringing excavation data together. Are we there yet and where is that?"* was delivered by Nenova and has since been made openly available online through Zenodo (https://doi.org/10.5281/zenodo.7117048). The presentation discussed the ARIADNEplus data aggregation strategy, presented the group to a wider research audience, introduced the problem of excavation data modelling and summarised the investigation directions of the group as well as the main conclusions of the virtual workshop.

3. Research results

The group initiated research to investigate the necessity for developing an Application Profile (AP) for excavation data ito facilitate sub-collection or item-level data mapping and integration. During the process, and after subsequent experimentation and discussion, it became evident that the current problems encountered in excavation data modelling would not benefit from such a prospect.

Excavation research has been at the forefront of archaeological data interoperability activities. This is understandable as it still provides the principal process for scientific data collection and interpretive reasoning in the archaeological domain. Data recording systems and methodologies with the aspiration of standardising the excavation process have appeared even before the digital turn in the discipline (e.g. MoLAS, Harris Matrix, J.-C. Gardin), followed by numerous digital documentation systems and underlying data models (e.g. IDEA, ArchéoDATA, IADB, SYSAND, Intrasis). The advent of semantic modelling was recognized as a way to align different documentation data resources, and this had been most apparent in the interest of archaeologists to employ CIDOC CRM and provide domain-specific implementations or facets. As part of the increased focus on digital archaeological data preservation, sharing and reuse, CRMarchaeo has been developed specifically to support archaeological data management during excavation projects to be followed by several extensions (such as CRMba and others) that interface excavation-related activities to a greater or lesser extent (like building archaeology).

In spite of the general agreement that excavation datasets cannot be fully exploited unless they can be combined or connected with wider-scale archaeological research data bodies, so far, it has proved very difficult to integrate different excavation datasets. Still, very few examples of interoperable datasets are out there, and this has greatly to do with the complexity of excavation data archives, which are compiled with different tools and methodologies and for different research purposes, employ distinct conceptual descriptions at variable granularities, can often be unfinished or open-ended and may be linked to all sorts of digital data types, each with its complicated production workflow.

Although CRMarchaeo, the domain-specific extension of the CIDOC CRM, has provided increased expressiveness, excavation research interfaces with several archaeological areas depending on the research stage, requires concepts from multiple CRM family models to achieve modelling coherence. This increases implementation difficulties as modellers have to navigate within an ever-expanding universe of concepts, properties and inheritances, while multiple implementation paths with complementary concepts mean that there are more than one ways to express certain relationships.

As a result, we are still some way from a common semantic description of the excavation universe. In this respect, it was realised that rather than attempting to create yet another AP for excavation data modelling, which could potentially add confusion, a better strategy would be to bring together already existing modelling attempts that try to model the excavation domain, compare and find their similarities and differences. Working in this direction helped us communicate our internal conceptualizations of the excavation domain and its data using a shared language. The organisation of the virtual workshop was also key towards this end, as it provided an opportunity to discuss individual experiences, contextualise the current state of research, understand available options with respect to tools that can help the modelling process and increase awareness of the benefits of conceptual data modelling in wider archaeological audiences and stakeholders.

Several practical and theoretical aspects with respect to data aggregation and semantic modelling were raised during the meeting and the participants shared advice and tips while engaging in further collaborations after the event. Five (5) research areas were highlighted as having the potential to consolidate and coordinate the excavation data modelling community. These are thoroughly described in the virtual workshop report (https://doi.org/10.5281/zenodo.7112917) and are summarised here.

a) **Models**, by developing those that have the capacity to describe the application domain consistently (e.g. CRMarchaeo etc.),
b) **Questions**, by identifying meaningful queries that can be pose on integrated archaeological excavation datasets,

c) **Methods**, by comparing existing excavation modelling examples, establishing modelling patterns and basic application scenarios, encouraging digital pedagogy and ontological thinking,

d) **Workflows and Tools**, by employing the major software tools that can be used in ontological modelling and data mapping and addressing their potential expansion or combination.

e) **Learning and training** by providing educational material, digital facilities and teaching opportunities for semantic modelling.

Concerning **models**, the evolution of CIDOC CRM has allowed domain-specific definitions that are more successful in capturing domain-specific meaning, such as CRMarchaeo, CRMba and CRMinf. However, the fact that an archaeological excavation is a multi-level process, happening on multiple fields, using different methodologies and documentation media or tools, illustrates the difficulties still encountered in its complete description. We found that the meaning of several real-world entities or their documentation proxies can be mapped to different model concepts depending on the research context or stage. For example, what may be identified as a *feature* during excavation, may be referred to as a *stratigraphic interface* that separates stratigraphic entities, as a *filled morphological building section* in architectural studies, as a *belief* to be a neolithic wall part in interpretive statements, as a *digital data object* that substitutes the original physical object in post-excavation research (Fig.4). In this respect, we need to distinguish between different contexts or stages of the excavation process and establish common descriptions that facilitate a basic level of agreement.

