Università degli Studi di Torino
Dipartimento di Informatica
Ciclo XXXIV

DOCTORAL THESIS

Digital Data Curation Through Semantic Encoding:

An operational proposal for the journey of archaeological data

by TUGCE KARATAS

Supervisor:
Vincenzo LOMBARDO

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Turin
Department of Computer Science
2022

Abstract

Digital Data Curation Through Semantic Encoding:

An operational proposal for the journey of archaeological data

by Tugce Karatas

Digital curation in cultural heritage organisations has become more and more established as empirical research for tools, techniques, skills, and standards for making curators able to manage the related digital data. It supports specific applications in diverse contexts of cultural heritage management. This thesis addresses the archaeological domain, a particular challenge since projects span from the planning of the excavations to the analysis of the findings, their interpretation, and the display of the results in a final exhibition. Further, in archaeology, digital curation must account for the relationship between physical materials and their digital twins. Our approach formulates a comprehensive definition of digital curation for the archaeological domain and devises a unified model based on the semantic organisation of the data. The methodology that is employed in this study is to i) abstract a general model for digital curation from the analysis of cultural heritage domains with a particular focus on archaeology ii) validate the model on some case studies in the archaeological field, and iii) apply the model to an archaeological project, with a preliminary evaluation of the approach and the suggestions of about the merge of the semantic encoding of archaeological data with a transdisciplinary approach.

Acknowledgements

I would like to begin by thanking my supervisor, Vincenzo LOMBARDO, for imparting his knowledge and domain expertise for three years. In addition, I would like to thank the members of scientific committee for my viva; Vito MESSINA from Università degli Studi di Torino; Daniela PETRELLI from Sheffield Hallam University; Angeliki ANTONIOU from University of West Attica.

Also, I want to thank the scientific board, the advisory board, mentors and especially management team of Tech4Culture[1] programme as well as fellow scientist from BeArchaeo[2] project especially Monica GULMINI and her research team for their collaborations and support. In addition, I am grateful to Daniel PLENTICKX, Ivan VARRIALE and Vittorio LAURO for providing the case studies and sharing their experiences with me.

I personally would like to thank Costis DALLAS, Tsvi KUFLIK and Davide NADALI for their reviews, assistance, guidance and feedback to improve my thesis.

I thank all of the people I got in touch with during this PhD, especially I would like to thank my friends all over the world for supporting me and making this journey remarkable.

My biggest thanks go to my family for their immense love, support and encouragement without them nothing would be possible.

---

[1]*Tech4Culture project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754511*.

[2] *BeArchaeo project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823826*.

TABLE OF CONTENTS

Chapter 1

Introduction

In this chapter, we introduce digital curation and its fundamental tenets, in the context of cultural heritage processes and data. We start from general definitions of the digital curation process and provide some background information and related terminology. We begin by exploring the digital technologies for representation, creation, visualisation and maintenance of cultural heritage data. Then, we address the specific area of archaeology. Archaeological investigations are challenging because of their multidisciplinarity in the contemporary era and the intense contribution of archaeometric disciplines. There are many individual approaches to digital curation in the archaeological field since every archaeological project is going digital. In addition, there are general technologies and tools for ontologies for Cultural Heritage that have specialised in archaeology (such as the CIDOC-CRM family, including CRM-archaeo). Given its transdisciplinary and reflexive methodology, the Semantic Web approach has been introduced to support the modelling, representing, preserving and sharing of archaeological data. The research problem addressed in this thesis is to devise an effective digital data curation process for archaeology, based on Semantic Technologies. Here, we introduce the research aim of this thesis in its contribution to the field, i.e.we sketch our approach to digital data curation through Semantic Technologies and conclude with the organisation of the thesis.

1.1 Digital Curation

Almost two decades after its debut[3], Digital Curation has had a significant impact on the management of digital assets as a field of intellectual inquiry concerning emerging pervasive curation practices in the digital environment (Dallas, 2015b). While it can be used as a broader term for the creation and management of digital assets during their entire lifecycle

---

[3] in 2001 during the "Digital Curation: digital archives, libraries and e-science" seminar http://www.ariadne.ac.uk/issue/30/digital-curation/ (last visited on 15 January 2022)

(Lee and Tibbo, 2007; Yakel, 2007), Digital Curation Centre[4] (DCC) defines digital curation as "actively managing data [...] from its point of creation until it is determined not to be useful".

Digital Curation has received much attention in recent years due to the rapid advancement of digital technologies, which are inevitably having an impact on cultural heritage. Cultural heritage is a very comprehensive field with an important role in the representations of societies' identities, belongings and behaviours which, in its broader meaning, includes tangible, intangible, and digital heritages (Jokilehto, 2017). The management of the cultural heritage assets is a complex challenge that addresses many competencies from diverse disciplines (such as, e.g., conservation science, archaeology, and chemistry), with information technologies that are more and more pervasive, transversely.

By its definition, digital curation inevitably addresses digital cultural heritage that is linked to many other cultural heritage domains, which are stored in different ranges of formats over several large datasets as well as a vast amount of documents handled by libraries and archives. Therefore, it serves as a digital infrastructure (Taylor 2001; Lord and Macdonald 2003) that provides the glue between research, representation, practice, and training across nations, disciplines, institutions, repositories, and data formats (Gold, 2010; Ray, 2009) in cultural heritage management.

Nowadays, there is a vast amount of digital technologies that cultural heritage specialists have to engage in including digital data visualisation, information analysis, and sharing results. On one hand, the advancements in technologies since the '80s with their increasingly complex and fruitful implications are helping to achieve the cultural heritage institutions' ultimate goal which is not only to preserve but also to share knowledge. On the other hand, according to Harvey's defining work (2010), technical obsolescence or fragility, lack of resources, ignorance of good practices, and uncertainty over appropriate infrastructure constitute serious risks to the practice of cultural heritage specialists considering the responsibilities involved in digital curation are to be shared across different institutions and communities.

4 http://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20What%20is%20Digital%20Curation.pdf (last visited on 5 January 2021)

In this context, digital curation has emerged as a viable solution in the coordination of the representation and management of digital information in the field of cultural heritage: in particular, digital cultural heritage assets that are a part of the complexity of selection, preservation, maintenance, collection and archiving, with the further requirement of adding value for subsequent exploitation (Yakel, 2011).

1.2 Digital Technologies and Archaeology

As archaeology is a transdisciplinary effort, the research questions emerge through collaboration and cross-fertilization of the knowledge exchange between colleagues from diverse disciplines. At the same time, archaeological datasets are also becoming increasingly available online: projects such as the Digital Archaeological Record[5] (tDar), The Central Institute for Cataloguing and Documentation of the Italian Ministry of Cultural Heritage[6] and the Archaeological Data Service[7] are making a range of archaeological data available for quantitative testing and processing, and re-use. For example, being used widely in the archaeological field, the Digital Archaeological Record (tDAR) serves as an international online digital repository for the records of archaeological investigations as well as related images, geospatial data and 3D scans. This online repository is housed by the Center for Digital Antiquity and is a secured data bank that has relational capabilities to combine various datasets from different archaeological projects. It was established to improve and assist in the preservation, dissemination, and access of archaeological data, which is a nonrenewable resource.

Inevitably, the improvement in digital technologies creates possibilities to fulfil many kinds of new projects in the archaeological field but also controversies. According to Niccolucci et al. (2009), the researchers have not been keen on the documentation of the archaeological excavations. The archaeological research data are represented through different modalities and acquired by different tools such as digital photography, 3D modelling and visualisation of the excavation, GIS mapping, diaries and video recordings, usage of tablets, computers, digital and analogue cameras, digital video recorders, and specialised 3D capture, modelling and georeferencing hardware and software (Forte *et al.*, 2015). For example, 3D modelling is required to build effective models from point clouds yielded from scanning and the semantic organisation of the data has contributed to effective database schema design. As digital technologies are growing faster, the need for a theoretical debate and principles with

[5] http://www.tdar.org/ (last visited on 15 January 2021)
[6] http://www.iccd.beniculturali.it (last visited on 21 December 2021)
[7] http://archaeologydataservice.ac.uk/ (last visited on 8 November 2021)

practical implications is also growing. As a result, numerous projects and documents are aimed at creating efficient guidelines and a set of principles of computer-based visualization to ensure intellectual and technical integrity, reliability, documentation, sustainability and accessibility (Demetrescu and Ferdani, 2021). Especially on the topic of virtual reconstruction and visualization of the archaeological data, various documents published over the years by different initiatives such as the London Charter[8], Seville Principles[9] and V-Must[10]. Particularly the London Charter points out the importance of structuring and documenting not only the sources used and their metadata but also the interpretation, namely paradata, made to achieve the visual representation (Hugh, 2012).

Additionally, in the last decade, as digital technologies evolved and made a transition from the Web 2.0 phase into a Semantic Web (3.0) era, users have been enabled to participate in the creation, sharing and aggregation of contents and data interoperability (O'Reilly, 2007). Consequently, this shift has supported the open, linked and meaningful data that can be shared by diverse scientific communities. In this context, Linked Data has emerged as a term that was first proposed by Tim Berners-Lee (Berners-Lee *et al.* 2001) and formally defined by the World Wide Web Consortium (W3C) in 2011, is also widely adopted/adapted in the digital humanities and heritage communities. The use of Web 2.0 and the possibilities offered by the semantic tools has eased the process of cataloguing each item and the related analyses, establishing statistics that provide several joint results characteristic of metadata handling platforms (Isaksen, 2008).

As we move towards the implementation of the idea of the Semantic Web and Linked Data, the field of archaeology slowly progresses in the same direction (Richards 2006; Wright 2011). Although emerging tools and standards benefit modelling and publishing of the data, managing archaeological data is particularly challenging considering its interdisciplinary characteristics which include e.g. anthropology, chemistry, biology, history, and physics as an outcome of multi-organizational and multilingual research activity. Related to these challenges, many initiatives have tried to create platforms, tools and systems that enabled to describe of different kinds of data and apply the benefits of the Semantic Web technologies to Cultural Heritage and archaeology such as, e.g., Functional Requirements for Bibliographic Record (FRBR) (Tillett, 2005), Dublin Core (DC) metadata Elements and DC Terms (Powell *et al.*, 2007), Simple Knowledge Organisation System (SKOS) (Miles &

[8] http://www.londoncharter.org/preamble.html (last visited on 11 March 2022)
[9] http://sevilleprinciples.com/ (last visited on 11 March 2022)
[10] http://www.v-must.net/ (last visited on 11 March 2022)

Bechhofer, 2009), Lightweight Information Describing Objects (LIDO) (Coburn et. al, 2010), MIDAS Heritage standard (Forum on Information Standards in Heritage (FISH), 2012), Europeana Data Model (EDM) (Meghini *et al.*, 2016), CIDOC Conceptual Reference Model (CIDOC-CRM)[11] and OAI-ORE[12] (Open Archives Initiative Object Reuse and Exchange).

Among the types of metadata standards/schemas, there are generic ones (such as Dublin Core) while others are domain-specific (such as the schemas that are known as the CIDOC-CRM family). Generic schemas tend to be easy to use and widely adopted, but specific schemas (such as CRMsci) have a much richer vocabulary and structure but tend to be highly specialised and only understandable by researchers in that area. However, the CIDOC CRM ontology family stands out because it is particularly concerning the archaeological data model, as well as the issues emerging from its actual usage in real application scenarios. By definition, the Conceptual Reference Model(CRM) is a formal schema intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information (Crofts *et al.*, 2009). The model was created by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) on empirical bases from real-world datasets. It allows schematic representation and in-depth analysis of an archaeological data recording system by providing a common and extensible semantic framework for evidence-based cultural heritage information integration. Therefore, the analysis and experimentation of the CIDOC CRM are particularly interesting, as ongoing research contributes to the debate about the evaluation of the archaeological data model's interoperability in Semantic Web languages.

Although all of the tools below are evolving rapidly, there is one that is particularly associated with cultural heritage, namely the comprehensive Europeana Data Model (EDM). EDM is an ontology that enables cultural heritage institutions to structure collection data, so they may be utilised by the data aggregator Europeana[13] project. Europeana features over fifty-eight million cultural heritage items from around 4,000 institutions by aggregating data from other EU-funded projects and frameworks such as CARARE (Connecting Archaeology and Architecture)[14] and ARIADNE (Advanced Research Infrastructure for Archaeological Data Networking in Europe)[15]. This model is designed to provide a generic data model for the core

---

[11] https://www.cidoc-crm.org/ (last visited on 21 January 2022)
[12] http://www.openarchives.org/ore/1.0/toc (last visited on 21 December 2021)
[13] https://www.europeana.eu/en (last visited on 11 December 2021)
[14] https://www.carare.eu/en/about/ (last visited on 10 January 2022)
[15] https://ariadne-infrastructure.eu/ (last visited on 19 January 2022)

categories (e.g., object type, media type, date, place) and be compatible with specific data models which can be utilised by individual galleries, libraries, archives and museums (GLAM in short). The generic data model draws from some existing top-level semantic tools and standards, as mentioned before, namely OAI-ORE, Dublin Core and SKOS and CIDOC-CRM.

In this regard, notable results of general value have been obtained so far, but the problem of representation of archaeological data, information and knowledge in contexts of semantic integration of heterogeneous metadata schemata is still a fundamental concern (Hyvönen 2009; Hacıgüzeller *et al.* 2021; Buranarach *et al.* 2022). There is a great potential for further research that mainly focuses on two crucial aspects of archaeological data curation: the creation of a semantic backend based on semantic repositories (using semantic relations); the design and development of user-centred interfaces based on semantic technologies, and their potential application in the dissemination through different forms such as exhibitions and scientific publications. In this research, we would like to examine further these issues to create an operational workflow and its application in an archaeological project.

1.3 Research aim and contribution

Recent research on the conceptualisation and practice of digital curation shows that there are both overlaps and gaps between digital curation tasks and digital tools in various cultural heritage institutions (Post, 2019; Poole, 2016; Tibbo, 2015). In recent years, there have been a significant number of researchers working on the development of digital curation workflows and frameworks (Post *et al.,* 2019). They have been provided with a way to assess the evolution of the e-resource management processes (Anderson *et al.*, 2010). These frameworks consider the research data in their general sense and provide guidelines that are not necessarily related to the field of cultural heritage or specifically archaeology. The specific needs and peculiarities and complexity of archaeological data are far from being addressed in their specific issues. However, to create a successful digital curation infrastructure and implement it in complex archaeological projects, there is evidence of a need for a unified workflow that conceptualises the major entities that form the tasks of the entire process, from the acquisition of digital data to their exploitation (Benardou *et al.* 2010; Poole 2016; Post *et al.* 2019). Creating a unified framework for shared workflow between teams could facilitate a ‘common language’ and create the foundation for more powerful cross-discipline analysis (Ridge *et al.,* 2005). Following Costis Dallas’ account of Gardin’s logicist approach, it has become even more important to obtain an understanding of how researchers interact with digital resources in constructing knowledge, as previous studies in the humanities have highlighted (Dallas, 2016).

However, recently, there have been influential contributions in the field of archaeology specifically on the importance of the semantic tools and ontologies for cultural heritage objects (Kakali *et al.* 2007; Havemann *et al.* 2009; Hyvönen 2009; Niccolucci *et al.* 2015). We acknowledge the networked methods for cultural heritage data archiving and dissemination (Power *et al.* 2017; Seifert *et al.* 2017) as well as the visualisation and annotation of archaeological models in real-time (Poyart *et al.* 2011; Snyder 2014). But, currently, problems in the archaeology field are mainly: i) vague ideas of the semantic relations held within the archaeological data, ii) every institution has its solutions and software tools but most of them fail to provide platforms for researchers to use digital technologies both for visualise data and documentation, iii) there is not a unified single framework to encompass the digital curation activities for archaeologists, iv) researchers are not very aware of the issues of Semantic Web and Linked Data (Bouchenaki, 2003; Carboni and de Luca, 2016; Lemonnier, 2012). The efforts to aggregate the different structures and languages with different epistemological traditions, without a semantically enriched framework, have led to reductive implementations that are of limited value long term. Therefore, an operational workflow is needed to demystify the entire digital curation approach in archaeology to align semantic technologies with the various tools, software and techniques that are used by specialists from diverse disciplines. The main aim of this research is to bridge this gap using a digital data curation approach to archaeological data concerning its transdisciplinarity.

Analysing the experiences with concrete projects, developed in the cultural heritage field, in this research we provide a semantic approach to solve the pragmatic concerns of digital curation, its methods, and its applications in archaeology. Therefore, we propose a comprehensive model for archaeology that concerns the journey of the archaeological data from scientific research to exhibitions. The model mostly focuses on digital data curation of an archaeological investigation, especially how the knowledge is linked to the form interfaces, for collecting the data as the excavation goes on, to be continued in the analysis labs, and eventually with the design and curation of the exhibitions. We have identified the major entities that are required for a reflexive methodology of archaeology, especially in its relationship with archaeometric knowledge. In particular, we investigate the establishment of a transdisciplinary approach to archaeology and archaeometry, interlinked through a semantic model of processes, relations and objects while compassing the archaeological excavation, the following archaeometric analyses of the site and the excavated materials, the interpretations of the findings, and the dissemination of the results through physical and virtual exhibitions (environments). In terms of the interdisciplinarity and diversity of the archaeological data, we use Kansa and Kansa's approach to "data literacy" for archaeologists and examine techniques to take care personally of their data (Kansa and Kansa, 2021). However, about data sharing, we aim to design user-friendly interfaces for the semantically enriched digital publication of archaeological excavations (Opitz and Johnson 2016; Lercari *et al.* 2017).

Given the above, we propose and demonstrate a novel framework for archaeological data curation workflow. We believe that a unified operational digital data curation workflow is a solution to manage the archaeological data effectively during its journey from scientific research to all the digital data curation processes namely conceptualisation, acquisition, processing, modelling, publication, and dissemination (Karatas and Lombardo, 2020). While aligning archaeological ontologies with real-life archaeological projects, we reflect on the aspects of archaeological digital infrastructures and related exhibition environments concerning community engagement, interpretation and meaning-making of the archaeological record (McDavid 2002; Moser *et al.* 2002).

1.4 Organisation of the thesis

The thesis is organised as follows. In the next chapter, we will provide a short historical perspective of the concept of "digital data curation", to understand the problems that raised the notion and the community that is engaged in the solution. Then we will address the existing gaps through some case studies which later we will use in the thesis. In Chapter 2, we will briefly examine existing digital curation frameworks with particular attention to two of them. Then, with the findings, we will introduce an abstraction of the Digital Data Curation workflow, which represents the outcome of the literature review mentioned above (Karatas and Lombardo 2020). We will explain each step and component of the workflow. Then we will apply this abstract workflow to the case studies to get some results for improvement of the workflow. In Chapter 3, we will approach knowledge representation in archaeology by adopting the Semantic Web paradigm. We will examine some archaeological projects that use semantic tools and discuss the necessities of semantically interoperable information sharing in archaeology. In Chapter 4, we will revise and refine the proposed Digital Curation Model by integrating semantic technologies from the beginning of its creation until the end of its lifecycle to provide a backbone. Later, we will test the Digital Data Curation Model on a methodological project, namely BeArchaeo, for the establishment of a transdisciplinary approach to archaeology and archaeometric disciplines, interlinked through a semantic model of processes and objects. The simplified and modified model, driven by a semantic conceptualisation, will be focused on data processing and interpretation up to the exhibition of the results while leaving the management and preservation issues for the next chapters. In Chapter 5, in particular, we will discuss how the connections between the archaeometric analyses and archaeological interpretations can be represented by carrying on the data collection through digital curation. While conceptualising the relationship between digital twins of the archaeological findings we aim to examine the connection between archaeological knowledge and scientific discoveries of archaeometric analysis in the catalogue records of the archaeological form. We will particularly discuss Omeka-S, a Content Management System that allows for the creation of semantically-driven database design. We will provide an evaluation of the workflow based on the feedback from the scientists as well as the technical components of the system. In Chapter 6, we will draw the concluding remarks on the research concerning both the more general and theoretical elements, and the technological and technical ones. The thesis will end with overall comments and conclusions and then suggest the possibilities for future work.

Chapter 2

Digital Data Curation Model

In this chapter, we provide a historical perspective on the notion of Digital Curation, intending to build an abstraction of the major component processes. We examine the potential benefits of digital curation as well as its limitations specified in the field of archaeology to find a solution for the creation and maintenance of archaeological data in long-term projects. We discuss some findings on digital curation frameworks to understand the similarities and the differences. We give particular attention to the DCC and respectively DCC & U model. Then, we introduce the Digital Data Curation model, which represents the outcome of the abstraction from the literature review and empirical findings. Later, we apply the digital data curation model to case studies to validate the Digital Data Curation model. Then, for validation purposes, we apply the Digital Data Curation Model to three case studies; Virtual Electronic Poem, Ename Crosier, and the Venus Pompeiana which are chosen from the archaeological field based on their spatial and temporal peculiarities. We conclude with the findings from the case studies.