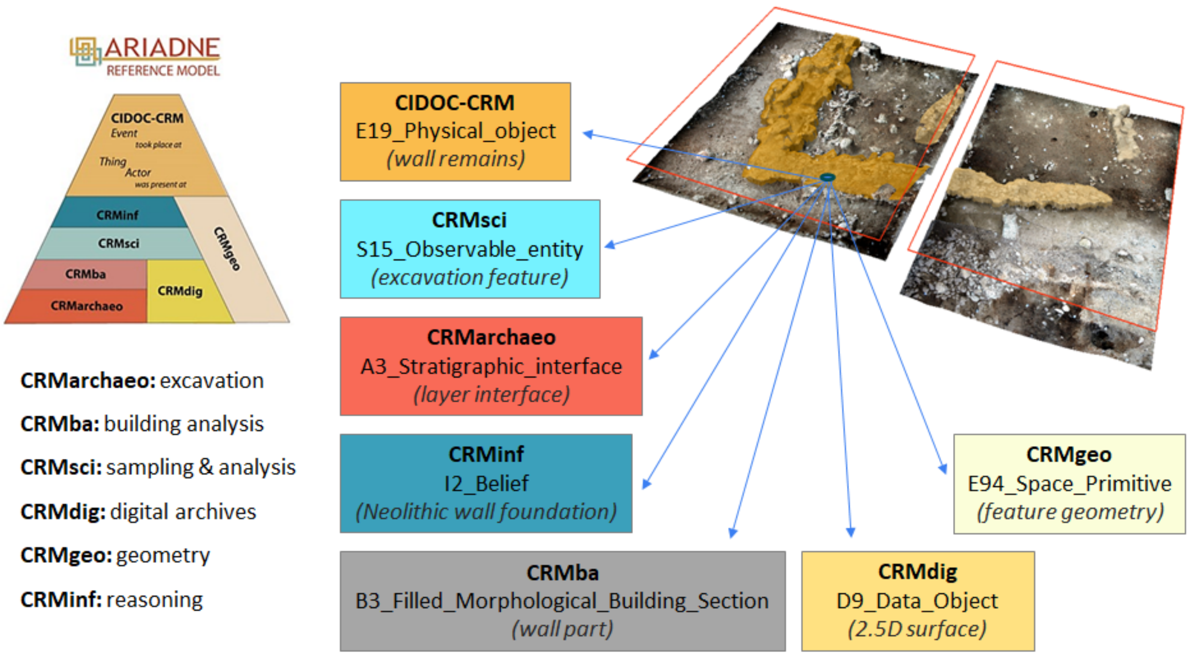

Figure 4. The CIDOC CRM family of models and an example of complementary domain-specific descriptions of an excavation feature. Image provided by the Paliambela Kolindros project.

The possibility for the multiple instantiations of concepts, i.e. the inclusion of multiple conceptual tags for the same object, was acknowledged as a useful approach for mitigating these problems. In addition, it was recognised that the more specific and all-encompassing a description of the excavation process gets, the more difficult it becomes to stick to baseline descriptions. In this respect, there are many occasions where datasets are partly compatible with CIDOC CRM descriptions and employ custom ontologies or complementary semantic standards (e.g. Dublin Core) to fit existing excavation methodologies or dissemination needs better. This is not necessarily a drawback, as data interoperability is required for information fields that make sense to be connected.

To understand what we need when searching aggregated excavation datasets, as an additional step, we need to explore the potential **questions** that researchers would formulate to retrieve existing reusable data. The formulation of questions can reveal what makes sense to ask of multiple excavation datasets and what each question involves regarding the respective syntax[2]. By posing questions and analysing their syntax, we may succeed in identifying a minimum baseline for useful excavation data descriptions. This exercise can also provide minimal and economic descriptions that can be standardised and re-used in new data integration processes. Finally, it may help in figuring out what an overall excavation domain description would mean in terms of basic concepts and linkages and how these could relate to more exhaustive descriptions of the excavation domain or other related sub-domains (see Annex D).

With respect to **methods**, we advocate the purposeful bringing together of data mapping examples to explore similarities and differences, compare their scope and meaning and decide upon standardised semantic descriptions or semantic data "patterns". Equally, 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 (Fig.5). Such a strategy can start from the core or generic entities involved in the excavation process and subsequently be out branched to cover more specific meanings.

In all cases, these data patterns can then be documented and made available within the community as standardised data *modelling recipes* in the form of a modular data description-building process or a *modelling cookbook*. 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 interpretive procedures (see Annex B).

[2] (e.g. the Question: "*Where can I find palaeolithic flint blades?"* involves the following linkage between semantic concepts: *E27 Site → AP21 contains → E22 Human-made Object,* which needs to include information on *type* [blade] (*P2 has type → E55 Type*), *chronology* [Palaeolithic] (*P108 was produced by → E12 Production → P10 falls within → E4 Period*) and *material* [flint] (→ *P45 consists of → E57 Material*).

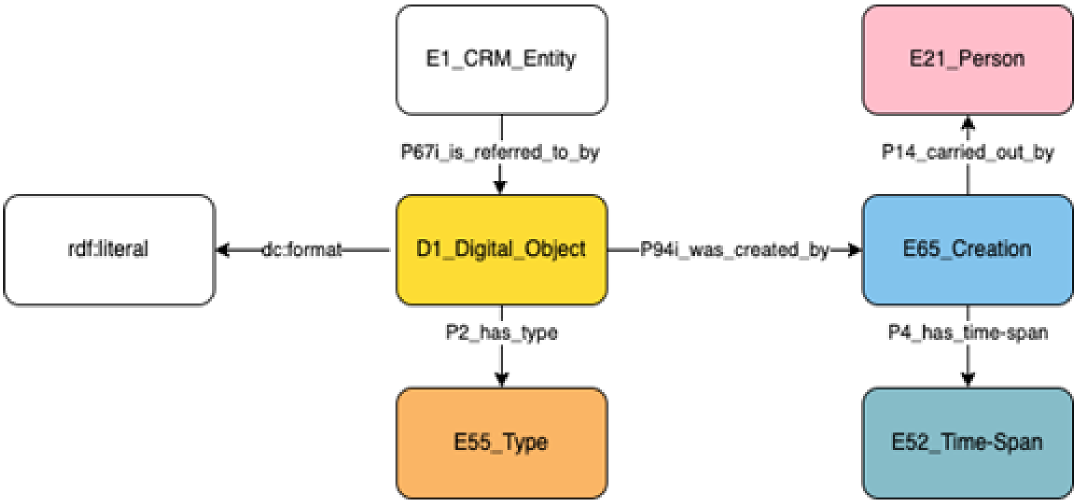

Figure 5. Modelling spatial data documentation files.