2.1 A Brief History of Digital Curation and Models

Digital Curation has received much attention in recent years due to the rapid advancement of digital technologies, which are inevitably having an impact on cultural heritage. While the study of cultural heritage involves different types of activities (including visualisation, information analysis and sharing results), digital curation concerns the management of scientific records or measurements of any type of data or digitally encoded information (Gold, 2010), including various types of medium (text, video, audio, etc.), the processes, models, tools, and software involved in these processes. One of the most significant areas of current research involves the digitisation of cultural heritage, in particular the development of technologies and systems to design and implement digital tools specialised in cultural heritage. The necessity for the digitalisation of cultural heritage is obvious, but to prepare the backbone for a digital archive it is necessary to recognize the importance of semantic representation of the data (Dallas, 2007).

On the other hand, while the research and practice of digital curation have been continuing to mature, relatively little empirical, comparative research and the consequent awareness of the tasks that compose the digital curation process has been achieved to date (Post, 2019). In order to respond to this need, some researchers have attempted a systematisation of the best practices and have built digital curation frameworks and models (Aliaga 2011; Dallas 2015). The notion of digital data curation has been revised and updated several times, with a recent focus on motivations and big data (Pouchard, 2015).

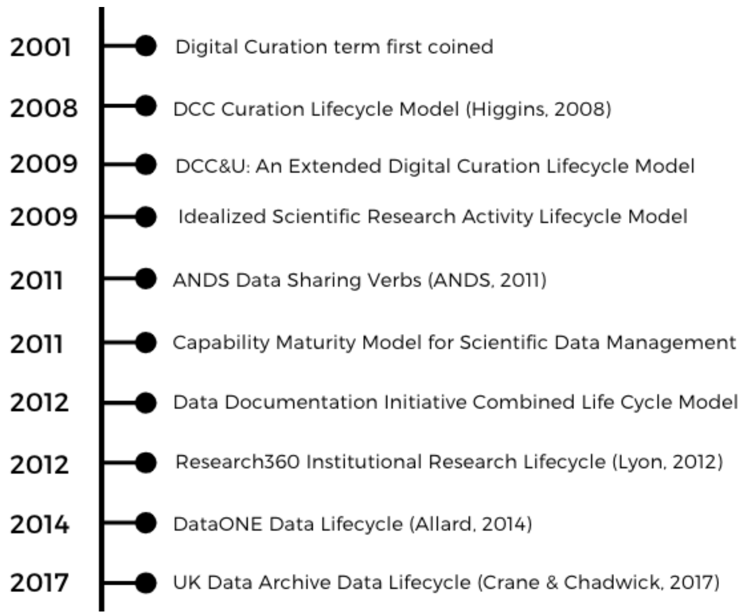


Figure 1: Brief timeline of digital curation and significant models.

Over two decades, a number of frameworks and lifecycle models[16] (as seen in Figure 1) have appeared in the literature from many institutions, namely Digital Curation Center Curation Lifecycle Model (Higgins, 2008), Extended Digital Curation Lifecycle Model (Dallas *et. al*, 2009), I2S2 Idealized Scientific Research Activity Lifecycle Model (Patel, 2009), Data Documentation Initiative (DDI) Combined Life Cycle Model (Gregory, 2011), ANDS Data Sharing Verbs (ANDS, 2011), DataONE Data Lifecycle (Allard, 2014), Research360

[16] In addition, the Open Archival Information System (OAIS) model broke new ground in the late 1990s and became ISO standard 14721: 2003. However, it is excluded from the list because it does not constitute a full-fledged lifecycle model and neglects to specify guidelines for creating or (re)using data (Lee, 2005; Lee, 2009).

Institutional Research Lifecycle (Lyon, 2012), Capability Maturity Model for Scientific Data Management (Crowston and Qin, 2011), UK Data Archive Data Lifecycle (Crane and Chadwick, 2017).

These frameworks consider the research data in their general sense and provide guidelines that are not necessarily related to the field of cultural heritage or specifically archaeology. Although they present some differences during their steps of digital data management, they overlap in the definitions of the tasks that compose the digital curation workflow. As the majority of frameworks mentioned above are created to deal with the complexity and diversity of some specific cases of data management, there are two of them that deserve particular attention. Therefore in the next subsection, we examine the Curation Lifecycle Model of Digital Curation Center (DCC) and Extended Digital Curation Lifecycle Model (DCC & U) respectively. The DCC & U model in particular is different from the others because the lifecycle of digital curation undertakes cultural heritage as a collection-driven domain from an information lifecycle perspective (Dallas *et. al*, 2009).

2.1.1 DCC and DCC & U Curation Models

Since its establishment in 2004, the Digital Curation Center[17] in the United Kingdom has contributed to the field with a detailed lifecycle model of digital curation. The Curation Lifecycle Model of Digital Curation Center (DCC) is devised by Sarah Higgins, in 2008, in order to be used to plan activities within an organisation or consortium to ensure that all necessary stages are undertaken, each in the correct sequence mostly by curators and digital librarians to organise their way of work in the long-term projects.

[17] https://www.dcc.ac.uk/ (last visited on 20 January 2021)

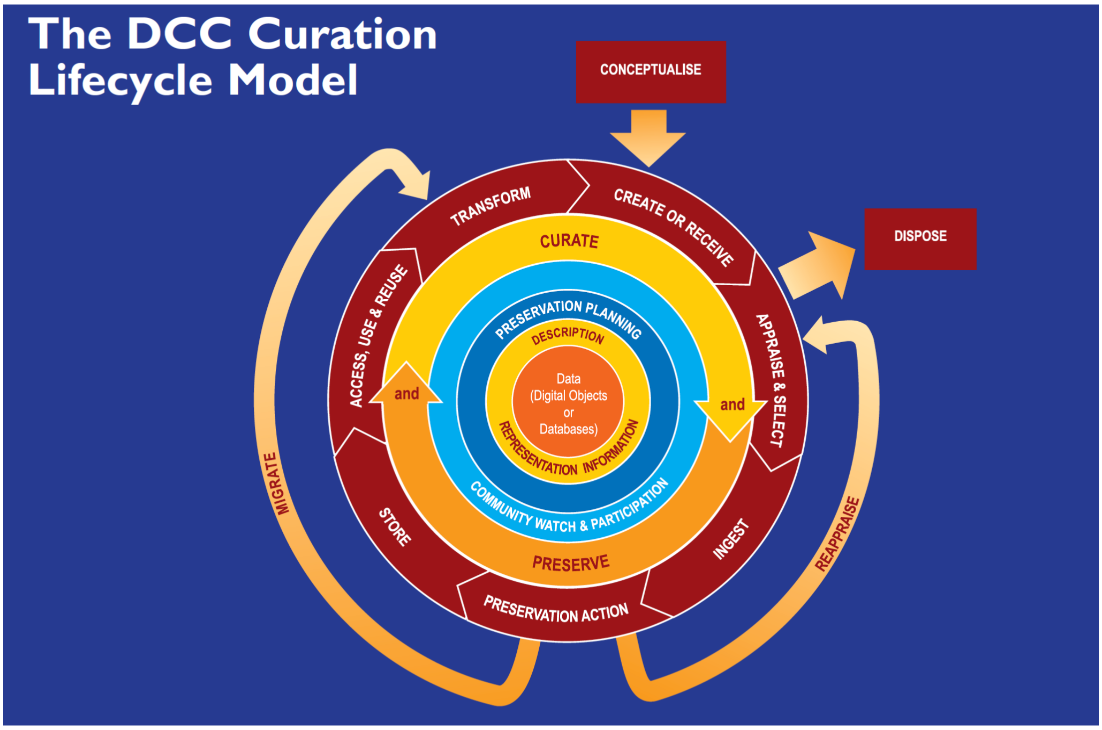


Figure 2: Digital Curation Center (DCC) Curation Lifecycle Model (Higgins, 2008)

The ***Description and Representation Information*** layer of the DCC Curation Lifecycle Model (Figure 2) consists of the administrative, descriptive, technical, structural and preservation that are necessary to adequately describe a digital object or database (shown at the centre of the figure) in the long term. It is the process of gathering all the information necessary for the understanding and rendering of the object and its metadata that consists of the creation, collection, preservation and maintenance of sufficient metadata to enable the data to be used and reused for as long as they have value to justify continued curation. ***Preservation Planning*** includes the necessary administrative and management plans for the actions of the lifecycle model. It involves strategies, policies and procedures for all curation actions. ***Community Watch and Participation*** helps in the development and evolution of appropriate standards and tools and their usage. The observation of what the OAIS Reference Model[18] terms a Designated Community which is a predetermined group of stakeholders in the data, in order to track changes in their requirements for the participation in the development of standards, tools and software relevant to the data. The ***Curate and Preserve*** include the need to be aware of, and undertake all the management and administrative actions planned to promote curation and preservation throughout the curation lifecycle. It properly describes most of the actions in the model but is used here to represent

[18] Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). Blue Book CCSDS 650.0-B-1. Also published as ISO 14721:2003. url: http://public.ccsds.org/publications/archive/650x0b1.pdf ((last visited on 20 December 2021)

the execution of the planned management and administrative actions supporting curation. After a conceptualization phase (top of Figure 2), where the specific cultural heritage domain is conceptualised and formalised, the curation lifecycle starts with the ***Create or Receive*** operation, where ‘create’ refers to original data generated and recorded by the researchers, and ‘receive’ refers to pre-existing data collected from other sources. The curation activities at this stage concern the verification that all the data are accompanied by the necessary administrative, descriptive, structural and technical metadata. It continues with the ***Appraisal and Selection*** of the data for long-term preservation using well-documented guidelines, policies and legal requirements. The ***data ingestion*** is about transferring the data not only to appropriate repositories but also ensuring that appropriate standards are used during this action. In the appraisal process, the development of criteria for the evaluation of potential resources as well as the actual selection of the resources may become subject to subsequent curation processes. Again, the ***disposal*** step is an occasional action but should be driven by documented guidance, policies or legal requirements. The remaining data are sent for ***Ingest*** by the custodians to an archive, repository, data centre or some other service. This process immediately leads on to the ***Preservation Action*** stage, which involves an array of different activities: quality control, cataloguing, classifying, generating fixity data, registering semantic and structural metadata, and so on. The preservation process aims at safeguarding against longevity risks. Any data that fail quality control checks are returned to the originator for further appraisal. This should result either in improvements in the quality of the data (e.g. corrections to data transfer procedures, improved metadata, repackaging of data) and reselection, or disposal. Some data may need to be migrated to a different format, either to normalise it within the system or to reduce risks arising from hardware or obsolescence.

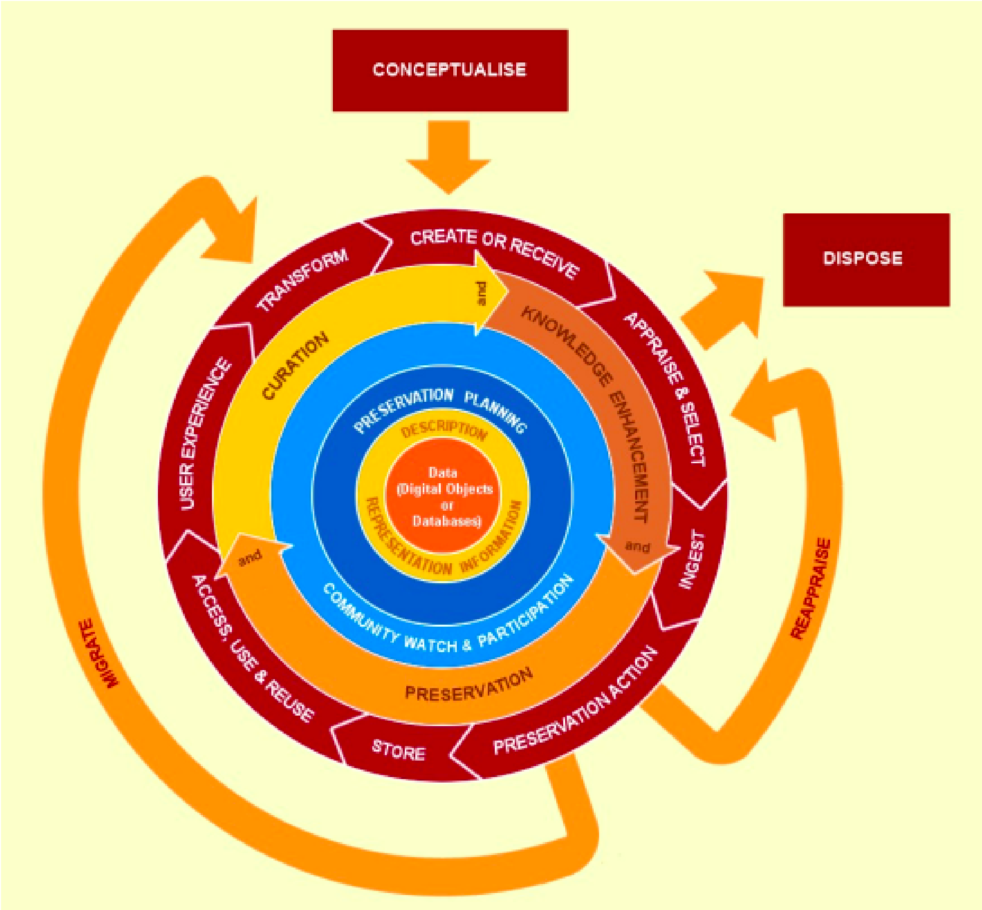


Figure 3: The DCC&U model provides a graphical high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation. (Constantopoulos et. al, 2009)

As seen in Figure 3, the Extended Digital Curation Lifecycle Model (DCC & U) is a combination of the UK Digital Curation Center's Curation Lifecycle Model and the Digital Curation Unit (DCU) model proposed by the Athens Research Centre in Athens (Constantopoulos et al., 2009), created by Panos Constantopoulos and Costis Dallas. It added the existing "Curate and Preserve" of the DCC model into the "Preservation, Curation and Knowledge Enhancement" action as well as the User Experience step (Dallas *et. al*, 2009).

Knowledge enhancement refers to entities, situations and events represented by digital resources in the real world, their wider context and domain, and the digital resources themselves; for example, annotating documents with the entities of an ontology they refer to, representing formally the situations or events mentioned in documents, and linking documents to other documents that support or contradict them would all be cases of knowledge enhancement. Presentation, publication and dissemination processes include the generation of new artefacts (scientific, scholarly, artistic, etc.) from existing primary or secondary digital resources. As a unique approach, the ***User Experience*** process captures the interaction between users and resources, as well as the effects of this interaction. Concerning both physical (centralised or distributed) and virtual interaction points, this step concerns ensuring the data are available to users as well as their experiences.

In conclusion, the models explicitly address the increase in the scope and complexity of data and have direct applicability to digital curation. While the DCC model is designed for data curators, data creators and data users, not specifically in the cultural heritage field, the DCC&U model emphasises the characteristics of the cultural heritage domain and context management. Both models represent the complex processes found in digital curation in a comprehensive and generic model that shows the lifecycle model of the data. They both provide a graphical high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt.

Strictly related to this research, the user-centred approach of the DCC&U model reflects the involvement of several disciplines which raise issues of personalization, mostly neglected by other most spread approaches (Pennock 2007; Post 2019). Because digital curation is a long-term endeavour that spans from the excavations to the interpretations and exhibition of the results, it also includes revisions and relationships with the societal context. In particular, to support the interdisciplinary collaboration between specialists in the field of archaeology, digital curation should encompass all the activities not only the management of the data from its creation to its possible discard but concern also reusability, authenticity and interoperability alongside long-term accessibility (Yakel *et al.*, 2011; Pryor, 2012). We also follow DCC&U's "Knowledge Enhancement" aspect in integrating semantic tools into the digital curation practice in order to encode the archaeological knowledge through an operational model as a procedure for addressing the transdisciplinary endeavour. By developing a prototype model that involves all the tasks, and tools and conceptualising the entire process, we define the roles and responsibilities and software applications for collaborative work. The abstraction, driven by a semantic conceptualization, will be focused on data processing and interpretation up to the exhibition of the results while leaving the management and preservation issues for future work.

2.2 Digital Data Curation Model

In this section, we present an operational schema of the Digital Data Curation model that concerns background knowledge and disciplines that populate the galaxy of cultural heritage domains through an abstract representation of the tasks, adapted from previous studies (Karatas and Lombardo, 2020). It serves as an abstract workflow of digital data curation, with the digital tasks and data formats that are concerned with the management and workflow of the digital assets. All activities involved in managing data, including, but not limited to, planning, creating, digitising at the best format, documenting, ensuring its availability, accessibility, safe storage, applicability for interpretation processes, and re-use in the future are part of digital curation. As we examined in previous sections, recent research on the conceptualisation and practice of digital curation shows that there are both overlaps and gaps between digital curation tasks and digital tools in various cultural heritage institutions (Post, 2019; Poole, 2016; Tibbo, 2015). Now, we examine the model and the individual tasks of digital curation in-depth, together with the technologies and activities involved in archaeological practice particularly; then, we see how they operate in the context of the entire workflow by analysing actual archaeological projects that have been implemented the Digital Data Curation model empirically.

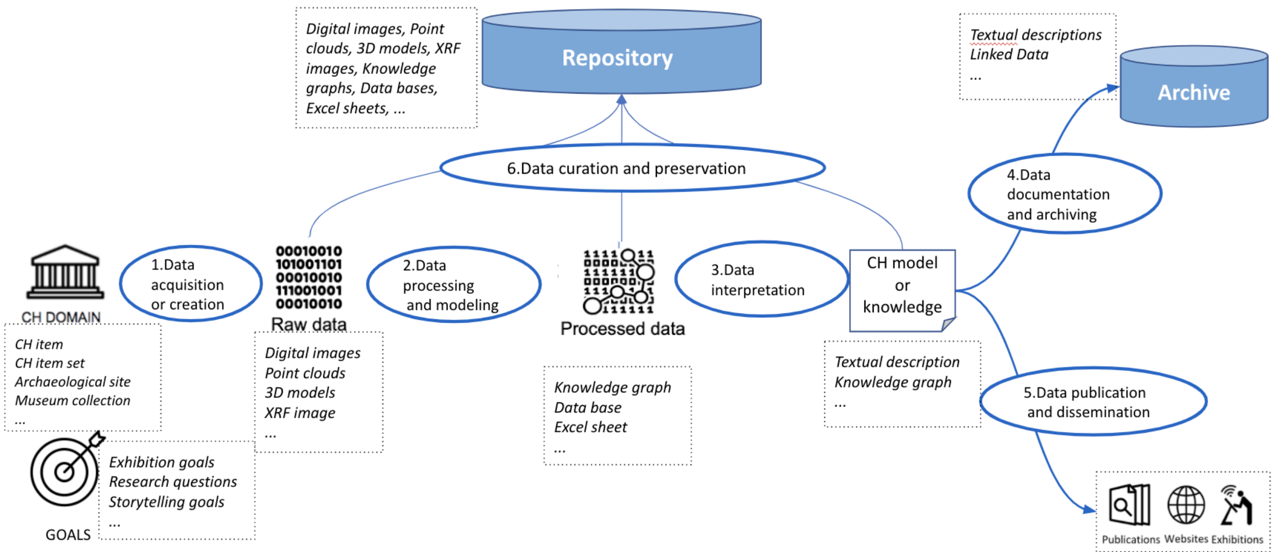


Figure 4: An abstract Digital Data Curation model

The Digital Data Curation Model consists of six common tasks, as highlighted with the blue circles in Figure 4, from the cultural heritage asset to the final outputs of the digital curation process. Each task has the tools and the possible components as bordered by the dotted lines. These six tasks are namely:

1. Data Creation or Acquisition
2. Data Processing and Modelling
3. Data Interpretation
4. Data Documentation and Archiving
5. Data Publication and Dissemination
6. Data Curation and Preservation

Data acquisition or creation

Digital data curation typically starts with the data creation or acquisition (numbered 1 in Figure 4) by focusing on what data is acquired, how, and why. Data acquisition brings data that have been created by a source outside some organisation into the organisation, for production use. This means that a number of activities, supported by tools, must be carried out, namely identifying, sourcing, understanding, assessing, and ingesting raw data. Instead, data creation is the process that samples signals that measure real-world physical conditions and converts the results into digital numeric values. Archaeology usually includes operations such as laser scanning or photogrammetry, while archaeometry includes scientific tests, such as radiography or observation under an electron microscope. The growing involvement of archaeometry in archaeological research is generating huge sets of digital entities from a variety of instrumental measurements, which can be performed either on the archaeological objects or samples detached from them. It is here that data start their life cycle; algorithms and instruments must be annotated for subsequent reference and processing and interpretation purposes. Such technical metadata are stored with raw data for possible revisions, in case of acquisition errors.