With respect to **workflows** and **tools**, several main pathways were identified for semantic mapping.

a) Within the ARIADNEplus project a data mapping workflow is based on the X3ML toolkit, which comprises a set of small, open-source software components for information integration. These include the X3ML Mapping Definition Language (https://github.com/isl/x3ml/blob/master/docs/x3ml-language.md), the 3M Mapping Memory Manager (https://demos.isl.ics.forth.gr/3m/), the X3ML Engine (https://github.com/isl/x3ml) and the RDF Visualiser (https://demos.isl.ics.forth.gr/RDFV-Demo/). Together these tools can be used within a complete workflow for transforming XML exports of datasets/databases into CIDOC CRM-compatible RDFs. This process can be combined with tools for achieving vocabulary homogenization such as the Vocabulary Matching Tool (VMT) by the University of Wales (accessed from within the ARIADNEplus VRE services and https://heritagedata.org/vocabularyMatchingTool/), which maps concepts to Getty's Art and Architecture Thesaurus and PeriodO - https://perio.do/, which is used to match chronological periods with absolute date ranges.

b) Under the MASA consortium, a workflow that covers the data lifecycle of archaeological excavation data includes several steps and multiple tools, for dataset cleansing (OpenRefine - https://openrefine.org/), ontology structuration and vocabulary alignment into an interoperable dataset (e.g. mySQL, PostgreSQL) that is mapped (Protégé-Ontop, (https://protege.stanford.edu/ & https://ontop-vkg.org/) and validated (SHACL, https://shacl-play.sparna.fr/play/) as an RDF TripleStore with a SPARQL endpoint (https://sparnatural.eu/). The generic backbone allows the linkage of several excavation datasets into an interoperable data pool at the expense of the specificities of each dataset.

c) Another approach - useful for aligning finalised or closed datasets - attempts the direct creation of RDFs from spreadsheets or databases. Datasets are analysed and rearranged into sets of spreadsheets or data tables that correspond with a combination of semantic standards. Tables are then integrated within a database (Postgres) that allows further data structuring (e.g. URI addition). Afterwards, semantic tools (e.g. Karma, https://usc-isi-i2.github.io/karma/ or OntoRefine) are used for RDF creation. In many cases, though, similar processes that involve the alignment of several datasets with respect to both their transformation to a conceptual reference model and among themselves may need to be complemented by decisions on what to include to the final deposition files, as well as significant manual data cleaning, conversion and alignment (e.g.the ADED project)

d) A different route, showcased by the *Archaeological Interactive Report* (AIR) attempts to standardise the archaeological publication process by providing an extensive alignment of archaeological information fields with multiple semantic models. This solution standardises the entire informational potential of the publication platform data schema to allow further data integration with data aggregators using RDF files.

Several other data mapping solutions exist between or outside these three main workflows. It seems that the current ecosystem of data modelling and knowledge organisation tools is far from standardised. Certain steps towards the further standardisation of certain workflows, the further purposeful development of selected tools by the archaeological community and the explanation of the benefits or overlaps of specific tool sets with actual examples is required.

Accordingly, **learning and training** should be given attention to attract other audiences for building new or sharing existing semantically compatible archaeological excavation datasets. At the rookie level, the *CIDOC CRM Game*, either in the table (https://www.cidoc-crm-game.org/) or its online edition (https://ontomatchgame.huma-num.fr/), can provide an entry point for domain experts who want to understand the mechanics and benefits of semantic structures. Also, the catalogue of bibliographic references and educational material (see Annex A). At a more intermediate level, the identification of excavation data modelling patterns (Annex B) can eventually comprise a *cookbook* for data modellers, allowing the creation of a kind of marketplace for archaeological semantic data. At an even more advanced level, this marketplace can include data modelling examples and workflows complemented by structured tutorials and targeted workshops. As discussed by many experts during the event, training in archaeological knowledge organisation may bring together different views of the excavation universe and even impact archaeological theory and data management methods.

4. Relevant work by individual partners

This section serves to provide descriptions of some of the projects that have been included in this work or have provided modelling and data examples. Some of these projects employ datasets that are being ingested within the ARIADNEplus knowledge base and effort is directed to include examples that will be accessible through the ARIADNEplus Lab VRE (https://ariadne.d4science.org/web/ariadneplus_lab/). Using tools included in the virtual environment, such as *GraphDB*, researchers could explore the ARIADNEplus knowledge base for excavation data with the available web GUI or programmatically with SPARQL queries.