Data processing and modelling

The data processing and modelling phase (numbered 2 in Figure 4) focuses on creating a conceptual model for the data to be stored in a database or spreadsheet, together with the associations between different data objects and set of rules (many projects employ E-R Model and UML format). The goal is to support the effective exchange of knowledge and interoperability. This phase can be iterated and/or concerning several acquired data objects.

As an example, we can consider the realisation of 3D models from point clouds of an archaeological finding and its chemical elemental composition. Even by employing the same scientific technique for determining the chemical elemental composition (for example, X-ray fluorescence), the composition can be produced as a qualitative table, a quantitative table, or a chemical map of the surface, according to the equipment that is used for the investigation. Different digital objects are therefore produced and each of them gives different information. The role of the data processing and modelling phase is therefore crucial to clarify this point and to enhance the quality of the subsequent phase of interpretation.

Data interpretation

Data interpretation (numbered 3 in Figure 4) is the process of making sense of data that have been collected, analysed, and presented. This phase has a strong connection with the reflexive methodologies addressed above. Interpretation can be carried out by humans or machines; the result can be an explanatory text in natural language, a revealing diagram, or, in the case of semantic reasoning, a chain of inferences or a knowledge graph. The members of the project can access a holistic overview of the data and the interpretations can concern individual items, sets of items, or higher-order categories: the dating of an archaeological finding, its motivation (relying on other digital data) and the maps with the paths of materials from source locations to final locations, are two frequent examples.

Data documentation and archiving

The data documentation and archiving process (numbered 4 in Figure 4) manages the metadata about some data products (e.g., database tables) that enables one to understand and use the data. It concerns all the data that contribute to the interpretation and greatly supports reflexivity. Data and documentation can be classified by the type of content included in it (e.g., bibliographic, statistical, document or text) or by its application area (e.g., biological, geological, etc).

Data dissemination and publishing

On the other hand, data dissemination and publishing (numbered 5 in Figure 4) is the distribution or transmission of statistical data or of the knowledge arising by the overall process to end-users, made available in some online structured format or as paper publications (i.e., PDF files) based on aggregated data, as well as the exhibitions and websites of the collections owned by the cultural heritage organisations. Although the entire model is designed considering the User Experience approach of previously discussed models, this step particularly concerns the interpretations, expressed as narratives and models, as experienced by different scientific communities (such as archaeologists, historians, and chemists) and general audiences.

Data curation and preservation

“Data curation” is used for the curation of records or measurements of information (“data”). The scientific measurements or records (“data”) are further distinguished from the computer science meaning of “data” to refer to any type of digitally encoded information (Gold, 2010). In this sense, digital curation concerns the conceptualisation of the overall goal which shapes the data we collect and informs our efforts to model, interpret and publish the information. The task of data curation and preservation[19] (numbered 6 in Figure 4) records all the data and metadata created during the first three phases. Particularly related to the knowledge enhancement aspect of previously examined models, the semantic relations between artefacts and their constituents are crucial in this step. It also concerns the aspects regarding authorisation, persistent identification, data curation[20] and long-term archiving.

To summarise the whole model through a simplified fictitious example, we can refer to an archaeological item such as an archaeological finding. As soon as some asset is acquired (e.g., a container with a cup shape in a trench of an archaeological site), the digital curation process builds some digital assets; these can be acquired from the asset (e.g., by photograph) or created from scratch (e.g., artistic 3D modelling for illustration purposes); this will produce some data, that we can generally name raw data because they do not include a model yet (e.g., a digital photograph or X-RAY image). These data are enriched with metadata that reveals an interpretation of the asset at some level (e.g., region of the image, identified via a path joining the pixels, labelled with the tag "warrior"). Metadata can reveal hidden knowledge about the raw data. Together with data, metadata enriches the data and transforms it into processed data and is used for the interpretation process to become a

[19] Data preservation does not necessarily imply continued access. Keeping the data comprising data safely without actively managing them can result in data which still exist, but which are unusable. For example data integrity checks may not have been carried out, and data can become unusable.

[20]Data curation includes: selection, appraisal, preservation, documentation, enrichment, aggregation, disposal and transformation for example, migration to an updated format. Curation is the act of managing digital items held within an archive over the long term. It is an active and on-going process, implying action on the part of the curators so that items remain secure, discoverable and accessible. Whereas digital curation involves maintaining, preserving and adding value to archived items ‘throughout their lifecycle’.

proper model of the cultural heritage domain (e.g., the wine cup is a Kylix of 7th century BC). These processes also proceed as a part of the repository of the data curation and preservation tasks, which also consist of the creation of a repository. The model and the relative knowledge about the domain at hand, can be used in the archives as a part of the data documentation and archiving (e.g., the digital image also receives the identifier of the physical Kylix). Also, it can become a publication and dissemination outcome (e.g., the digital image is part of a virtual collection exhibited through a website or a catalogue, and inserted in a scientific paper for a journal).

Comparing the previously examined frameworks (mentioned in 2.1.1), the goal of the project refers to the step of selection and appraisal. It appears in different scenarios according to research design, setting parameters, initial queries before ingesting as well as selection criteria and processes. In either case, this process depends on the ultimate goal and generates data that are hugely relevant to preserve and access later.

However, since most of the digital curation models lack a thorough validation of concrete projects as examined in previous sections, we have decided to apply the model to three real-life case studies to discuss the digital technologies and tools within the digital data curation model to understand the needs and components of diverse archaeological projects to revise the digital data curation model afterwards.

2.3 Testing Digital Data Curation model

In this section, we apply the Digital Data Curation Model to three case studies that represent the archaeological domain. The Virtual Electronic Poem (VEP), a virtual reconstruction project of an archival event, was chosen because of its interdisciplinary approach to digital heritage. As another example, we examine an archaeological finding named Ename Crosier which was found during the archaeological excavation in Belgium. It provides a good example of the usage of pervasive digital technologies in archaeological projects. For the third example, we introduce the Venus Pompeiana (also known as the Cistern area) project which is a complex project that includes several datasets, data representation and interpretation. The examples are chosen based on their chronological significance as VEP is a good example for contemporary, Ename Crosier is related to modern archaeology while Cisterna is classical archaeology. They are also well distributed according to the object addressed namely VEP is the archaeology of multimedia installation, Ename Crosier is an object in a museum, and Venus Pompeiana (Cistern area) is an archaeological site in

Napoli, Italy. Finally, the selection of the case studies is also based on the kinds of knowledge representation aligned to the archaeological entities; formally (Venus Pompeiana), technically (Ename Crosier), and representationally (Virtual Electronic Poem) complex artefacts.

2.3.1 Case Study: Virtual Electronic Poem

In 2005, the Virtual Electronic Poem (VEP) project realised the virtual reconstruction of a 1958 event (namely Poème Électronique) which was originally designed by Le Corbusier for the Philips pavilion at the Brussels 1958 World's Fair (Lombardo *et al.* 2006; 2009). The project had many components (sound, light, time, and space) with complex challenges in finding the related documentation and providing a reconstruction of the event which had occurred a few decades ago. It is an important example of "archaeology of multimedia" and it also serves as a reconstruction example, concerning tangible and intangible heritage respectively, a temporary event of the 20th century (Lombardo et. al, 2006).

The research ***goal of the project*** was the reconstruction of the event virtually while regaining the experience of the installation and the multimedia show that consisted of visual effects and a film conceived by famous architect Le Corbusier, two music pieces composed by artists Edgard Varèse and Iannis Xenakis, respectively, and a stunning pavilion architecture conceived by Xenakis.

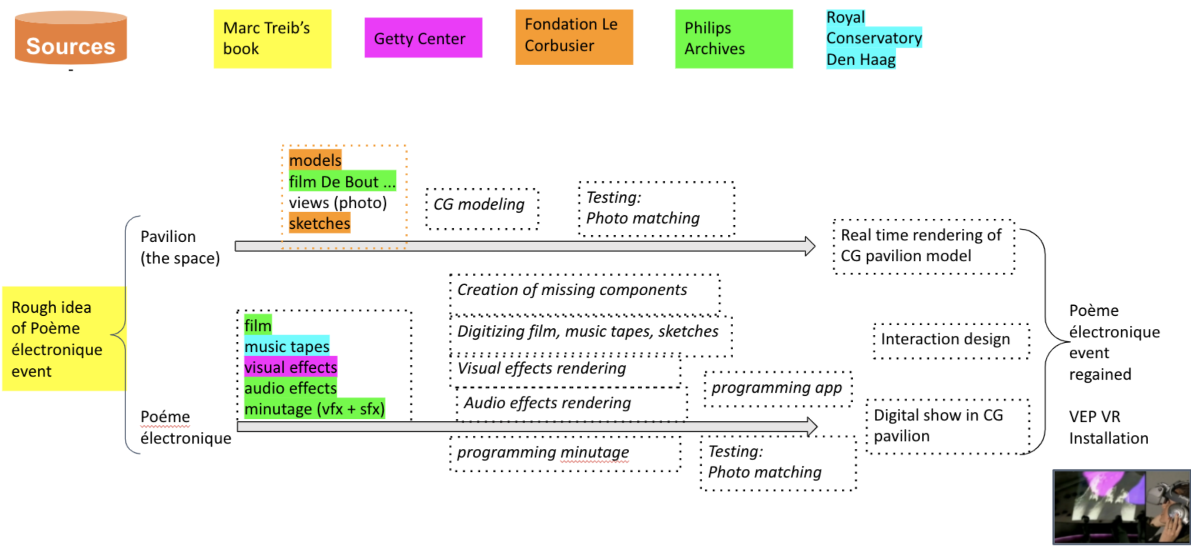


Figure 5: The visual representation of the related records and components from the archives shows from the idea until the realisation of the VEP project.

Figure 5 shows a visual representation of the related records and components from the archives from the rough idea until the realisation of the VEP project. It uses a goal-driven perspective, from the requirements of a virtual exhibition that regains the physical event through the acquisition, interpretation and visualisation of the related digital data on demand. The recorded audio consisted of tapes, stored in the archive of the Den Haag Royal Conservatory, digitised as individual tracks. The film, stored in Philips archives as a VHS tape, was digitised respectively. And finally, the visual effects were created from scratch in computer graphics, given the indications provided by the documentation available. The re-creation of the audiovisual show with all its components has become an immersive experience staged inside a computer graphics reconstruction of the Philips Pavilion.

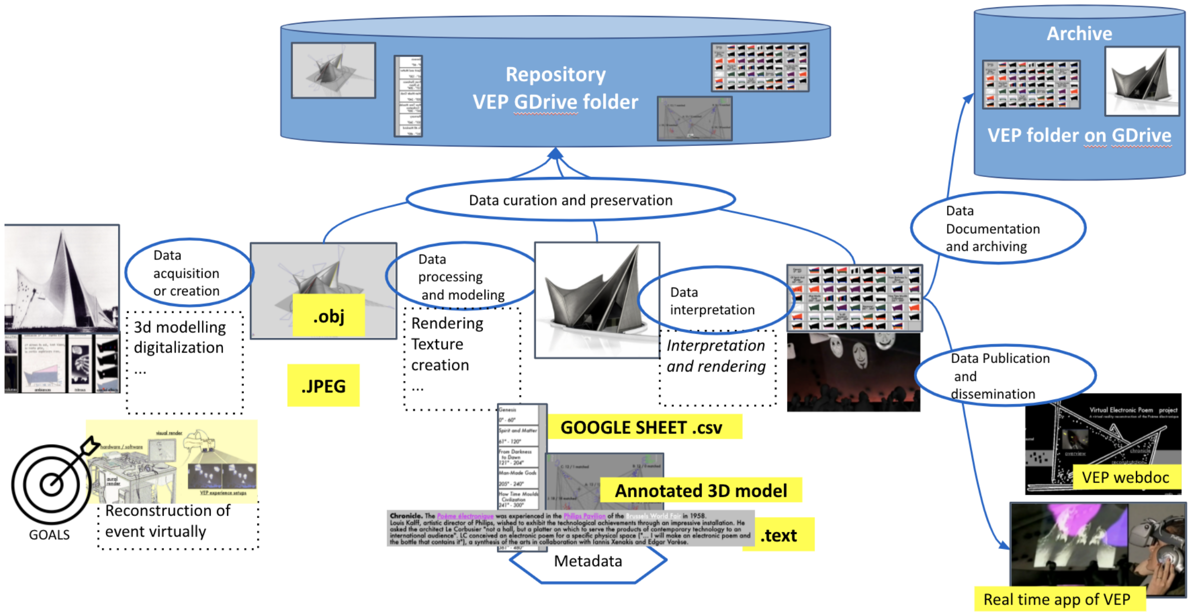


Figure 6: An abstract Digital Data Curation model for VEP

Relatively, the Digital Data Curation (Figure 6) of the VEP project starts with ***"Data acquisition"*** from institutions and archives (mentioned above and shown in Figure 5). It includes digital items such as textual files, images or sound files, along with their related identifiers and metadata. In the case of VEP, researchers collected and assigned the representation information required to understand and render both the digital material and the associated metadata which were later used in the reconstruction of the Poème Électronique experience using virtual reality and binaural audio techniques.

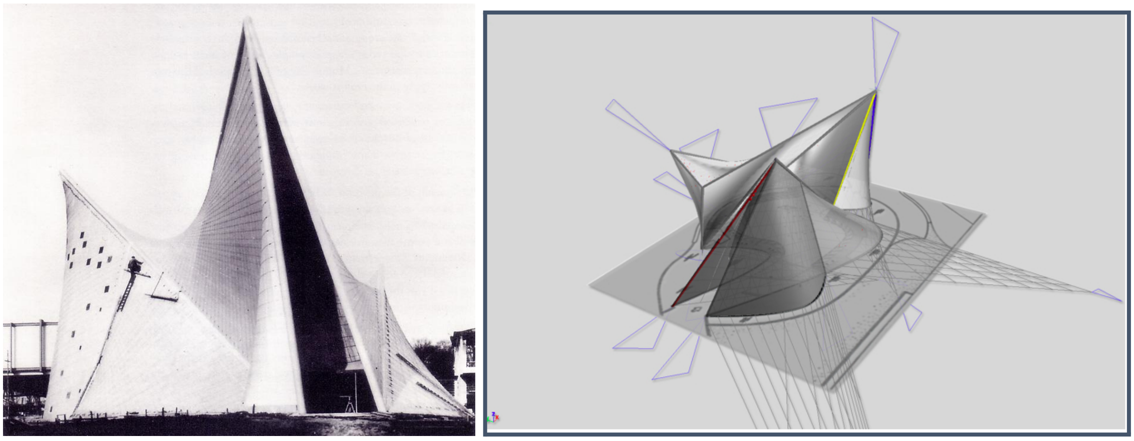


Figure 7: a) VEP project photo from the archives and b) 3d construction of the spatial components. Source: http://www.cirma.unito.it/vep

As with other elements of the Poème Électronique, the lack of documentation has made the reconstruction ***(creation)*** of the data and metadata quite a crucial task for the researchers. The reconstruction focuses on the original set-up of the installation with inconsistencies resolved by referring to the photographic material (for example the photo of the pavilion is shown in Figure 7a), sourced primarily from the Philips Company Archives and the Getty Center. In this context, sketches and archival material are used in the process of data creation.

The material for 3D construction (Figure 7b) is reconstructed from written and photographic information, the sound aspect of the project spatialization is also the result of a generalisation of the 30-second excerpt's spatial features and guessing of plausible solutions such as some longer sounds were split and sent to distinct clusters (Lombardo *et. al,* 2005). Once these have been agreed upon and communicated the validation of the final result by a committee of experts drawn from various areas is an important part of the lifecycle of the data. Therefore, receiving data also means to be in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata.

It is noticeable that audio and video technicians were in close contact with designers and programmers of the virtual exhibition, sharing formats and information about the digitised materials. The sharing of formats was of particular importance here, because of the software functioning.

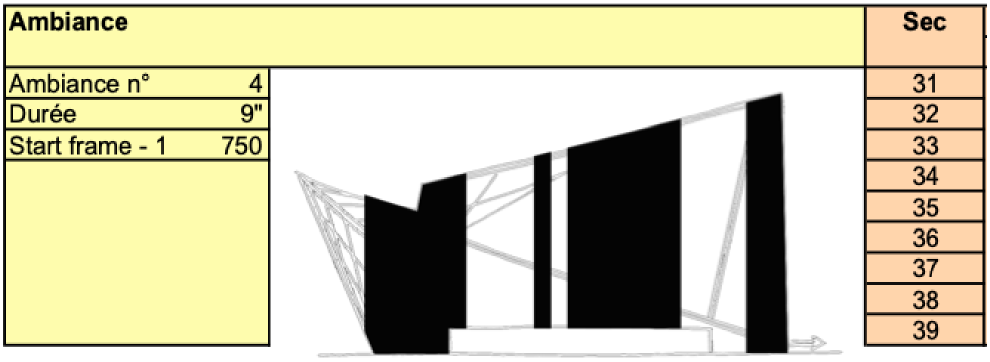


Figure 8: Interpreted and rendered minutage. Source: http://www.cirma.unito.it/vep

The sketch (Figure 8) from the archive shows one of the ambiences as the visual effect of the space in a specific time slot. The reconstruction of the minutage took into account the variations introduced by Le Corbusier, Xenakis and Phillips Company's artistic director M. Louis Kalff at the last minute in Brussels, while some ambiences were deleted or simplified concerning the original setting considering there was no evidence found of their implementation in the final exhibition accounting for multiple versions and related changing parameters.

The processing and modelling phase reflected on reading the annotations and the numerical data which provides metadata such as duration and the usage of the materials as well as the sequence of the ambiences within the event. The data interpretation is based on the findings and how to transform this knowledge into visual effects rendering and it becomes a part of the minutage that provides the sequences of all the components of the event. Also, some parts of the movie were changed because of the deletion of some sections. Therefore the visual effects are reconstructed by taking into account the original technical setup of the installation and the schemata (diagram) sketched by Le Corbusier aligned with interpretations and rendering.

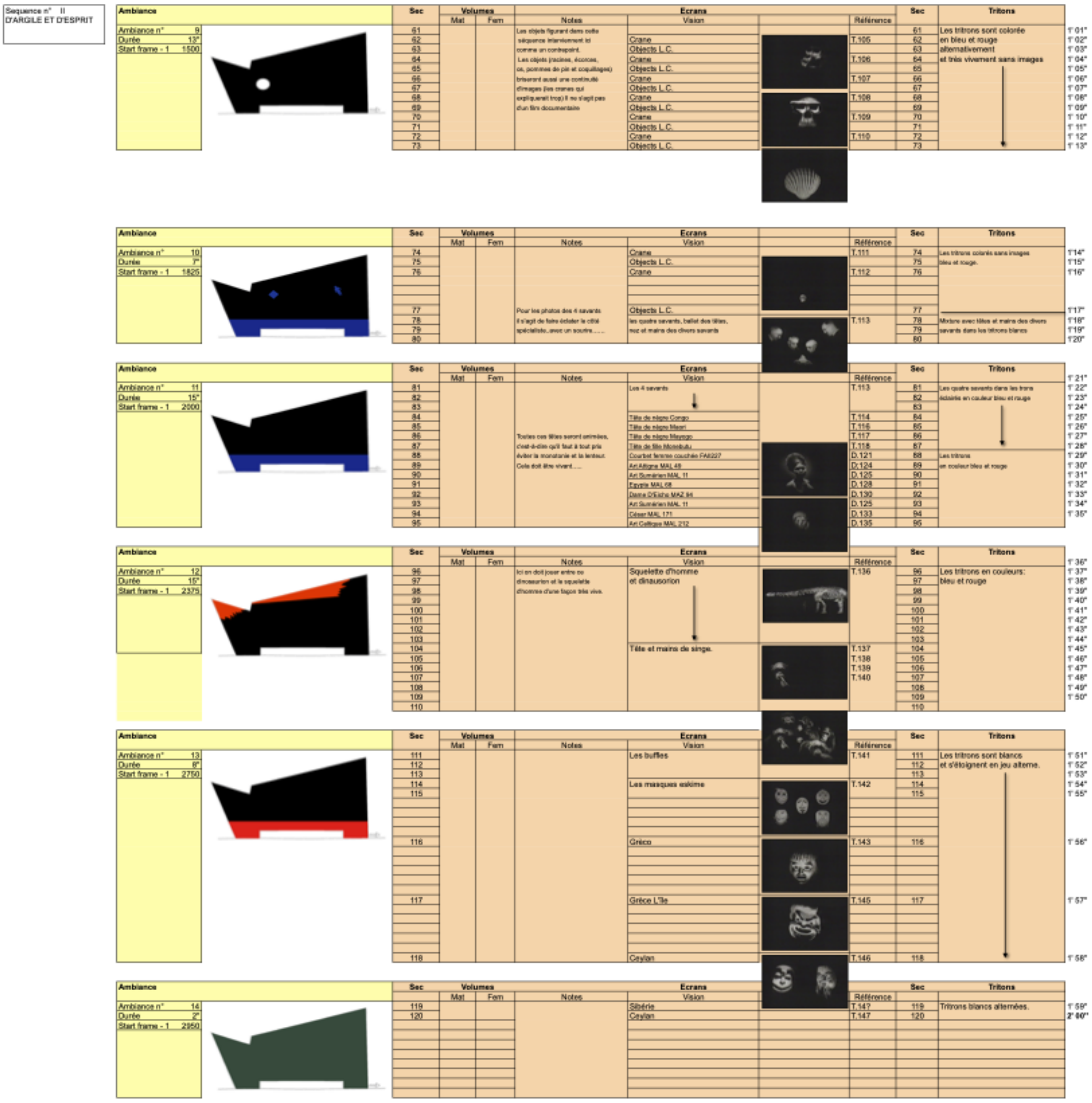


Figure 9: The reconstruction of the minutage. During the process, many ambiences were deleted or simplified concerning the original setting. Source: http://www.cirma.unito.it/vep

Then, sounds and visuals were spatialized and temporalized according to the control score, which assigned sounds to groups of loudspeakers in time (animated sound effects) and visuals to the walls of the pavilion. Even the control score, retrieved from a small excerpt on a journal paper, was reconstructed together with the pavilion structure as seen in Figure 9[21].

[21] http://www.cirma.unito.it/vep/minutager.html (last visited on 10 December 2021)

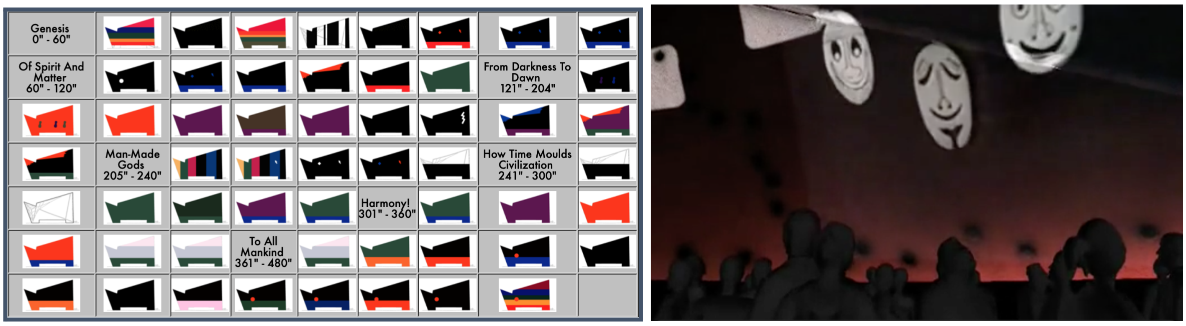


Figure 10: Interpreted and rendered ambiences a) sketches b) film excerpt. Source: http://www.cirma.unito.it/vep

Researchers undertake these actions planned throughout the ***data curation and preservation*** phase by considering the fragments (Figure 10a) of the various components, such as photographs and drafts of the architecture, the projected film (Figure 10b) from the Philips Archives and the recordings of Varèse's and Xenakis's music (Lombardo et. al, 2005). ***During the data interpretation process,*** as a result, there were four available settings to provide an appropriate user experience for the reconstructed Poème Électronique.

Figure 11: Screenshots from the event can be found on the website dedicated to the VEP project. Source: http://www.cirma.unito.it/vep

Although the most immersive setting was a single-user setup, in immersive virtual reality, there were also two settings to accommodate several people at a time: a multiple-user setup, with a large screen and binaural audio through headphones, and a multi-channel setup, with a large screen and multi-channel audio through loudspeakers (Figure 11).

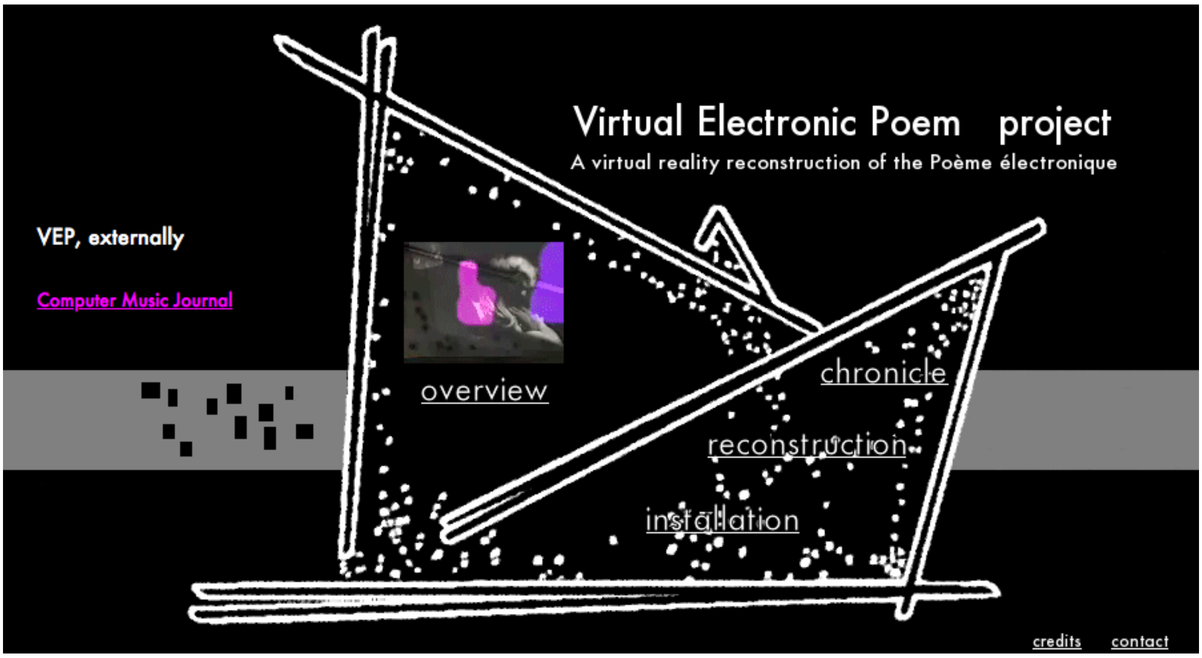


Figure 12: Screenshot of the homepage of the VEP website[22]

While the VEP installation is released through the major setups and a reduced web version (Figure 12), ***the documentation and archiving*** of the related documents and results of the analysis are in the repository of the scientists' personal storage. Considering the user experience aspect through ***the data publishing and dissemination purposes***, alongside its application as a virtual reality exhibit, several records and sections can be accessed through the website[23]: chronicles, reconstruction, installation of the Computer Graphics (CG) Pavilion as well as the architecture of the Philips Pavilion (Lombardo *et al.*, 2005).

2.3.2 Case Study: Ename Crosier

Ename Crosier[24] (Figure 13a) is an archaeological finding of the ivory head of an abbot's staff and was excavated in 1995 at the archaeological site of Ename, Belgium. In the case of the Ename Crosier, it is believed that it must have been broken because the two bronze bars that traverse the object horizontally have been added to keep the broken pieces together. To show how this object looked in its original state, it needed to be virtually reconstructed. At

[22] (http://www.cirma.unito.it/vep) (last visited on 14 December 2021)
[23] https://www.cirma.unito.it/vep/VEP_documentary.html (last visited on 16 November 2021)
[24] A crosier (also known as a paterissa, pastoral staff, or bishop's staff) is a stylized staff symbol of the governing office of a bishop or abbot and is carried by high-ranking prelates of Roman Catholic, Eastern Catholic, Eastern Orthodox, Oriental Orthodox, and some Anglican, Lutheran, United Methodist and Pentecostal churches (Chisholm,1911).

the end of the project, the object was reconstructed digitally (Figure 13b) and is currently on display in the PAM Ename Museum (Capurro *et al.,* 2015).

Figure 13: a) archaeological find, the 12th-century Ename crosier, made in ivory and its b) virtual representation (Credits: Pam Ename and Vidi)

Like a physical restoration, the virtual reconstruction of the object is also based on the extensive study of the object and its context from material and structural aspects, which need a multidisciplinary approach and full documentation. The most common techniques for virtual reconstruction include the use of depth maps of digital sculpting on digitised 3D models or 3D modelling. In 3D computer graphics and computer vision, a depth map[25] is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint. It simply creates a distance representation of the image from a reference point and provides details of depth based on how near and how far away, in terms of perspective, a part of the image is. For example, in the depth map, the background is considered the farthest away from the viewer.

[25] https://en.wikipedia.org/wiki/Depth_map (last visited on 16 December 2021)

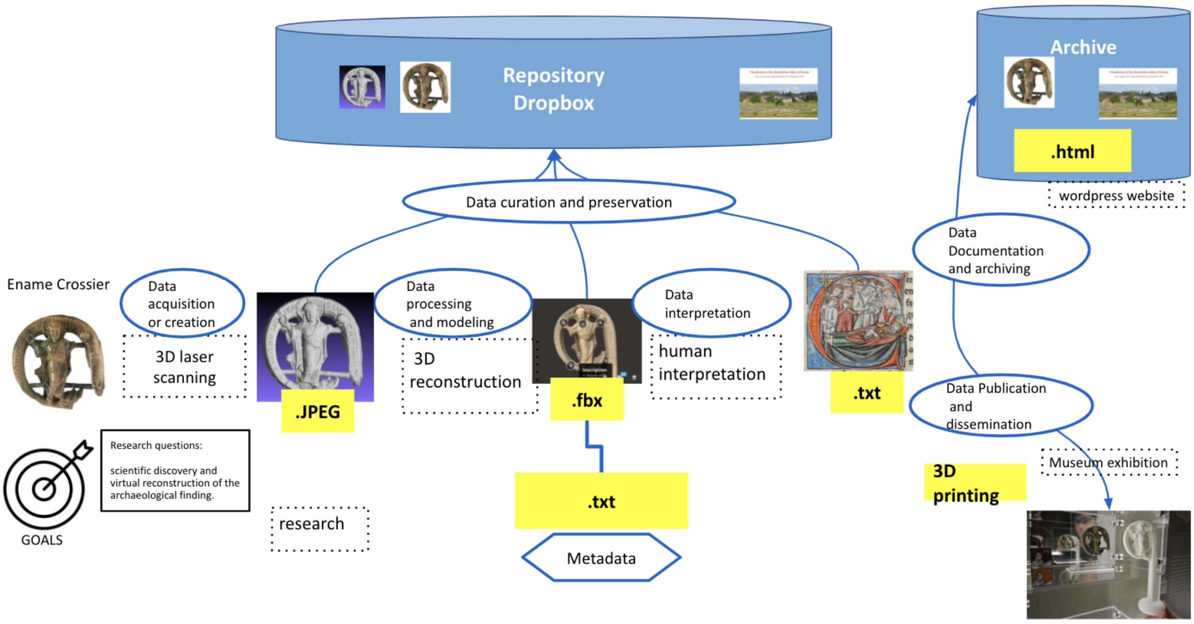


Figure 14: Digital Data Curation workflow for Ename Crosier

As seen in Figure 14, the proposed DDC workflow is based on the proposed digital curation of the project. For data acquisition, the depth map image (.jpeg) contains information about the distance between the surface of objects from a given viewpoint. A depth map is created from a source image and is typically in grayscale format. When merged with the source image, a 3D model is created. During the ***data acquisition process***, the crosier was 3D laser scanned in high resolution (Figure 15), which produces the resolution of about 50 micrometres, 1/20 of a millimetre alongside XRF measurements were also applied to understand if gold was applied to the object (further details can be found on website[26] dedicated for the visualisation of the Benedictine abbey of Ename).

[26] https://enameabbey.wordpress.com/2017/06/18/the-story-of-a-masterpiece-part-2/#comments (last visited on 16 January 2022)

Figure 15: The highly detailed 3D scan of both sides of the crosier (Credit: Visual Dimension bvba).

The processed data come from different techniques, such as XRF and 3D scan, stored in different software formats. The specialists collect all the data and produce reports to inform 3D designers as a part of ***the data modelling process***. At the end of the data modelling and processing phase, there is a reconstructed version of the physical object stored in the software called Blender[27] as a 3D drawing. The measurements are also collected during the modelling and processing of the data in a 3D format using the Blender and Zbrush[28] software and transformed to FBX[29] (Filmbox) which is a proprietary file format (.fbx) developed by Autodesk. It is used to provide interoperability between digital content creation applications.

For the ***data interpretation***, while the specialists analysed the object itself, they also looked at similar objects to understand the period it was made and used and the alterations that were made. The data processing manually (by humans not machine) between about 200 similar mediaeval crosiers circa 500 images from mediaeval manuscripts to produce knowledge. In this process, knowledge is stated as text in natural language with selected pictures and videos from the archives. Later, they used historical facts modelling and production of 3D models with the data interpretation of the object (based upon the laser scan) and data processing (based upon the detailed art history study). The result is both a 3D model as well as metadata in the form of annotation in the 3D model.

[27] https://www.blender.org/ (last visited on 16 March 2022)
[28] http://pixologic.com/features/about-zbrush.php (last visited on 16 March 2022)
[29] https://www.autodesk.com/products/fbx/ (last visited on 16 March 2022)

Figure 16: Metadata annotations on SketchFab. Source:
https://sketchfab.com/3d-models/ename-crosier-original-state-2d3500a046484714af9b5c533adcd7c5

As seen in Figure 16, there is a 3D reconstructed version of the Ename Crosier that shows the annotations which are based on the data interpretation process consisting of processed data which is enriched by specialist knowledge using the metadata which was gathered in the previous steps. After the data interpretation process, the virtual replica of Ename crosier is uploaded in the SketchFab[30] software. Sketchfab is an open-source online platform to upload, download, view, share, sell and buy 3D assets. Essentially, as it serves as a repository for 3D files on the web, annotated by the creators, the model contains the necessary metadata for the next steps.

During the data interpretation process, the 3D model is reconstructed based on archive searching and some research on the same typologies and the iconography from the relevant times which is stored in the SketchFab repository as well as Dropbox[31] of the virtual reconstruction production company in Belgium namely Visual Dimension[32].

In this project, the specialists also use Dropbox as a digital repository for ***data curation and preservation***. There is no database or API server. Since it is a museum object, the physical piece is always stored in the museum and ready for re-examination in future.

[30] https://sketchfab.com/ (last visited on 6 March 2022)
[31] https://www.dropbox.com/ (last visited on 6 March 2022)
[32] http://heritage.visualdimension.be/ (last visited on 6 March 2022)

Figure 17: 3d printed crosier is displayed next to the real object in the Pam Ename museum (Credit: Visual Dimension bvba).

For ***data publication and dissemination purposes***, 3D printed in polyamide on a scale of 1/1 (Figure 17) is created through digital sculpting based on the previous steps. This physical replica is based upon previous steps and turned into a storytelling device or interface to explore the cultural heritage objects in the museum's educational department as a part of the user experience of museum visitors.

The goal of displaying replicas in the exhibition setup is to create a multimodal experience where the user can hold the smart object, feel it, manipulate it in real-time and explore its story through images and sound. This setup gives the user full control of the object and its content while creating a personalised experience (Capurro *et al.*, 2015). Also in terms of data publication and dissemination, all the steps and the related metadata are archived on the website of the Ename Abbey and published on the WordPress website[33]. Considering the complexity of the project and the implementation of digital technologies within the archaeological context, the project is a great example of the journey of digital data from scientific research to the exhibition.

[33] https://enameabbey.wordpress.com/2016/10/18/the-story-of-a-masterpiece/ (last visited on 1 January 2022)

2.3.3 Case Study: Venus Pompeiana Project

The Venus Pompeiana Project (VPP) is a collaboration between the University of Missouri-Columbia and Mount Allison University, under the auspices of the Pompeii Archaeological Park in Napoli, Italy. In this project, the digital data curation addresses specifically the excavations of the cistern in Area IIS of Venus Pompeiana temple to understand the archaeological relations and possible virtual reconstruction of the findings.

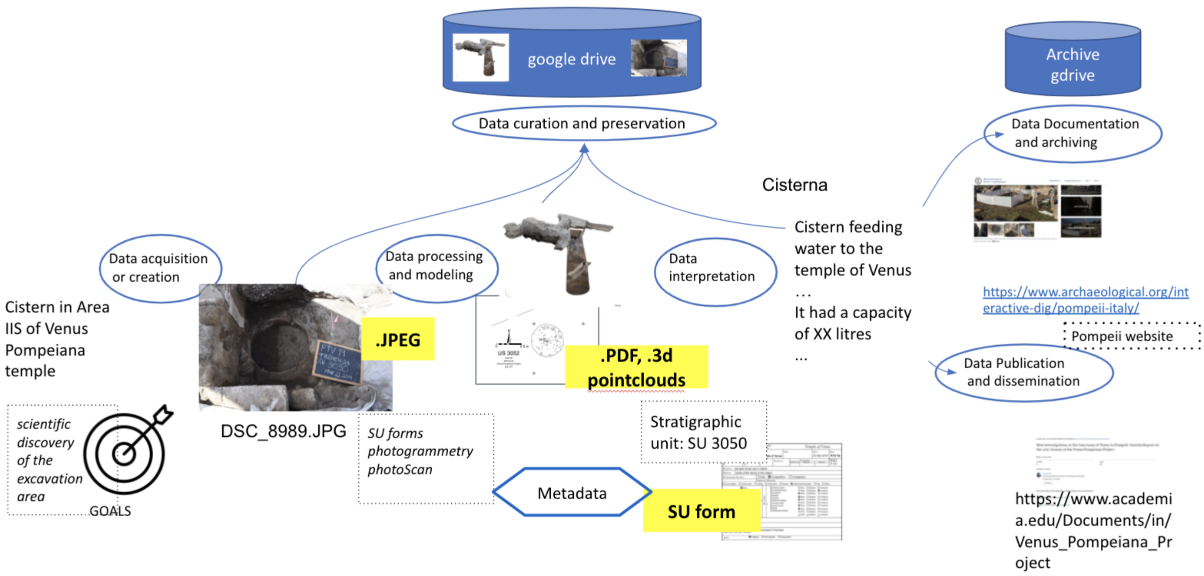


Figure 18: Digital Data Curation workflow for Venus Pompeiana Project

The general framework of digital data curation (Figure 18) illustrates the model with an example of digital curation of data which is related to digital data generated from the archaeological site. So, each identifier is built as follows: SU represents Stratigraphic Unit, VPP represents Venus Pompeiana Project whereas AS represents Archaeological Site.

Figure 19: Digital image is taken during the excavation of the VPP area.

The ***data acquisition*** of the area was based on photogrammetry taken using a Nikon D750 DSLR camera with a fixed 50 mm lens. The digital image, namely DSC_8989.JPG (Figure 19), is produced as a digital image. The digital image itself also has some annotations such as the date (i.e. May 22, 2019) and the number of the Stratigraphic Unit (i.e. SU 3050) during the ***data modelling*** phase. Even though this information can be interpreted by scientists, they are not machine-readable. The data based on photogrammetry is interpreted later by using Agisoft PhotoScan Professional to process the photos into 3D point clouds.

Figure 20: Virtual construction of cisterna for Venus Pompeiana Project.
Credit: Venus Pompeiana Project team.

Based on the measurements that are also collected in the data acquisition process, with the modelling and processing of the data in 3D point clouds, the 3D reconstruction (Figure 20) of the image is created. Later in the ***data interpretation process***, the scientist interpreted the cistern as feeding the area and the potential drainage placement etc. based on the findings.

The relevant data and media is ***preserved and curated*** through an external repository namely Google Drive and belongs to the scientist's personal accounts and external disks.

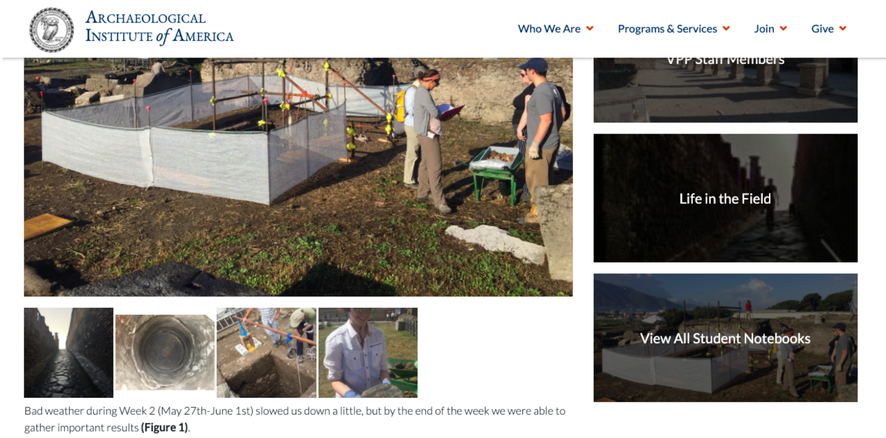


Figure 21: Website for Venus Pompeiana Project.