- The Paliambela Kolindros (Greece) excavation dataset is the product of one of the first 3D GIS documentation workflows that included conceptual modelling using CIDOC CRM to describe data components. Integrating this excavation dataset within the ARIADNEplus infrastructure at the item level has been a core incentive for participating in the consortium. The preparation of a partial and incomplete geospatial data archive for deposition has made numerous challenges apparent, some of which included the necessity to update existing semantic descriptions. Within the excavation modelling group, data modelling components have been produced and shared, while the final data mapping attempted to include to some degree several of the semantic patterns contained in Annex B and also to experiment with multiple instantiation (Fig.6). Data mapping was implemented using the 3M tool. Apart from the practical aspects of consolidating an excavation dataset, this process has led to many realisations concerning the digitally assisted excavation documentation and reasoning (see the presentation by Katsianis and Styliaras in the virtual workshop). The dataset's metadata have been aggregated within the ARIADNEplus knowledge base and are available via the portal at https://portal.ariadne-infrastructure.eu/resource/7ef8e39b437a44e9e92612f4e7f93b5631e8c5361e55a06abdddf94c8938f036.

Origin Table/ Field	Root Node	Target Table/Field	CRM Path to Target Node	Notes	Example
ExcavationProcessUnit	A1/AO_Activity	ExcavationUnit	AP5->A2/AO_Object		
ExcavationProcessUnit	A1/AO_Activity	ExcUnitIsRelatedToFeature	P9->S19		
Feature	E25/AO_Object	ExcUnitIsRelatedToFeature	O19i->S19		
Feature	E25/AO_Object	SpatialCoordinates	P53->E53->Q11i->SP6		
SpatialCoordinates	SP6	SpatialCoordinates:CoordX	Q10i->SP5->P1->E41->(p2->E55[x])		3245.12
SpatialCoordinates	SP6	SpatialCoordinates:CoordY	Q10i->SP5->P1->E41->(p2->E55[y])		-26732.34
SpatialCoordinates	SP6	SpatialCoordinates:CoordZ1	Q10i->SP5->P1->E41->(p2->E55[z])		59.59
Feature	E25/AO_Object	Feature:FeatureID	P1->E42->P190->rdf:literal	PK	F102
Feature	E25/AO_Object	Feature:FeatureType	P2->E55	CV	posthole
Feature	E25/AO_Object	Feature:FeatureDimensionWidth	P43->E54->(P2->E55[width:constant]/P91->E58[m.:constant])->P90->rdf:literal	(length /m.)	0.15 (m)
Feature	E25/AO_Object	Feature:FeatureWidthOrientation	P43->E54->P67i->E33->P190->rdf:literal	txt	N-S
Feature	E25/AO_Object	Feature:FeatureMaterial	P45->E57	txt	mudbrick
Feature	E25/AO_Object	Feature:FeatureGeometry	P67i->D1/AO_Data_Resource->P1->E41->P190->rdf:literal		F102.shp
Feature	E25/AO_Object	FeatureGroupIsFormedByFeature	AP16i->A6		
FeatureGroupIsFormedByFeature	A6	FeatureGroupIsFormedByFeature:FeatGID	P141->B5->P1->E42->P190->rdf:literal		FG01

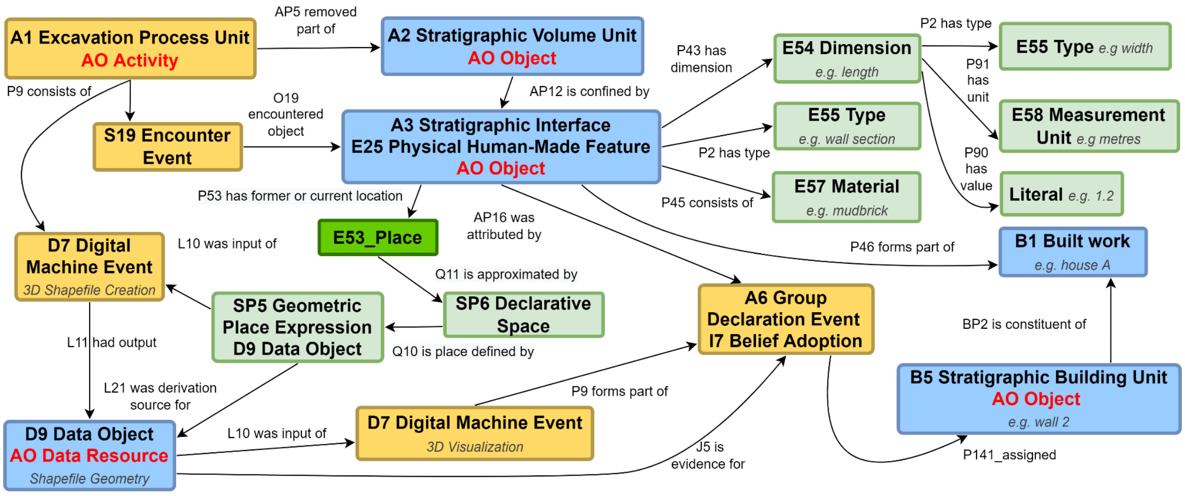

Figure 6. Data mapping example (table and graph) from the Paliambela Kolindros excavation project.

- In France, the MASA consortium is working to disseminate the FAIR principles and find solutions to help archaeologists bring their data to the Semantic Web. For this purpose, MASA has created a digital ecosystem with several tools to help researchers process their data, from structuring to dissemination on the Semantic Web. Within the workflow developed, MASA worked on two aspects in particular. On the one hand, the implementation of [OpenArchaeo](), a semantic Web platform for archaeological data: MASA has set up a SPARQL semantic web platform with a user-friendly interface (Sparnatural), allowing users to intuitively generate a query without having to write any SPARQL code. The OpenArcheo datasets have been aggregated within the ARIADNEplus knowledge base and are available through the portal (Fig.7). On the other hand, the training of archaeologists at applying CIDOC CRM is pursued through the card game by G. Bruseker and A. Guillem, whose online release, [Onto Match Game](), is managed by MASA. The digital online version of the card game is fully customisable (ontology, instances, pedagogical progression) and automates and systematises the pedagogical part of the game (Fig.8). Thanks to these two achievements, MASA is succeeding in mobilising more and more archaeologists around the issue of Open Science.