Source: https://www.archaeological.org/interactive-dig/pompeii-italy/

All the steps and the related metadata are ***archived*** on the website (Figure 21) of the project ***published*** on the web[34] as written in WordPress format. The results are also published as scientific papers in academic journals.

Conclusion of the chapter

In conclusion, using previous findings and insights from the literature review (special focus on User Experience and Knowledge Enhancement), this thesis discusses the related tasks, tools, technologies and software for digital curation in the field of archaeology. We examined three different projects using the schema of the Digital Data Curation model. We constructed the digital data curation models of these three projects through close collaboration and several discussions with the data creators, curators and custodians of the projects. In terms of the digital tools and techniques even though the Ename Crosier and Cisterna example uses the same methodology as virtual reconstruction the tools and software are different. Although each project has its own specific goal which affects the tools and techniques used throughout the project, they are representative of many projects in the field of archaeology. From the perspective of user experience, as seen in all of the examples, scientists use external repositories to preserve the data such as Dropbox (Ename), and GoogleDrive (VEP, Cisterna) which makes data publishing and dissemination difficult. In addition, the dependence on the external repositories (SketchFab and Dropbox) in addition to the lack of

[34] https://www.archaeological.org/interactive-dig/pompeii-italy/ (last visited on 19 November 2021)

digital curation strategy makes the data and metadata vulnerable in the preservation and curation phases[35]. On the other hand, VEP and Ename Crosier have been considering the user experience aspect quite extensively, the Cisterna project has a highly specialised scientific community as users. In this sense, we can easily access the data in two projects whereas in the other we had to dig into the personal archives.

In terms of knowledge enhancement, there are traces of semantic tools and relations but in fact, none of the projects has adopted any digital curation approach from the beginning of the conceptualisation of the project. Although digital curation helps manage properly there are still some issues that need to be solved; submission of new versions of the original asset and related metadata, enrichment with dynamically evolving linked data in an information delivery environment whether it is an online publication, complex database or exhibition (Gavrilis *et al.*, 2013). Therefore the semantic relation between the archaeological domain and their digital representations is not aligned with any standardisation. We believe that also Digital Data Curation can benefit from a Semantic Web approach. In the next section, we propose to extend the analysis perspective toward crucial elements that are emerging around and beyond the Semantic Web vision and technologies, with particular respect to applications of archaeology while examining the relationship between semantic tools and archaeological data through the latest technologies and tools to create fruitful representations. Because without regular, sustained processing of the data of the cultural heritage assets, digital curation activities are only put into practice intermittently which eventually creates a gap between practices and multiple actors.

[35] In fact, in the case of Ename Crosier, the metadata is annotated on a 3D model using the SketchFab platform. However, during the process of writing this research, New York-based startup Sketchfab was acquired# by Epic Games, another tech company behind online video game Fortnite and game engine Unreal Engine. Since the terms of the deal were not disclosed by Epic Games, the future of Sketchfab is unclear in terms of the preservation of the previously created 3D models. It's evident that the dependence on external software in preserving the data is quite important in terms of long use of the data in future.
Source: https://tcrn.ch/3BqRCRe (last visited on 15 January 2022)

Chapter 3

Semantic Web and Archaeology

In this chapter, we start with the examination of the archaeological data in the context of the Semantic Web. We introduce the basic characteristics of the technologies underlying key elements. Specific attention is paid to the technologies that are involved in the high-level Semantic Web architecture (the "Semantic Web pile"), with particular respect to the languages enabling semantic interoperability (RDF, RDFS and OWL) and query on RDF triples (SPARQL). We provide some technical definitions of the semantic tools, terms and concepts and then go further through exemplifying their usage. Later, we expand on the Semantic Web and Linked Open Data issues in the context of archaeology to understand their contribution to the field. We discuss the archaeological activities within the application domain in the context of semantic encoding, to achieve interoperability for digital data.

3.1 Digital Archaeology

Archaeological projects go digital in all their phases: data collection, curation, and visualisation (Berggren, 2015), data analysis (Forte, 2012), and the exhibition of the results (starting from the virtual archaeological reconstructions of the 1990s) and addressing general public outreach and participation. In this matter, archaeological data, additionally defined as "records"[36], is a primary outcome of observations made throughout the fieldwork, which reflects archaeological entities, their attributes and relations – as historical

[36] Scholars have frequently used in textual analogies such as 'record', 'source' and 'archive' to refer to material evidence of the past relatively the term 'archaeological record' probably originated this way, possibly via parallel concepts as geologic record from field of geology or fossil record from the field of palaeontology (Lucas, 2012).

sources, through formation theory, and as material culture – then reveals ways to connect these three domains through a reconsideration of archaeological entities and archaeological practice (Lucas, 2012). Archaeological documents preserve all sorts of data: structured data (e.g. tables), unstructured data (e.g. field drawings), semi-structured data (e.g. excavation diaries), and their numerous combinations (e.g. survey forms). The difference between context-based and feature-based excavation documentation is crucial when differentiating the archaeological data curation process. Context-based documentation is an approach to in-situ documentation during the archaeological excavation process.

For example, first discovered in the late 1950s, Çatalhöyük is amongst the very earliest urban settlements ever discovered and thus opens a crucial window onto an era that is arguably the most significant social shift in history that is the transition from hunter-gathering to farming societies[37]. In addition, since 1996, Çatalhöyük is one of the first major excavations to make its records available online. The Çatalhöyük Research Project, the 9,000-year-old Neolithic settlement in the Konya plain of Turkey, positioned itself as a flagship project for creating a long-term digital infrastructure based on its reflexive postprocessual methodology for data collection and interpretation in which archaeological information is permanently open to reinterpretation by both scholars and public.

Over more than two decades of excavation and analysis, the record of the archaeological site was constructed dynamically in situ and ex-situ as the excavation progressed, through a combination of single-context database recording (Berggren *et al.*, 2015). The relational database of Çatalhöyük consists of records from various multimedia and multi-genre material that was published both in print and in a searchable, hyperlinked form which eventually evolved into an online archive, with plans to develop it into a permanent sustainable open-access, interactive, linked-data research tool (Lukas *et al.* 2018).

Relatively, in 1999, in his book "The Archaeological Process: An Introduction" Ian Hodder mentioned the benefits of the interactive booklet of data from archaeological reports (Hodder, 1999). For example, over his administrative years for the Çatalhöyük project, thousands of printed reports had been effectively published concerning a Neolithic settlement in Turkey, carried out to maintain the data as long as possible. The Çatalhöyük

[37] From Claudia A. Engel's blog post on February 11, 2021 through https://library.stanford.edu/blogs/stanford-libraries-blog/2021/02/catalhoyuk-image-collection-released-searchworks (last visited on 15 February 2022)

Database[38] and the Çatalhöyük Image Collection Database make the documentation of the Çatalhöyük excavation site available online. These platforms allow for the search of data uploaded during every excavation season and then made available through the Çatalhöyük Living Archive[39] which tells about two decades of excavations and analyses. However digital publication and online dissemination of excavation reports, specially withinside the domain area of academic archaeology are nevertheless missing (Berggren and Hodder, 2003).

Unit: 16132

Property	Value(s)
Mound	East
Year	2007
Area	4040
UID	16132
grid_x	1060.02
grid_y	1197.08
Data category	Skeleton
Discussion	This skeleton seemed to be the latest in this part of the est platform. It was placed right in the middle with other completely disarticulated bones and arranged around it. PS X4 Gypsum bead broke and shattered completely.:
Dry sieve volume	0
Fastrack	true
Assoc. buildings	102
Assoc. spaces	17
Assoc. features	1630,3010
Assoc. features type	1630 (platform),3010 (burial)
Assoc. feature types	platform,burial
Assoc. Hodder levels	SCRAPE.?G
Conserved	07.053
Figurines	16132.H1,16132.H2
Figurine type(s)	,figurine,
Figurine primary fabric(s)	clay,clay
XFinds material(s)	Chipped Stone,Shell,Stone
Finds store material(s)	Clay,Human,Obsidian,Shell,Stone

Figure 22: Digital documentation of Stratigraphic Unit Sheet (numbered 16132) shows the scientific interpretation as well as conditions and characteristics of the stratigraphic unit. Source: http://catalhoyuk.ege.edu.tr/database/catal/FeatureSheet.asp?num=16132

In the Çatalhöyük example, Stratigraphic Units are the basic recording element of an archaeological site, which ideally represent a single identifiable depositional event and can be divided into several general categories (e.g. cut, fill, layer). As Units are architecturally determined parts of an excavation site, the Unit Sheet allows defining a whole burial, archaeological findings or architectural element, as opposed to its units to be described. As seen in Figure 22, retrieved from the Çatalhöyük database, digital documentation of the Stratigraphic Unit Sheet, numbered 16132, shows the scientific interpretation as well as the conditions and characteristics of the stratigraphic unit. It has much information as well as the

[38] http://catalhoyuk.ege.edu.tr/research/database (last visited on 5 March 2022)
[39]The website http://catalhoyuk.stanford.edu/ is currently inaccessible due to the construction (last visited on 5 March 2022)

relation with other features (16303 - 16302 - 16300 - 16302 - 16133) and several scientific annotations including the sketch of specialists. As seen in this example, a stratigraphic unit may contain many archaeological findings that potentially feature the same dating. Both these categories, Archaeological Finding and Stratigraphic Unit, must be represented, interpreted, and linked to each other.

Alongside database recordings, in contemporary digitally-enabled excavations such as Çatalhöyük, researchers engage in a range of activities encompassing the capture, description, annotation, classification, knowledge enrichment and dissemination of documents single-context sheets (such as unit sheets), digital diaries and audio memos, digital photography, reflexive video recording, sketching and drawing, stratigraphic sequence matrices, 3D models, GIS representations and hypertextual, diagrammatic and narrative interpretations of particular aspects of the archaeological record (Tung 2013; Dallas 2007).

Moreover, the transdisciplinary character of archaeological studies has increased significantly in recent years due to the growing importance of archaeometric analysis. Various scientific disciplines work together to obtain an objective description of material artefacts. For instance, laser scanning allows for the capture of archaeological sites and findings in 3D (Hammerle and Höfle 2018; Raun *et al.* 2018). On the other hand GIS, geoinformatics and computer sciences, have also served as essential tools in cultural heritage management for many years, particularly concerning the protection of archaeological sites (e.g. Ioannides *et al.,* 2014). It is evident (Levine 2014; Riley 2014) that a large amount of digital data supports the entire process and there is a great value in maintaining the consistency of information and knowledge contributed by each discipline involved.

It is necessary to model, not only texts and images but different types of objects and the data packages such as 3D models. It is possible to add value to digital objects, for example through in-depth indexing and incorporation of semantic structured data as Linked Open Data (LOD), and also by creating new contexts and developing new original services. Interoperability makes it necessary to harmonise different conceptual models for particular types of digital objects and at different levels of granularity.

3.2 Archaeological data

To apprehend the motivations behind the adoption of semantic technology and its benefits for scientific knowledge within the archaeological domain, we want to distinguish the terms: data, information, and knowledge[40].

Figure 23: The Data-Information-Knowledge-Wisdom (DIKW) pyramid or hierarchy was made a commonly used framework in 1989 by Russell Ackoff. Source: https://www.paperlesslabacademy.com/2020/01/07/knowledge-and-wisdom-in-rd/

The DIKW model or DIKW pyramid is a commonly used framework, with roots in knowledge management, to explain the hierarchy starting from the data (the ‘D’) to information (I), knowledge (K) and wisdom (W). As seen in Figure 23, each building block is a step toward a higher level and represents the relationships between data, information, knowledge and wisdom. Each layer in the figure adds value therefore the more the data is enriched with meaning and context, the more knowledge and insights are created to make better, informed and data-based decisions. The data, in data acquisition and creation, classified as being “raw,” consist of values and both quantitative and qualitative variables. It may be a spreadsheet as data: there's a column of numbers and, adjacent to it, a column with corresponding descriptions. The larger the file (data) is, the more information we will gain from it. The various and diverse amounts of information generate “knowledge”. In such a light, archaeological data are effectively seen as raw materials that, when brought together

[40] https://www.ontotext.com/knowledgehub/fundamentals/dikw-pyramid/ (last visited on 1 January 2022)

within a specific context or set of relations, become information which in turn builds into knowledge, in the classic data-information-knowledge-wisdom (DIKW) pyramid. Instead, based on their experience, research objectives etc., the archaeologist articulates their knowledge to identify and categorise information, and that information is systematised within a digital environment to create data (Huggett, 2020).

Similarly in the field of archaeology, the use of digital tools ranging from CAD (Computer-aided design) to Structure from Motion photography is increasingly employed as surrogates for traditional field drawing which, among other things, changes the nature of the engagement with the physical remains (Hacıgüzeller, 2019). The journey of digital data begins to exist the moment it is recorded by using a machine and it arguably obscures the role of human decisions in its creation (Rendgren, 2018). Furthermore, digital data can constrain and limit subsequent analysis through their structuring and organisation which ultimately determine what can and cannot be recorded, and through the set of procedures that shape the retrieval and processing of the data (Huggett, 2015).

In general, archaeologists struggle with the employment of digital tools when conducting excavations, whereas the perceived nature of the curation of collected data could be a typical existence as a separate process concerning the archives (Dallas, 2015). Although the task of managing and preserving archaeological data is not a straightforward one, digital tools can help the effective exploitation of archaeological knowledge. In particular, the development of the advanced technologies regarding the embodiment of all the archaeological domains corresponds to diverse fields from paleoanthropology to bioarchaeology, environmental archaeology and lately archaeometry. Archaeometry represents the interface between geosciences and archaeology and aims at contributing to the solution of cultural-historical questions (Wagner, 2007). In addition to this conventional perception, this research takes account of the digital dimension of archaeological projects, including its relation to archaeometric processes. In the latter case, although several representational models have been developed over the years to account for the registration of the archaeological process in databases, there is one upper reference model[41], called CRMsci, that accounts for the representation of scientific investigations in general.

[41] There is also notable research about Toulmin model on arguments construed as the semantic entities that are the outcome of processes of reasoning mainly focus on artificial intelligence application (check Katzav and Reed’s article on Argumentation Schemes and the Natural Classification of Arguments in 2004). However since the scope of this research is cultural heritage and archaeology specifically, we review the semantic web technologies in this context.

Also, recently there have been some initiatives such as FAIR (Findable, Accessible, Interoperable, Reusable) data principles, and the IIIF (International Image Interoperability Framework) for interoperability. In addition, there are some data standards and process models such as the OAIS reference model (ISO:14721:2003) and ISO 21127:2014[42] developed by the International Council of Museums (ICOM-CIDOC) which provide guidelines on how to support the effective exchange of knowledge and the retrieval to store data in the long-term. Specifically for the archaeological data, English Heritage's Center for Archaeology (EH-CfA) also established some principles over time (Cripps *et al.* 2004; May *et al.* 2008). It is still a common practice in archaeology to build and use a database without an explicit publication of the data structure and the relationships between the entities. This happens with legacy and modern datasets alike, but the effort of building a model of data structure is essential for the preservation and re-use of a dataset.

The structure of a recording system can be restored from the existing dataset. If an excavation uses a relational database, its conceptual model (generated by the software) may serve as a basis for the data model. Otherwise, there are data modelling languages and tools that may assist. In this context, several instruments for data modelling, interpretation and curation of cultural heritage knowledge were developed in recent years to account for the registration of the archaeological process in databases, and more recently several semantic models have been developed to account for the link between institutional forms and formal knowledge of excavation, archaeological and related finds. In the following section, we will investigate the semantic tools and technologies in general with their specific relation to archaeological information management.

3.3 Knowledge Representation in Archaeology using Semantic tools

After the diffusion of Web 2.0, many applications have been developed, and with them many data formats and query languages focusing attention on the semantics of information such as eXtensible Markup Language (XML) a standard format for data exchange; Web Application Programming Interfaces (APIs) that are methods to request data from a web server to use them in external applications; Web services standard protocols and architectures specified by the World Wide Web Consortium (W3C). The main idea of Web 3.0, also known as the Semantic Web, is the information shared on the web should be

[42] ISO Standard "Information and Documentation: A Reference Ontology for the Interchange of Cultural Heritage Information" (ISO 21127:2006 - 2014).

interlinked. According to the definition may be found in the article[43] "The Semantic Web" (Berners-Lee *et al.*, 2001) "... it is an extension of the current web in which information is given well-defined meaning, better-enabling computers and people to work in cooperation". Eventually, it follows the idea of interoperability and linking contents (data) embedded in documents (HTML pages) in a global information space that is considered an extension of the current web (Berners-Lee *et al.*, 2001).

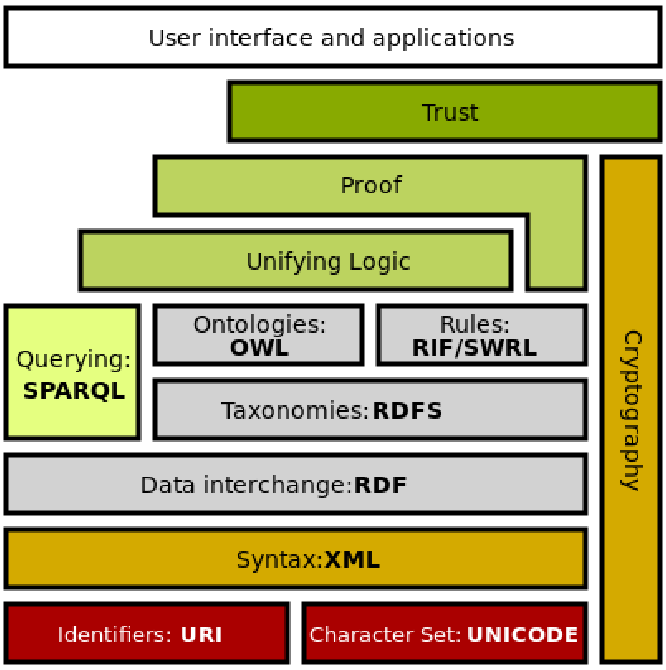


Figure 24: Semantic Web stack. Source: https://en.wikipedia.org/wiki/Semantic_Web_Stack

In this context, Semantic Web (also known as Web of Data) is a term first proposed by Berners-Lee in 2006 (Berners-Lee *et al.*, 2001) and formally defined by the World Wide Web Consortium (W3C) in 2011, is widely discussed in the digital humanities and heritage communities. The need to represent data and information in meaningful ways in order to allow better automated processing has been one of the main objectives for several years. The Semantic Web currently relies upon many technologies that represent the building blocks of a “stack or pile” (Figure 24), which is important to introduce here with particular reference to the languages that currently exist for expressing semantics at different levels. For example, Unicode[44] is the standard text encoding format that allows machines to cooperate because it ensures that text characters are independent of the software platform, the application, and the language used. In the “Semantic Web”, the need to express uniquely

[43] Scientific American Journal article "The Semantic Web" (Berners-Lee et al., 2001)
[44] http://www.unicode.org/ (last visited on 1 December 2021)

data and information on the Web is satisfied by the Uniform Resource Identifier (URI) that identifies the representation of a resource and a new W3C model: Resource Description Framework (RDF), an XML-based language that deals with semantic interoperability which is the possibility for machines to exchange information based on the meaning of information rather than its format.

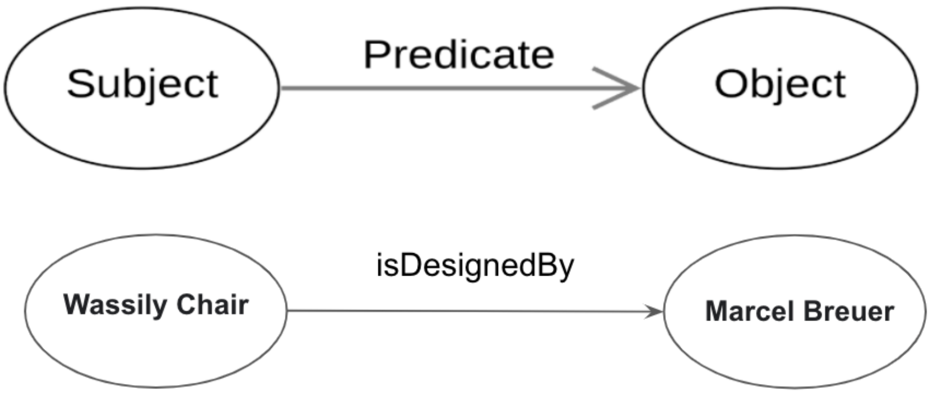


Figure 25: Top image shows the structure of the triple, and the bottom image represents the relationship between Wassily Chair and its designer Marcel Breuer.