  The OpenArcheo datasets have been aggregated within the ARIADNEplus knowledge base and are available through the portal:
  https://portal.ariadne-infrastructure.eu/resource/e41c59ca73ac30b3dd5932da08bedf33b2999303c399ad742fdb8b0ee0ed9a8d

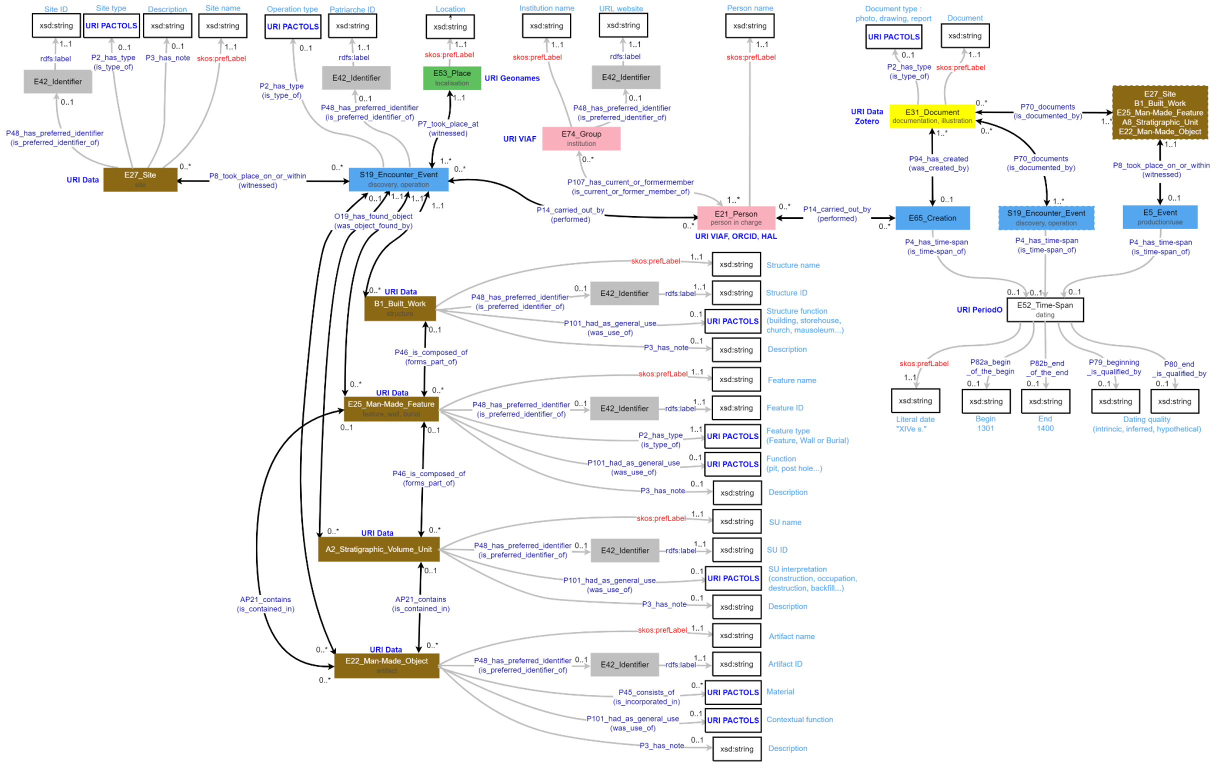

Figure 7. Generic model (CIDOC CRM based) to map archaeological data into OpenArchaeo.

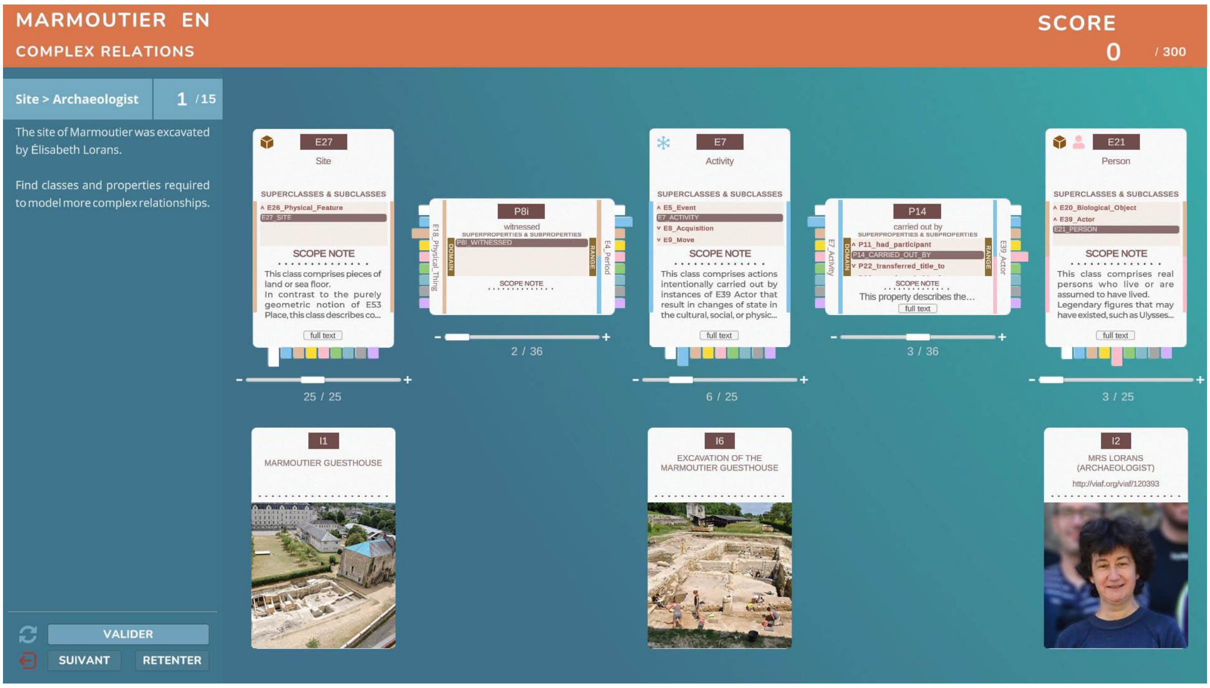

Figure 8. Learning CIDOC and data mapping with Onto Match Game.

- In Austria, at the Archaeological Department of the University of Innsbruck, in a specific Open Research Data Pilot project (https://www.uibk.ac.at/projects/ord4mining-archaeo/index.html.en), the excavation data from the project "Prehistoric Copper Production in the Eastern and Central Alps" was modelled using the CIDOC CRM ontology. The data was collected during several scientific research campaigns and is related to prehistoric mining activities in the eastern Alps of Austria. The documentations were done according to the guidelines of the Austrian Federal Monuments Office (BDA – Bundesdenkmalamt). The extension CRMsci was used to model S20 Rigid Physical Features, S4 Observations and S5 Inferences and CRMarchaeo to model A8 Stratigraphic Units (Fig.9).

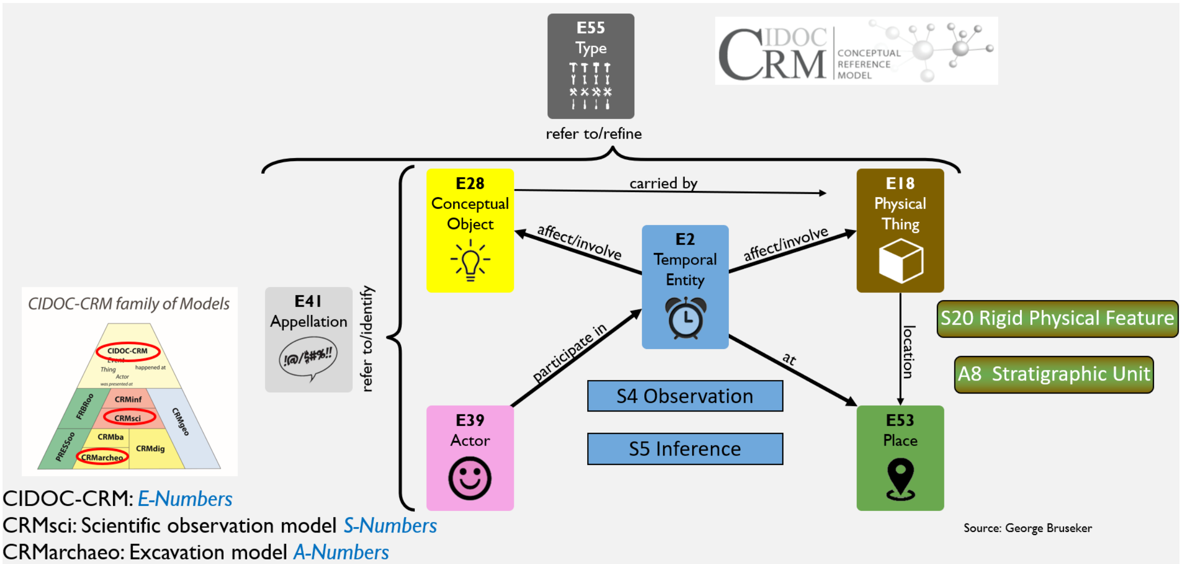

Figure 9. CIDOC CRM general classes and relations with specific CRMsci and CRMarchaeo classes.