RDF is based on triple stores and it consists of individuals, kinds of things, properties of things and values of those properties. In an RDF database, each original statement has at least three additional statements made about it: one to assert that its subject is the identifier of a resource, one to assert that its predicate is the property or attribute of the subject that has to be expressed, and one to assert that its object. RDF is a labelled, directed multi-graph by defining a set of nodes and linking them through oriented relationships; a triple is represented by the connection of two nodes by an arc. As seen in Figure 25, the top image shows the structure of the triple, and the bottom image represents the specific relationship between Wassily Chair and its designer Marcel Breuer. This conceptual representation shows simply that the Wassily Chair is designed by Marcel Breuer, in other words, the designer of the Wassily Chair. Additionally, the statements about the original statement may also exist as needed.

The ontologies represent the most advanced way of representing knowledge on the web, enriching RDF Schema (RDFS), and defining relationships between concepts (Bruseker, Carboni, and Anais, 2017). An ontology is described as “an explicit specification of a conceptualization”, formally defining a common set of terms that are used to describe and represent a domain area of knowledge (Gruber, 1993). Most ontologies describe individuals (instances), classes (concepts), attributes and relations which are commonly encoded using ontology languages (such as OWL in Figure 24). OWL (Web Ontology Language) is a family of languages that currently represents the most advanced way of representing domain knowledge on the Web in order to enrich the capabilities of RDFS in being used for automated reasoning (Mantegari, 2010).

Query, on the other hand, is a precise request to retrieve information retrieval from databases and information systems through a computer language (such as SPARQL). SPARQL (Simple Protocol and RDF Query language) is both a query language for RDF and respectively for OWL and it serves as a protocol that enables requests in a Web environment (Prud'hommeaux and Seaborne, 2008). There are other blocks in the pile, such as the RIF (Rule Interchange Format), which is a W3C proposal for the semantic Interchange of rules that may be useful for automated reasoning and are expressed in different languages (such as SWRL, WRL, FOL RuleML, and KIF). However, some blocks (coloured green in Figure 24) are not directly relevant to the context of this research, but important to conclude the formal definitions of the stack of Semantic Web. The Unifying logic considers the need for a logic upon which the mediation between the information and knowledge representation of the lower levels and the issues connected to its dissemination to users (shown in Figure 23 on upper levels) is built. The Proof considers the issues connected to establishing the truth of statements, whereas Trust refers to the different levels of fiduciary arrangement based on trustability during the information retrieval process from the Web. Finally, information is accessed on the Semantic Web through the components of the User Interface and applications. The user interface (UI) serves as the point of human-computer interaction and communication in a device (such as display screens), an application or a website.

Constructed through all these components mentioned above, one of the promises of the Semantic Web is semantic interoperability which means enabling different agents, services, and applications to exchange information, data and knowledge in a meaningful way. RDF and OWL are constructed to enable semantic interoperability for agents, services, and applications by modelling, querying, validating and linking machine-readable data. In other words, the structured machine-readable data (mentioned as Linked Data in previous sections) are interlinked concepts, people, and places coming from the different datasets and are more meaningful through semantic queries.

3.3.1 Metadata

Supporting nearly all of the steps in the digital curation lifecycle, metadata are tantamount to importance to the data themselves (Levine 2014; Riley 2014). For a document, metadata might include a collection of information like the author, file size, the date the document was created, and keywords to describe the document or the metadata whereas metadata of an image file might include the creator's name, the pixel qualities, and the year it was taken. Because without regular, sustained processing of the data of the cultural heritage assets, digital curation activities are only put into practice intermittently, and this eventually creates a gap between practices and multiple actors. For example, some interpretations provided by some archaeologists can be later disconfirmed by some archaeometric data, and the data system must record dependencies between data and interpretations to be revised later. So, in general, such information may include the scientific context underlying the data as well as who collected the data, why the data were collected, and where, when, and how the data were collected. The resulting metadata deals with attributes “that describe, provide context, indicate the quality, or document another object (or data) characteristics” (Greenberg, 2005).

In order to be useful, metadata are structured using their common components (such as dates, names, and places) schemas and standards. There are different sets of metadata standards used by diverse scientific communities to ensure the correct use and interpretation of the data by its owners and users, such as, e.g., EML (Ecological Metadata Language), and ISO 19115 (International Organization for Standardization Geographic information metadata). Standardised structure and consistent metadata are usually expressed in machine-readable extensible markup language (XML) that can be represented in other human-readable formats (e.g., HTML, pdf, etc.) (Curdt *et al.*, 2012). For metadata sharing of electronic resources, widely used for scientific articles, is achieved through the Digital Object Identifiers (DOI), also popular for scientific primary data (Brase, 2011). For example, in the DOI name 10.1000/129, the prefix is numbered 10.1000 and the suffix is 129. DOI names can identify creative works (such as texts, images, audio or video items, and software) in both electronic and physical forms, performances, and abstract works[45] such as licences, parties to a transaction, etc. Usually, the basic level of metadata is entrusted to the Dublin Core Metadata Element Set (DCMES) which is a widely accepted

[45] https://www.doi.org/faq.html#1 (last visited on 12 November 2021)

standard, which covers all general requirements and information to describe a dataset. Basic information is typically Title, Creator, Subject (e.g., the name of the project), Description (abstract), Publisher, actual Contributor of some record, Date, Type of dataset, Identifier (e.g., DOI), and primary Language.

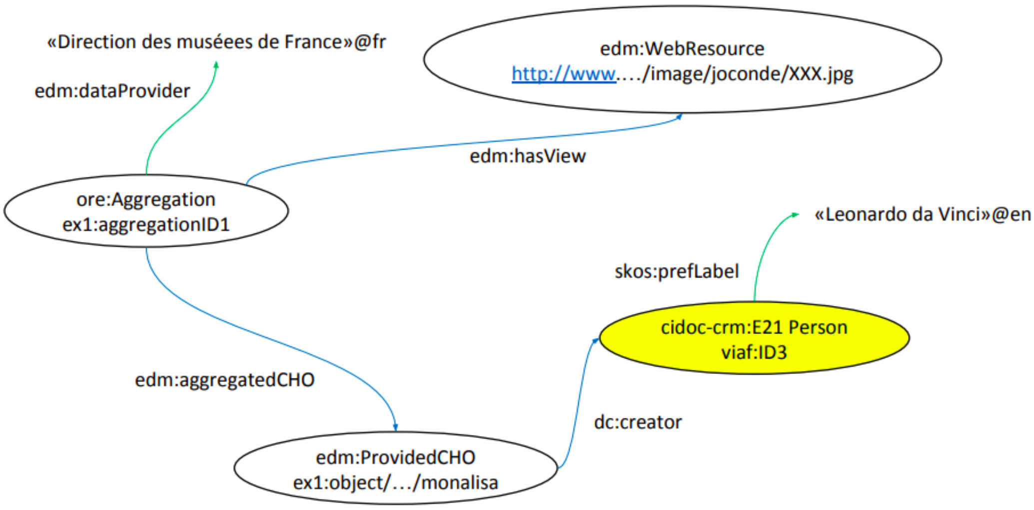


Figure 26: An example of Cultural Heritage domain and descriptive metadata on EDM Source: https://www.slideshare.net/antoineisaac/data-modelling-at-europeana-and-dm2e-smw13

Figure 26 shows a fictitious example of the creation of the conceptual model of a cultural heritage object. In this specific example, some data properties of "Mona Lisa" are shown. A new Resource is introduced to be able to unambiguously refer to a set of Web resources. This new resource, named *ex1:aggregationID1*, has a URI and a machine-readable representation that provides details about the Aggregation. It lists the Aggregated Resource here as the OAI-ORE[46] provides class *ore: aggregation* relationship. This URI *(ex1:aggregationID1)* can be seen through the related web source *(http://www.…/image/joconde/XXX.jpg)* which can be reached through the relation or property *edm:hasView*. Here, also the resource is mapped through the provider relationship (*edm:dataProvider*). Dublin Core terms are used for the authorship of the creator (*dc: creator*) which leads to a person specified as *E21* with CIDOC-CRM. Classification into this can be expressed with SKOS *(skos:prefLabel) in English for the representation of the name of Leonardo da Vinci*.

On the other hand, several cross-discipline metadata standards have been adopted in archaeology and Cultural Resource Management (CRM) in the 1990s. Dublin Core

[46] https://www.openarchives.org/ore/ (last visited on 7 December 2021)

(http://dublincore.org) is an ISO standard schema that allows for providing a basic description of any digital resource. It consists of fifteen core elements that may be repeated or extended (Gladney, 2007; Wise and Miller, 1997). The Dublin Core[47] schema can be easily applied to archaeological resources and validated manually or using one of the tools for metadata creation, extraction or conversion, as mentioned previously.

3.3.2 Taxonomy, thesaurus and vocabulary

The management of metadata additionally involves controlled vocabularies, taxonomies, and thesauri to support the effective exchange of knowledge and interoperability and keep a consistent global database. A controlled vocabulary is an established list of standardised terminology for use in indexing and retrieval of information to support consistent, accurate, and quick indexing and retrieval of the content of the digital asset using descriptive metadata fields. A taxonomy is a controlled vocabulary with a hierarchical structure. It was originally founded as the science of classification that originally referred only to the classifying of organisms. Now, it is regularly used in a more general setting, referring to the classification of things or concepts, as well as the schemas underlying such a classification. In addition, a taxonomy typically has some hierarchical relationships embedded in its classification.

A thesaurus may be understood as an extension of taxonomy because it takes taxonomy as defined above, allowing subjects to be organised in a hierarchy and in addition, it adds the possibility to allow other statements to be made about the subjects. A thesaurus contains hierarchical relationships but is not necessarily constructed as an overall hierarchy but other relationships as well. More precisely, an ISO-standard thesaurus has, in addition to BT/NT, Related Terms (RT) and scope notes. It's important to note that their hierarchical relations (BT/NT) may represent both the generic relationship (genus/species, or superclass/subclass) and the mereological relationship (part of). Also, they can be used for geographic inclusion relationships, and for thematic inclusion relationships which are not quite the same as superclass-subclass (e.g. in a library). More directly relevant is that the distinction between data structure standards, such as data models/schemas, and data content standards, such as thesauri etc., becomes tenuous the moment that one introduces a semantic approach.

For example, the Getty The Art & Architecture Thesaurus (AAT) is a controlled vocabulary used for describing items of art, architecture, and material culture used by, among others,

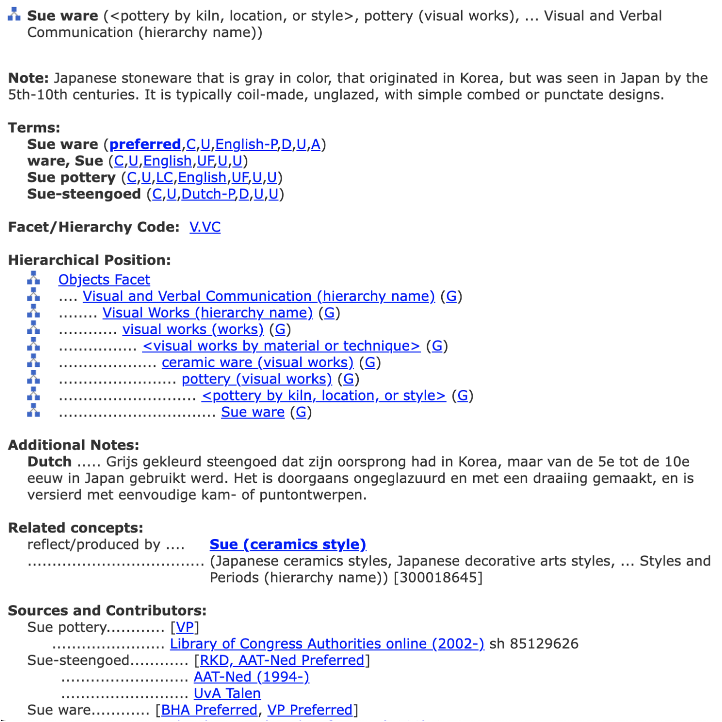
[47] http://dublincore.org/tools/ (last visited on 1 November 2021)

museums, art libraries, archives, catalogers, and researchers in art and art history. The AAT is a structured vocabulary of more than 55,600 concepts including 131,000 terms, descriptions, bibliographic citations, and other information relating to fine art, architecture, decorative arts, archival materials, and material culture.

The AAT is a faceted classification system as well as a hierarchical one and there are mainly seven facets (Whitehead, 1989):

- ***Associated Concepts*** consist of abstract concepts, such as beauty, balance, connoisseurship, metaphor, freedom, and socialism (Hierarchy: Associated concepts)
- ***Physical Attributes*** consist of perceptible or measurable characteristics such as size, shape, chemical properties, texture and hardness, such as strapwork, borders, round, waterlogged, and brittleness. (Hierarchies: Attributes and Properties, Conditions and Effects, Design Elements, Colour)
- ***Styles and Periods*** consist of stylistic groupings and distinct chronological periods, such as French, Louis XIV, T'ang Dynasty, Chippendale (Hierarchy: Styles and Periods)
- ***Agents*** consist of people, groups of people, and organisations such as printmakers, landscape architects, corporations, and religious orders. (Hierarchies: People, Organisations)
- ***Activities*** consist of areas of endeavour, physical and mental actions or methods, such as archaeology, engineering, analysing contests, exhibitions, running, drawing (image-making), and corrosion. (Hierarchies: Disciplines, Functions, Events, Physical and Mental Activities, Processes and Techniques)
- ***Materials*** consist of physical substances, such as iron, clay, adhesive, emulsifier, artificial ivory, millwork, and nylon. (Hierarchy: Materials)
- ***Objects*** consist of objects either fabricated or given form by human activity, such as paintings, amphorae, facades, cathedrals, Brewster Chairs, and gardens (Hierarchies: Object Groupings and Systems, Object Genres, Components; Built Environment: Settlements and Landscapes, Built Complexes and Districts, Single Built Works, Open Spaces and Site Elements; Furnishings and Equipment: Furnishings, Costume, Tools and Equipment, Weapons and Ammunition, Measuring Devices, Containers, Sound Devices, Recreational Artefacts, Transportation Vehicles; Visual and Verbal Communication: Visual Works, Exchange Media, Information Forms)

The AAT contains generic terms, such as "cathedral[48]", but no proper names, such as "Cathedral of Notre Dame." In this case, the hierarchical position can be Objects Facet Hierarchy followed by Built Environment and Single Built Works respectively.



Sue ware (<pottery by kiln, location, or style>, pottery (visual works), ... Visual and Verbal Communication (hierarchy name))

Note: Japanese stoneware that is gray in color, that originated in Korea, but was seen in Japan by the 5th-10th centuries. It is typically coil-made, unglazed, with simple combed or punctate designs.

Terms:
Sue ware (**preferred**,C,U,English-P,D,U,A)
ware, Sue (C,U,English,UF,U,U)
Sue pottery (C,U,LC,English,UF,U,U)
Sue-steengoed (C,U,Dutch-P,D,U,U)

Facet/Hierarchy Code: V.VC

Hierarchical Position:
Objects Facet
.... Visual and Verbal Communication (hierarchy name) (G)
.... Visual Works (hierarchy name) (G)
.... visual works (works) (G)
.... <visual works by material or technique> (G)
.... ceramic ware (visual works) (G)
.... pottery (visual works) (G)
.... <pottery by kiln, location, or style> (G)
.... Sue ware (G)

Additional Notes:
Dutch ..... Grijs gekleurd steengoed dat zijn oorsprong had in Korea, maar van de 5e tot de 10e eeuw in Japan gebruikt werd. Het is doorgaans ongeglazuurd en met een draaiing gemaakt, en is versierd met eenvoudige kam- of puntontwerpen.

Related concepts:
reflect/produced by .... **Sue (ceramics style)**
.... (Japanese ceramics styles, Japanese decorative arts styles, ... Styles and Periods (hierarchy name)) [300018645]

Sources and Contributors:
Sue pottery............ [VP]
.... Library of Congress Authorities online (2002-) sh 85129626
Sue-steengoed............ [RKD, AAT-Ned Preferred]
.... AAT-Ned (1994-)
.... UvA Talen
Sue ware............ [BHA Preferred, VP Preferred]

Figure 27: Getty AAT metadata for Sueware, a Japanese ceramic style. Source: https://www.getty.edu/vow/AATServlet?english=N&find=sue+ware&logic=AND&page=1&note=

For example, in Figure 27, there is a type of ceramic prominent in the Kofun period called “sue ware”[49] represented in Getty AAT, namely stoneware with natural ash glaze and incised decoration. Its hierarchical position is defined as Objects Facet followed by Visual and Verbal Communication and Visual works respectively. The entry is enriched by metadata that shows hierarchical position, related concepts and terms on the Getty AAT website. The additional notes and some explanations are also available in Dutch. Controlled vocabularies support various metadata fields and thus vocabulary design needs to be integrated with the metadata strategy that may require not just a single person, but rather a multidisciplinary team of experts to implement (Hedden, 2010).

48 http://www.getty.edu/vow/AATFullDisplay?find=cathedral&logic=AND&note=&english=N&prev_page=1&subjectid=300007501 (last visited on 13 November 2021)

49 http://www.getty.edu/vow/AATFullDisplay?find=sue+ware&logic=AND&note=&english=N&prev_page=1&subjectid=300387403 (last visited on 11 November 2021)

3.3.3 KOS, SKOS, RDF

In recent years, many applications, data formats, and query languages have been developed, which focus attention on the semantics of information. For example, the eXtensible Markup Language (XML), a standard format for data exchange, the Web Application Programming Interfaces (APIs), i.e. methods to request data from a web server in order to use them in external applications, and many standard protocols and architectures for the implementation of web services, specified by the World Wide Web Community (W3C).

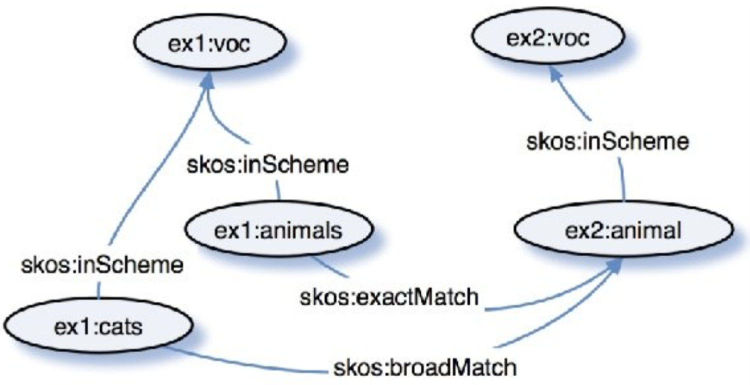


Figure 28: an example of the SKOS relationship between cats and animals.

Knowledge Organisation Systems (KOSs) and ontologies are mostly used for encoding knowledge. The distinction is that KOSs are less rigorous than ontologies, and no formal reasoning can be carried out by simply having KOSs. SKOS, which stands for Simple Knowledge Organisation System, is itself an OWL ontology and can be written out in RDF syntax to represent KOSs (Isaac and Summers, 2008). KOS searches for the concepts and individuals more robust. For example, Figure 28 shows a simplified representation of a cat as an animal using just SKOS. However, KOSs can not match up with ontologies when it comes to completely representing the knowledge in the solutions prompted by the Semantic Web paradigm. SKOS is one of the many Semantic Web Standards (also including e.g. OWL, RDF, SPARQL). It is used to port existing KOSs into the shared space of the Semantic Web; therefore a SKOS can be published on the Web and be machine-readable as well as exchanged between software applications. In basic SKOS, conceptual resources (concepts) can be identified with URIs, labelled with lexical strings in one or more natural languages, documented with various types of notes, semantically related to each other in informal hierarchies and association networks and aggregated into concept schemas.

An ontology formally defines a common set of terms that are used to describe and represent a domain. So an ontology is domain-specific and is used to describe and formally represent an area of knowledge. It contains the terms and the relationships between these terms. Relations can be taxonomical (subclass), membership (type) or properties. The latter describes various features and attributes of the classes and individuals. By having the terms and the relationships among these terms clearly defined, ontology encodes the knowledge of the domain in such a way that the knowledge can be machine-readable and processable.