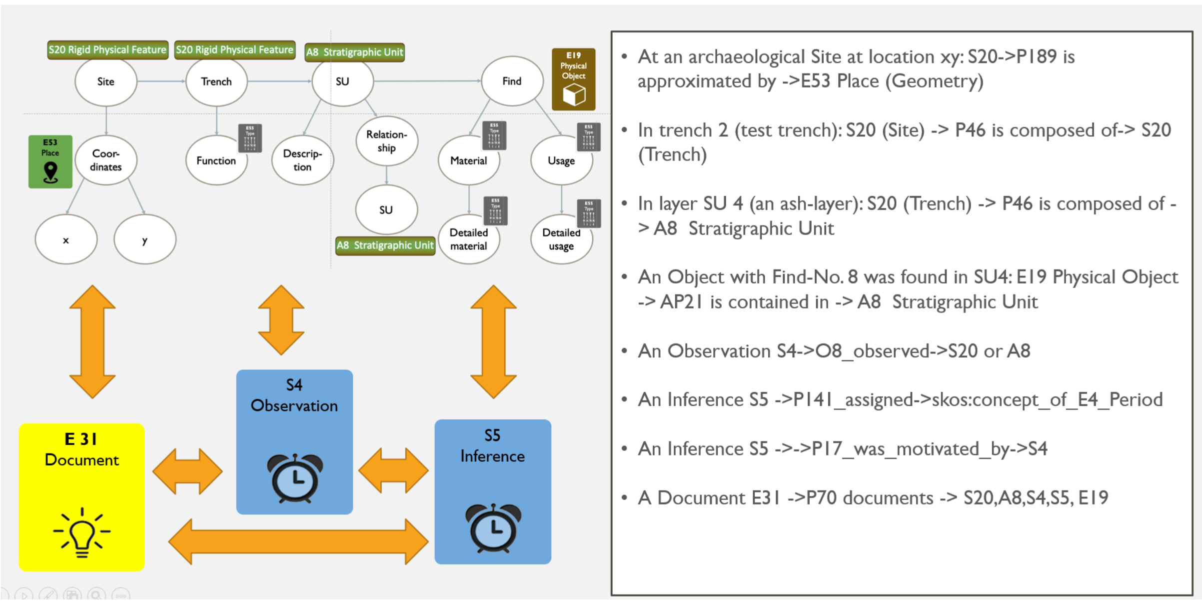

Figure 10. CIDOC CRM general classes and relations with specific CRMsci and CRMarchaeo classes.

Figure 10 shows the basic modelling of an excavation with the mentioned classes and the main properties used to relate these classes. The modelling is implemented in a GraphDB Triple store at https://disc-semantic.uibk.ac.at/ in the repository "Open_Research_Data_Mining" and details of the modelling can be explored there.

For example, the following link shows the site "Gratlspitz-1": https://disc-semantic.uibk.ac.at/graphs-visualizations?uri=http:%2F%2Fuibk.ac.at%2FORD%2FS20%2FGratlspitz-1 (Fig.11). The visitor can choose the "Open_Research_Data_Mining" repository on the top right first and reload. Afterwards it is possible to log in as "guest" with pwd "guest" to adjust viewing settings. It is recommended to turn off "Include Inferred Statements" and set the language tag to "en" (English).

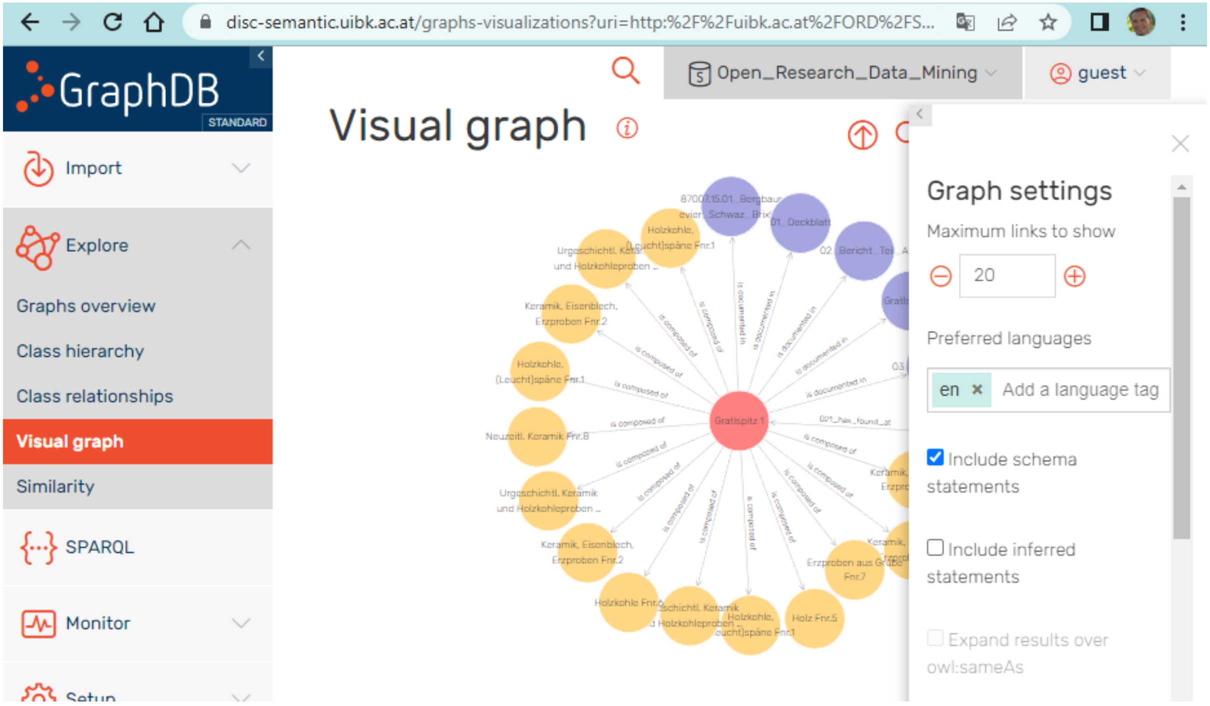

Figure 9. Graph DB Repository with excavation data

- A set of semantic patterns concerning different aspects of modelling the excavation and archaeological study domain has been developed as a part of two ongoing projects with the participation of Takin.solutions. The Common Grond project (https://isvroma.org/sv/2021/11/03/common-ground-a-research-platform-for-digitized-archaeological-collections-and-archives-at-the-swedish-institutes-in-athens-rome-and-istanbul-2) led by the Swedish Institute in Rome aims to integrate nearly a century of archaeological data originating in different formats, with different applied methodologies andusing various data solutions. The end product will be an integrated platform using Arches (https://www.archesproject.org/), which will unite decades of research performed at the Swedish archaeological institutes in Athens, Rome and Istanbul. Another project (SEMAFORA), led by the University of Groeningen, aims to develop a strategy for expressing the specifics of archaeological survey practices. Three base datasets are used for data mapping in consideration with other common and methods and techniques. The patterns devised based on these endeavours are developed to provide the core branches for each primary activity, from which each archaeological project can further develop depending on individual methodological and practical specifications.

5. Conclusions – Next steps

We are still some way from the seamless reusability of archaeological excavation datasets. The centrality of the excavation process as the most significant data generator of archaeological practice requires novel attention in terms of its ontological significance and interfacing with sub-domain-specific practices (like digital fieldwork documentation, laboratory studies and interpretive processes). We have tried to show that it is time for the analytical work in the description of the excavation universe to be complemented by synthetic attempts that can foster a more stable and user-friendly ecosystem of methods and tools for all those researchers who want to structure their data in meaningful and interoperable ways. Towards this end, as a group, we have moved away from the need for a separate AP for excavation research and tried to identify some key areas (*models*, *questions*, *methods*, *workflow & tools* and *learning & training*) where further work can provide all the necessary aspects for excavation datasets to be both FAIR and especially Re-usable.

An additional and quite important benefit of the group's activities was bringing together what has largely been a fragmented community of scholars working in the same direction. Members of the group include researchers that, apart from developing, try to *implement* archaeological excavation semantic modelling in existing datasets and examples. This interest group, which would not have been possible had it not been for the encouragement of the ARIADNEplus Steering Committee, the past two and a half years managed to exchange ideas and advance concrete lines of thought within a largely unstructured research environment.

In terms of the future of the group, several options were discussed. It was felt that this initiative should continue as the group's survival could help approach a wider archaeological community. The most realistic option after the conclusion of ARIADNEplus is to maintain this group as an informal "discussion" or "special interest" group, and probably in relation to similar initiatives like the Semantic Heritage Research Data (SHeRD) (https://twitter.com/sherd_athens)[3]. Communication channels with relevant CIDOC-SIG will also be sought to provide feedback from our experiences in trying to implement the model and informon new CIDOC-related resolutions. To stay active, the organisation of relevant sessions in subsequent conferences (like the CAA or CHNT) will be pursued to bring together this community (in person for a change!). For the foreseeable future, PP and Katsianis can continue to steer the initiative even after the end of the year. Of course, further members from the ARIADNEplus community are welcome to join the group's work and also take over or share its steering.

[3] See also the series of online seminars by the SHerd group at the respective Youtube channel https://www.youtube.com/channel/UCVKjdvN5CUaerqmE9DbvlWw/about.

6. References

Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S. & Velios, A. June 2022. *Definition of the CIDOC Conceptual Reference Model. Version 7.1.2.* CIDOC CRM Special Interest Group. https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.2.pdf.

Doerr, M. & Theodoridou, M. 2014. *CRMdig. An extension of CIDOC-CRM to support provenance metadata*. Version 3.2.
http://www.cidoc-crm.org/crmdig/sites/default/files/CRMdig3.2.pdf.

Doerr, M. et al. 2019. *Definition of the CRMarchaeo. An extension of CIDOC-CRM to support the archaeological excavation process*. *Version 1.4.8.* PIN, University of Florence, Italy. http://www.cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.8.pdf.

Doerr, M. et al. 2020. *Definition of the CRMsci. An Extension of CIDOC-CRM to support scientific observation. Version 1.2.8*.
http://www.cidoc-crm.org/crmsci/sites/default/files/CRMsci%20v.1.2.8.pdf.

Felicetti, A., Meghini, C., Richards, J., Theodoridou, M. 2021. Towards the AO-Cat Ontology. Version: 1.1.

Hiebel, G., Doerr, M., Eide, Ø. Theodoridou, M. et al. 2015 (September). *CRMgeo: a Spatiotemporal Model An Extension of CIDOC-CRM to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement.* Proposal for approval by CIDOC CRM-SIG Version 1.2. https://cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf.

Katsianis, M., & Styliaras, G. (eds) 2022. *Virtual Workshop on Semantic mapping of archaeological excavation data (1.0)*, with contributions by Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Katsianis, M., Kritsotakis, V., Marlet, O., Nenova, D., Nurra, F., Styliaras, G., Ore, C. E., Theodoridou, M.. Zenodo. https://doi.org/10.5281/zenodo.7112918.

Nenova, D., Bruseker, G., Derudas, P., Hiebel, G., Hivert, F., Katsianis, M., Marlet, O., Opitz, R., Ore, C.-E., & Uleberg, E., 2022 (September 27). Bringing Excavation Data Together. Are We There Yet and Where is That?. *28th EAA Annual Meeting (EAA 2022)*, Budapest, Hungary. Zenodo. https://doi.org/10.5281/zenodo.7117049.

Paveprime and collaborators 2018. *CRMinf: the Argumentation Model. An extension of CIDOC-CRM to support argumentation*. Version 0.10.
https://cidoc-crm.org/crminf/sites/default/files/CRMinf%20ver%2010.1.pdf.

Ronzino, P., Niccolucci, F., Felicetti, A., Doerr, M. et al. 2014 (December). *Definition of the CRMba An extension of CIDOC CRM to support buildings archaeology documentation*. Proposal for approval by CIDOC CRM-SIG. Version 1.4. PIN S.c.r.l. https://cidoc-crm.org/crmba/sites/default/files/2016-12-3%23CRMba_v1.4.1_UR.pdf.

Annexes

Annex A - Excavation modelling reference list

Annex B - Excavation modelling patterns

Annex C - Excavation domain semantic issues

Annex D - Queries for retrieving excavation data across different projects