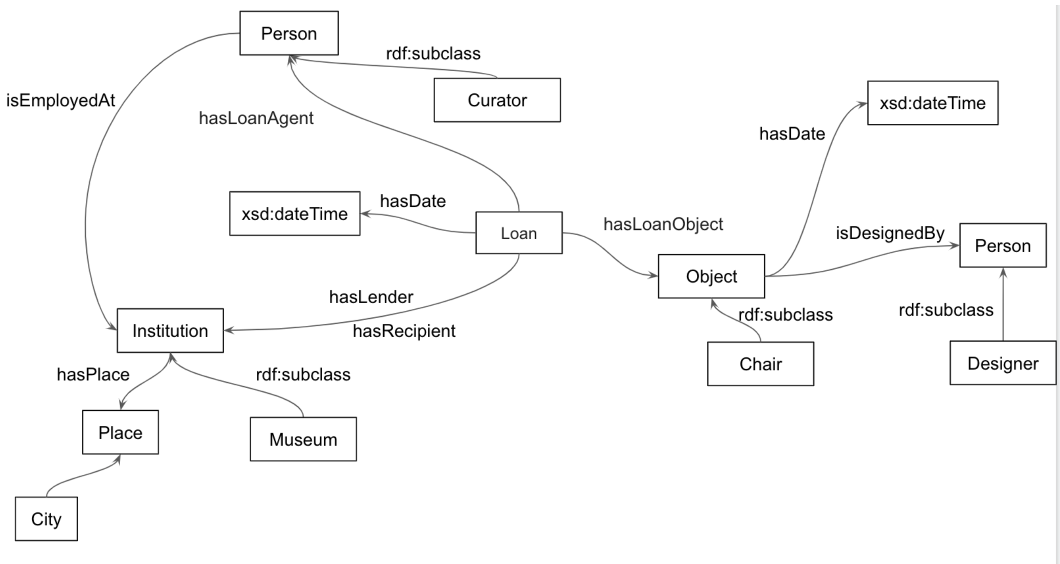


Figure 29: General knowledge representation of “A chair designed by designer Hans J. Wegner in 1955, is loaned from MoMA museum in New York to Design Museum London on the date 29.10.2020 by Paola Antonelli who works at MoMA as curator.”

For example, as seen in Figure 29, we have represented the general knowledge of the statement “A chair designed by designer Hans J. Wegner in 1955, is loaned from MoMA museum in New York to Design Museum London on the date 29.10.2020 by Paola Antonelli who works at MoMA as curator.” In this case, a loan (class Loan) concerns a loaned object (property *hasLoanObject*, range class Object), a loan agent (property *hasLoanAgent*, range class Person), a lender and a recipient institutions (properties *hasLender* and hasRecipient, range class Institution). It also occurs at some date (property *hasDate*, range *xsd:dateTime* type), LoanCH88, a chair designed by designer Hans J. Wegner in 1955, was loaned from

MoMA museum in New York to Design Museum London on the date 29.10.2020 by Paola Antonelli who works at MoMA as curator.

3.3.4 CIDOC-CRM

The conceptual Reference Model (CRM in short) is a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information (Crofts *et al*. 2009). The model was created by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) on empirical bases from real-world datasets. The latest version 7.2.1[50] (Bekiari *et al.*, 2022) consists of more than 85 classes and 140 properties, and it is currently being extended by users for more specific domains. The distinction between a temporal entity and a persistent item is a fundamental one for classical ontology as well as an important aspect of the CRM class hierarchy (as mentioned in section 3.3.2). It is a separation between temporal entities and persistent items which allows the modelling of historical events as well as the actors participating in related events.

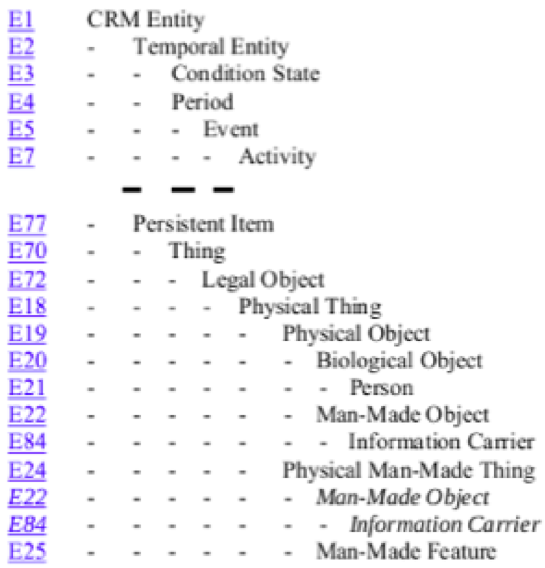

E1 CRM Entity
E2 - Temporal Entity
E3 - - Condition State
E4 - - Period
E5 - - - Event
E7 - - - - Activity
E77 - Persistent Item
E70 - - Thing
E72 - - - Legal Object
E18 - - - - Physical Thing
E19 - - - - - Physical Object
E20 - - - - - - Biological Object
E21 - - - - - - - Person
E22 - - - - - - Man-Made Object
E84 - - - - - - - Information Carrier
E24 - - - - - Physical Man-Made Thing
E22 - - - - - - *Man-Made Object*
E84 - - - - - - - *Information Carrier*
E25 - - - - - - Man-Made Feature

Figure 30: A partial CIDOC-CRM class hierarchy. Source: https://jyx.jyu.fi/bitstream/handle/123456789/51319/towardssemanticmodellingofculturalhistoricaldata.pdf?sequence=1&isAllowed=y

50 https://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_7.2.1%20%28updates%20from%207.2%20and%207.1.2%29_0.pdf (last visited on 7 March 2022)

Figure 30, above, shows a partial CIDOC-CRM class hierarchy. The most general class is *E2 Temporal Entity,* which is defined as being abstract and used as a superclass for all the classes that have a temporal component. In particular, the *E4 Period*, which ". . . comprises sets of coherent phenomena or cultural manifestations bounded in time and space" and the *E3* Condition State, which ". . . comprises the states of the objects characterised by a certain condition over a time-span" (Crofts *et al.*, 2009). *E3 Condition State* is not further specialised, while the *E4 Period* is the superclass of several classes (such as the E5 Event, which can be seen in Figure 30). Specialised classes are in turn characterised by specific properties, such as the ones of the *E4 Period* that make it possible to define spatio-temporal relations between different instances of *E4*.

CIDOC CRM is based on the event-centric approach which has several advantages compared to traditional approaches. First, events provide a semantically meaningful way of describing links between the physical things and the actions of humans (Doerr and Kritsotaki, 2006). It is intended to be a common language for domain experts and provides the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, galleries, libraries and archives (also referred to as GLAM in previous sections).

Also, the event-centric model allows a very flexible structure for an individual record. Events describe the history of an item, and new events can be added at any time which allows in practice, based on the decision of the person creating the documentation, the related documentation can be very detailed or just on a general level. Therefore, documentation can be constructed based on the specifications and the goal of the documentation rather than a rigid structure. It also simplifies the design of the data structure. In this context, it is possible to represent several types of creating a cultural heritage item, the construction of a building or having a scientific seminar. Because events are individual records, split into smaller units, the documentation is semantically more precise and more accessible. Last, and, from the perspective of documentation, a very important benefit is that using explicit events in documentation makes events themselves documentable as units. In the traditional item-centric approach this would need specific fields for every event type. To be able to define the cultural context of an item, it is beneficial if the cultural object can be separated from the physical carrier object or objects.

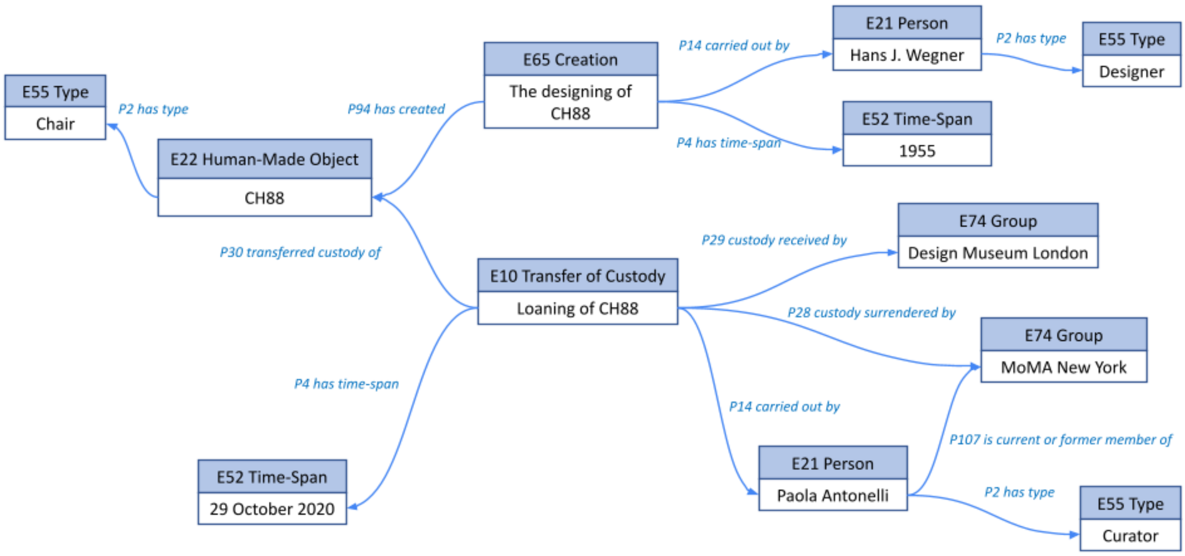


Figure 31: CIDOC-CRM relation of “CH88, a chair designed by designer Hans J. Wegner in 1955, is loaned from MoMA New York to Design Museum London on the date 29 October 2020 by Paola Antonelli who works at MoMA New York as curator.”

Figure 31 shows the CIDOC-CRM representation of ‘CH88 a chair designed by designer Hans J. Wegner in 1955, which is loaned from MoMA New York to Design Museum London on the date 29 October 2020 by Paola Antonelli who works at MoMA as curator.’ To simplify the process, there are some Classes and individuals boxed in white such as Human-made Object/Chair: CH88, Group: MoMA New York, Design Museum, Date: “29.10.2020”, Year: 1955, Person/Curator: Paola Antonelli, Person/Designer: Hans J. Wegner. There are object properties such as P14 carried out by who is responsible for that activity or event. We can represent the sentence through CIDOC-CRM (coloured in blues) relations as follows: CH88 has a type *(E55-type)* which was designed (*E52-timespan*) in 1955 by Hans J. Wegner who is an actor (*E39*) of the designing (E65 Creation) of CH88. Also, the Loaning of CH88 is represented as the transfer of custody (*E10*) to (P29) the London Design Museum whereas its custody is surrendered by (P28) MoMA New York. The museum is represented as E74 Group in CIDOC CRM. It is loaned from MoMA New York which is a group (*E74*) to another group, the Design Museum London (E74). The date (*E52*) of the loan is 29 October 2020 and carried out by (*P14)* curator Paola Antonelli (*E21*) who works at MoMA (*E53*).

3.3.5 Database management Tools

Proposed by E. F. Codd in 1970, the relational database is based on the relational model of data, maintained by a relational database management system (Codd, 1970). Many relational database systems have an option of using SQL (Simple Query Language), a native query language for relational databases. While SQL is a well-known and established query language, it is complex for novice users (non-IT people) and requires some knowledge about internal data structure. Another problem with SQL is that queries can become very complex with the recursive database structures. Now, there are several in which organisations can release data to the public for example electronic format or paper publications (PDF files based on aggregated data) and virtual exhibitions and websites of the collection of the archives of the cultural heritage organisations. The digital curation of data is extremely vital with what concerns the consequences of the interfaces for cultural heritage assets, specifically concerning recent technological changes.

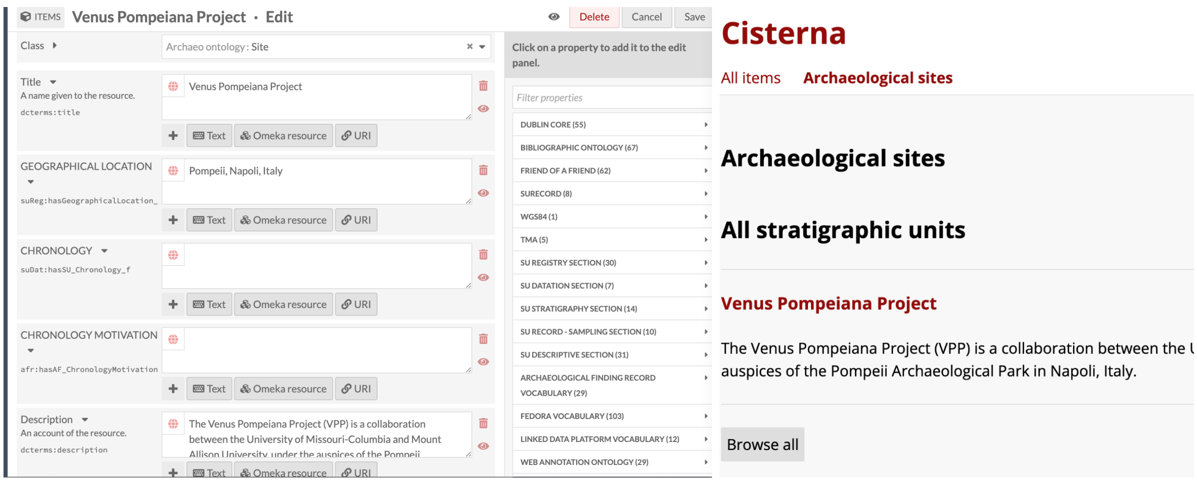


Figure 32: a) View of an example of an Omeka-built website's backend and b) frontend interface. Source: https://bearchaeo.di.unito.it/omeka-s/s/cisterna/page/welcome

A Content Management System (CMS) is software that allows for creating and managing websites. The core of CMS is a database where the structure and content of the sites are stored. Content Management Systems for Cultural Heritage are oriented to the creation of digital collections such as Arches, GlamKit, CollectiveAccess, and Omeka. For example, the Content Management System Omeka-S[51] offers several user interface metaphors which can be well suitable for presenting and interacting with rich media content. Omeka-S is a tool for

[51] https://omeka.org/s/ (last visited on 7 October 2021)

creating and publishing digital content accompanied by metadata and documentation which allows organising items into collections. A collection is an archive of material, related to event representation as we discussed in the previous section. As visible in Figure 32, the Venus Pompeiana Project is an exhibit, a selected set of materials from a collection, related by theme, topic, or other curatorial decisions. Using Omeka-S software, we created a website[52] for archiving and disseminating the Venus Pompeiana Project (also known as Cisterna which was mentioned in the previous chapter). In Figure 32a, there is the item that was created to represent the Venus Pompeiana Project digitally and Figure 32b shows the frontend interface of the website which was created by the Omeka-S web publishing platform. Participatory practices and tools changed the notion of archiving and practices in cultural heritage institutions due to the character of the audience involved and the established structures between and within institutions through linked open data, and other participatory practices. It opens up new opportunities and practices yet to be explored.

3.3.6 Digital Repositories

The archiving purpose of data and repositories are used to support their long-term value and mitigate digital obsolescence (Doerr, 2007). The repository includes a critical reflection on the digital materials and builds upon open and sustainable formats, semantic relations among artefacts and their constituent parts, in addition to aspects regarding authorisation, persistent identification, data curation and long-term archiving.

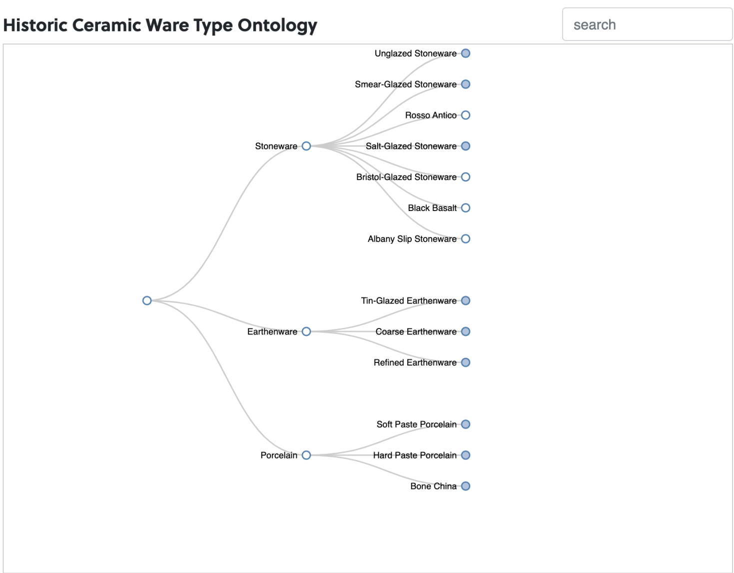


Figure 33: An excerpt tDar with a graphic representation of Historic Ceramic Ware Type Ontology. Source: https://core.tdar.org/

52 https://bearchaeo.di.unito.it/omeka-s/s/cisterna/page/welcome (last visited on 7 March 2022)

For example, as mentioned in previous chapters, the Digital Archaeological Record (tDAR) is an archaeology-specific repository that archives a wide variety of archaeological media. tDAR's use, development, and maintenance are governed by Digital Antiquity, an organisation dedicated to ensuring the long-term preservation of irreplaceable archaeological data in order to broaden access to these data. The goal of this endeavour has been to develop tools for synthetic and comparative research based on a novel, on-the-fly, ontology-based data integration to be deployed and tested in the context of the prototype infrastructure. In the semantic encoding of tDar, as seen in Figure 33, taxonomy is shown as ordinarily hierarchical (tree-like) that represents an arbitrary number of levels of class-subclass relationships[53]. It shows the result of the query function as the graphic representation of Historic Ceramic Ware Type Ontology. The tDar uses query-driven, ad-hoc data integration based on relevant taxonomies to conceal the semantic demands through a query from the semantic content as available datasets. Many perspectives can influence the creation of databases through ontologies. In particular, the personalisation aspect of the project is related to the collaboration of the agents such as data curator, curation professional, or end-user (including algorithms) in different granularities that includes taxonomies, data sets, data collection sheets, or interdisciplinary repository. Good digital preservation takes some thought in advance based on what to preserve, and to what extent: data (tables, images, other files), entire documents (reports, publications, theses), or displays (websites, interactive content). All of those kinds of information would require different concerns for preservation. Additionally, it is possible to investigate options to find an appropriate repository to tap into existing infrastructure and expertise. This repository might be a part of a university library ecosystem or something specifically designed for archaeological data, like Open Context[54] or tDAR, or it might be very broadly defined, like Zenodo[55]. The selection of the repository might dictate the kinds of metadata that are collected and also the way they are collected. Therefore, in the digital curation lifecycle, the sooner long-term storage and preservation options are considered, the better for the success of the project.

[53] https://slideplayer.com/slide/7914481/ (last visited on 13 January 2022)
[54] https://opencontext.org/ (last visited on 13 December 2021)
[55] https://zenodo.org/ (last visited on 13 January 2022)

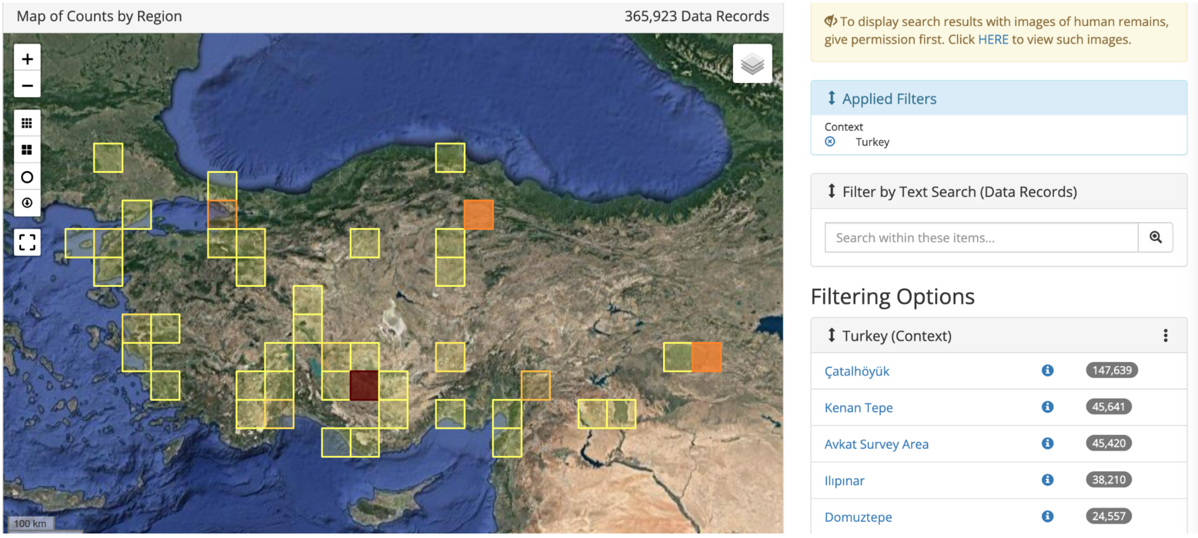


Figure34: A screenshot of the Open Context website that shows nearly 150,000 records of Çatalhöyük in Turkey. Source: https://opencontext.org/subjects-search/Turkey#6/39.096/33.398/9/any/Google-Satellite

Open Context is an open access data sharing system developed to support enhanced knowledge in archaeology and related disciplines. With an approach of "publishing by sharing" interlocking datasets on a common platform, Open Context has a highly abstracted and generalised global schema for representing archaeological data. In Figure 34, there is a screenshot of the Open Context website that shows nearly 150,000 records of the ancient site Çatalhöyük in modern Turkey. It is necessary to highlight that Open Context's data are published on aggregation in another scholarly information system, namely Pelagios[56], as "Linked Data". In order to contextualise data with rich metadata, Open Context regularly references Uniform Resource Identifiers for concepts in data stores such as PeriodO[57] (linking periods to define chronology) and the Getty Art and Architecture Thesaurus (Kansa *et al.*, 2020). For example, using PeriodO as seen in Figure 34, we can make searches with particular modern geographic boundaries ("Turkey") or ancient or modern sites ("Çatalhöyük"). Although the current internal data model of Open Context largely looks like a graph-database structure commonly used for RDF triple-stores, in fact, it mainly emphasises RDF and Linked Open Data to relate the data it publishes with the data curated by external sources. It is crucial to apprehend that the information is archived safe and sound in an appropriate digital repository with the purpose that the data are findable, referenced, interoperable or available to re-use by other researchers (FAIR principles mentioned above). Since no "best" repository choice fits all the digital humanities data projects, it is necessary

[56] https://pelagios.org/ (last visited on 18 January 2022)
[57] https://perio.do/en/

to understand the strengths and disadvantages according to the needs and specifications of each project.

3.4 Semantic interoperability for archaeological projects

Nowadays, digital archaeology is evolving fast while the archaeological objects call for an ontological approach to the representation of the archaeological domain, such as the CIDOC Conceptual Reference Model (CIDOC CRM) as described in previous sections. However, as Leif Isaksen (2008) observes that there is still a certain confusion and some resistance (Niccolucci *et al.*, 2009) inside the "digital archaeology" community about the Semantic Web. Based on his experience in the process of integrating different datasets from different sources, Isaksen questioned to what extent is contemporary archaeological practice able and willing to meet the social and technical requirements of expressing data with semantic technologies concerning the necessity and potential benefits of the Semantic Web technologies (Isaksen, 2008). This is certainly true if we consider the broad cultural heritage community that often showed a superficial approach to the deep understanding of the nature and limits of new technologies and approaches mentioned previously in this chapter; however, there have been some successful initiatives that show how the proper set of Semantic Web ideas and technologies is permeating the field of archaeology (Mantegari, 2010).

Some projects successfully integrate semantic tools into their research infrastructure such as the PARTHENOS[58] project which is a research infrastructure whose objective is to strengthen the cohesion of research across several related fields associated with the digital humanities. The project includes linguistic studies, cultural heritage, history and archaeology and is built on existing research structures such as CLARIN[59] (languages) and DARIAH[60] (arts and humanities). However, in this research, we will focus on ongoing and ever-evolving research projects specifically in the realm of archaeological data. For example, current work in the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (as mentioned before, namely ARIADNE[61]) provides an event-centric ontological representation of archaeological excavation namely CRMarchaeo, a lower ontology extension of CIDOC CRM and the CRMdig digital provenance mode as well as RDF Schema for encoding metadata about the steps and methods of production (also called

[58] https://www.parthenos-project.eu/ (last visited on 3 January 2022)
[59] https://www.clarin.eu/ (last visited on 8 January 2022)
[60] https://www.dariah.eu/ (last visited on 11 January 2022)
[61] https://ariadne-infrastructure.eu/partners/ (last visited on 13 January 2022)

provenance) of digital products and virtual representations such as 2D, 3D or animated models.

Considering the complexity of building semantic infrastructure and services for a specific domain such as archaeology, cooperation between domain data creators and curators, aggregators and service providers is crucial. In order to establish interoperability, collaboration is for sharing datasets through a domain portal (i.e. the ARIADNE data portal[62]) as well as using common or aligned vocabularies (e.g. ontologies, thesauri) for describing the archaeological data. As an extension to the preceding ARIADNE Integrating Activity by including archaeological statistics infrastructures in Europe, the ARIADNEplus undertakes and includes archaeological statistics infrastructures by indexing in its registry approximately 2.000.000 datasets (ARIADNE portal). Constructed on the results of the ARIADNE project, ARIADNEplus extends and supports the research community by creating similarly growing relationships with key stakeholders along with the most critical European archaeological associations, researchers, heritage professionals, national heritage organisations and so on. It includes leaders in specific archaeological domain names like palaeoanthropology, bioarchaeology and environmental archaeology in addition to different sectors of archaeological sciences.

There are other examples of multi-year, multilingual, multidisciplinary archaeological projects such as the STAR[63] (Semantic Technologies for Archaeological Resources) project, which was undertaken by English Heritage, the University of Glamorgan (Wales) and the Royal School of Library and Information Technology (Denmark). The general aim of the STAR project was the application of Semantic Web technologies for the integration and exploitation of digital archive databases, vocabularies and the associated grey literature (i.e. fieldwork reports, building surveys etc.) in archaeology. One of the important outcomes of the STAR program has been the creation of an extension of the CIDOC CRM to describe the archaeological datasets, with particular respect to the archaeological excavation. The extension, namely “CRM-EH”, provides 125 sub-classes and 4 properties, created in 2004 by the English Heritage Research and Standards Group in collaboration with CIDOC (Cripps *et al.*, 2004). For example, the central entity of the excavation methodology is defined as ‘context’ in CRM-EH (CRM-EH 2014). From the user experience perspective, the

62 http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_archaeological_LOD_study_10-2016-1.pdf (last visited on 3 January 2022)

63 http://www.methodsnetwork.ac.uk/resources/casestudy13.html (last visited on 25 January 2022)

development of the CRM-EH mapping/extraction tool has been considered essential within the context of the project to assist users in the process of data mapping, cleansing and extraction (Binding *et al.*, 2008) to allow the mapping of RDF entities to database columns, constructing structured SQL queries, cleansing data, and outputting the result to RDF files. In addition, the prototype search/browse application allows users to cross-search and explore integrated data extracted from the previously separated databases.

The creation of the CRM-EH extension was an important step forward for archaeologists, mostly because it broadens the scope of the CIDOC CRM beyond the community of cultural heritage institutions or museums (Isaksen, 2008). On the other hand, the evaluation of the tools developed in the context of the STAR project was restricted to the available demonstrators that provided limited capabilities. Considering completed projects such as STAR, a full release of the data sources would possibly provide very interesting contributions in the field of Semantic Web applications for archaeological research.

Conclusion of the Chapter

In conclusion, these projects ultimately support an abstract and conceptual model of digital data curation for archaeology with semantic encoding. They employ a user-centred approach that can accommodate the needs of various users and refine the description and expression of digital data accordingly. Third and last, they suggest extensions of the existing archaeological knowledge that should be considered important to produce and maintain descriptions and expressions of resources when diverse research areas or institutions construct and develop long-time preservation of the archaeological data. It is necessary to outline the principal characteristics of these projects that are considered relevant to discuss the current situation of digital archaeology concerning digital curation. The Digital Curation field could also take into account the evolving Semantic Web, which should be considered as data-centred processing of knowledge and information management by merging the web of human-readable documents with the web of machine-readable data (Lee *et al.* 2001; Shadbolt *et al.* 2006).

In the following chapters, we will use these insights and approaches in digital archaeology, alongside the literature review and abstracted model which we examined in the first chapters, to construct an extended digital data curation model to implement and test through a multidisciplinary, multilingual, multinational archaeological research project named BeArchaeo.

Chapter 4

Digital Data Curation Model for Archaeology

In this chapter, we will elaborate on the conceptualisation aspect of the digital data curation model in order to accommodate the needs of the various users and refine the description and expression of digital twins, in the field of archaeology. We will redesign the digital data curation workflow according to findings from previous chapters. We will exemplify the digital data curation model with the archaeological finding SH1 from a digital-born project called BeArchaeo. BeArchaeo is a suitable project because we want to reflect on this new challenge in the context of Archaeology and Archaeometry since archaeologists and archaeometrists (e.g., chemists, physicists, ...) with a truly transdisciplinary vision. In particular, we will discuss the digital data curation applied to the integration of the archaeological investigation with archaeometry. Archaeometry involves the development and application of natural scientific methods and concepts to the solution by measuring and evaluating facts and interpretations. Therefore this chapter also discusses the disciplines that may contribute to archaeology (e.g., chemistry, biology, physics, and geological sciences). The contribution of the results from archaeometric investigations can reveal hidden properties of the domain.

4.1 Digital Data Curation Model for Archaeology

Nowadays, the journey of archaeological data starts not only at the excavation site but possibly also at surveys, or even in objects within legacy collections. Later follows through the museum's curatorial practices, accompanied by labels in exhibitions and records in digital repositories and archives. While archaeological investigations have been relying more and more on reflexive methodologies during the excavation and producing data by using web-based interfaces, up to filling the database entries for the excavation, the contribution of archaeometry as a practice is also acknowledged by many archaeologists as an essential and integral part of these archaeological investigations. It involves the development and application of natural scientific methods and concepts to the solution of cultural-historical questions. Since the applications of natural sciences have impacted archaeology since the quantitative analysis of Roman coins in 1799 by Martin Heinrich Klaproth, it is evident that archaeometry is similar to archaeology by its ultimate aim while it is considered also natural science by approach (Lombardo *et al.*, 2020). It includes all the disciplines that may contribute to archaeology (e.g., physics, chemistry, biological sciences, anthropology, geological sciences), by measuring and evaluating facts and interpretations (Artioli, 2010) Although there is some confusion among diverse practitioners (Post, 2019), we have realised that to enrich the collaboration between interdisciplinary specialists, digital curation in the field of archaeology needs to include the conceptualisation phase into the lifecycle of the projects as soon as possible (Dallas *et al.*, 2009).

Therefore, in traversing our schema (Figure 4 in chapter 2.3), considering knowledge representation and semantic tools, it is important to conceptualise the entire process, decide about the tasks, define the roles and responsibilities and choose related digital tools and software applications for collaborative work. We have added "Conceptualisation" into the Digital Data Curation model for archaeology in order to accommodate the needs of various users and refine the description and expression of digital twins, in the field of archaeology. Because without regular, sustained processing of the data and metadata of the cultural heritage assets, digital curation activities are only put into practice intermittently which eventually creates a gap between practices and multiple actors.

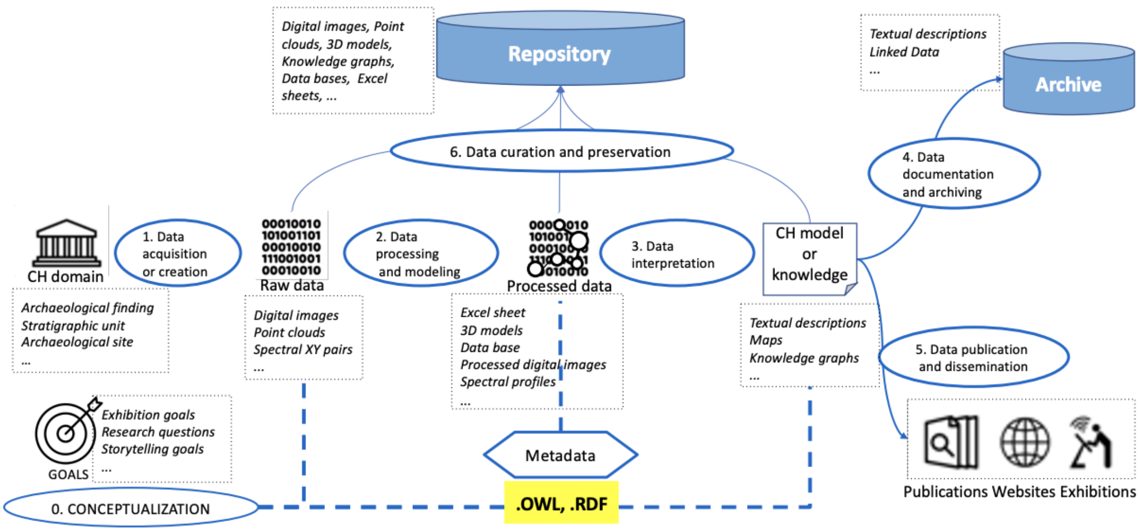


Figure 35: Digital Data Curation model for Archaeology

The conceptualization phase (numbered 0 in Figure 35) provides a knowledge management framework to define the operational model for the digital data that is produced during the ongoing project implementation. The conceptualization depends on both the cultural heritage domain (an archaeological site, a specific finding, …) and the goal intended to be achieved by the project (the chronology of the finding, the material provenance, …) which is encoded in an ontology expressed through the Semantic Web languages (e.g., OWL, RDF). It provides the backbone for the database schema design that will account for the description and encoding of the digital data produced throughout the project.

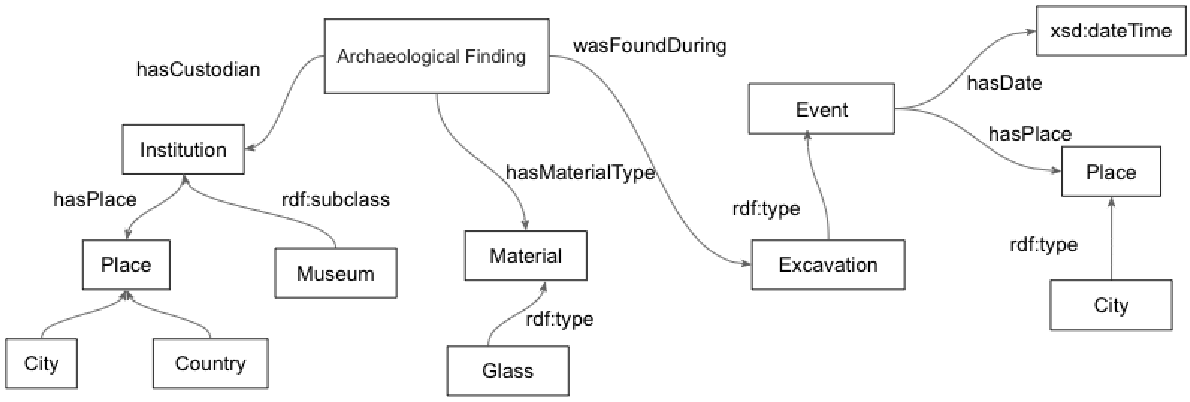


Figure 36: General knowledge representation of the “AF SH1" is an archaeological finding made of glass that was found during the Tobiotsuka Kofun excavation in Okayama, Japan.

For example, Figure 36 shows the general knowledge representation/conceptualisation of “AF SH1 is an archaeological finding made of glass which was found during Tobiotsuka Kofun excavation in Okayama, Japan.” We can examine it as AF SH1 is an archaeological finding which was found during the excavation. AF SH1 has a material type which is glass. Excavation has a location in Okayama, Japan. Okayama University has custody of AF SH1. Shimane Museum of Ancient Izumo is a museum. Okayama University lends AF SH1 to the Shimane Museum of Ancient Izumo for an exhibition.

Now we illustrate the Digital Data Curation model with the example from an ongoing archaeological project. The digital data was generated from an archaeological finding, recorded during an archaeological excavation and analysed afterwards. It is a forward-looking perspective, from the acquisition of the digital data after the discovery of a finding to interpretation and storage for future use.

4.2 BeArchaeo as a DDC-born archaeological project

Beyond Archaeology (BeArchaeo[64]) is a RISE European project that consists of the archaeological excavation, archaeometric analysis, interpretation of the findings, and eventually dissemination of the results about the Tobiotsuka Kofun (Soja city in Okayama Prefecture), together with other Kofun burial mounds and the related archaeological material in ancient Kibi and Izumo areas (present Okayama and Shimane Prefectures), in Japan. The project activities and outcomes will be accessible to the general public through engaging media communication along with the project development and two final exhibitions in Italy and Japan.

Together with other Kofun burial mounds and the related archaeological material in ancient Kibi and Izumo areas, scientists aim to work on a transdisciplinary vision of archaeology combined with archaeometry; the project activities and outcomes are accessible to the general public through engaging media communication along with the project development. In the next section, we will examine the overall Digital Data Curation process of the project through an archaeological finding, namely SH1, from the perspective of a methodological project. The BeArchaeo project has a transdisciplinary approach to archaeology and archaeometric disciplines and is interlinked through a semantic model of processes and objects.

4.3 Digital Data Curation for Archaeological Finding SH1

The BeArchaeo project carries out archaeological excavation and the related archaeometric analyses of the Tobiotsuka Kofun as well as other Kofun burial mounds and the related archaeological material in ancient Kibi and Izumo areas. The researchers aim to develop a transdisciplinary vision in studying the archaeological site and other archaeological materials now stored in museums and laboratories, in Japan. The conceptualization of the knowledge in the BeArchaeo project is driven by the design principle of recording the archaeological/archaeometric activities and the collected data that occur both on the archaeological site and in the lab. The data are recorded in a database filled by the scientists to be employed in interpretation processes and exhibition organisation.

[64] https://www.bearchaeo.com/be-archaeo-project (last visited on 4 January 2022)

Figure 37: An archaeological finding(SH1) is a fragment of proto-historical pottery that has been recovered by the archaeological excavation.

Archaeological finding SH1 (Figure 37) is a fragment found in a trench of an archaeological site during the BeArchaeo DDC-born archaeological project. The digital curation process starts as soon as the fragment is found. The goal of digital data curation is to support the scientific research on the composition of the findings and to examine their relationship with the question of their similarities and differences. In this specific example, the research question is to find the provenance of a set of similar potteries through a comparison of the component materials, including elemental composition, morphological features, presence, typology and composition of inclusions such as minerals or rock fragments.

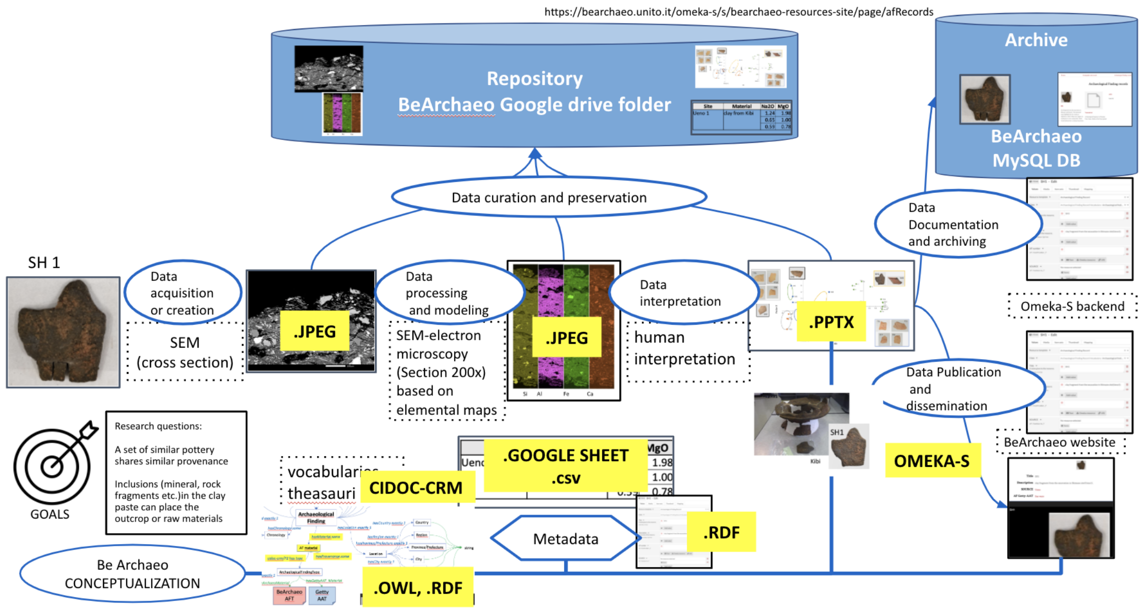


Figure 38: An archaeological finding(SH1) is a fragment of proto-historical pottery that has been recovered by the archaeological excavation.

Figure 38 instantiates the Digital Data Curation model addressing the digital data originated since the discovery of the archaeological finding named SH1, undergoing a specific investigation path, at the current stage of development. As we have seen above, interpretations are recorded in some digital format and then revised or updated, also encoding other formats, going formally when possible. The scientists carried out later in the lab, after mounting a subsample as a cross-section where scientists analyze the structure of the fragment through a process of data acquisition. Images of the cross-section of the fragment are analysed utilising an SEM-BSE (Scanning Electron Microscope backscattered electrons) to analyse the structure of the fragment through data acquisition.

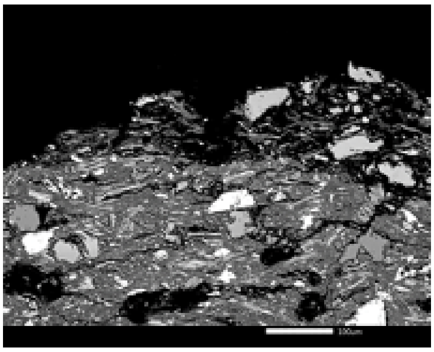
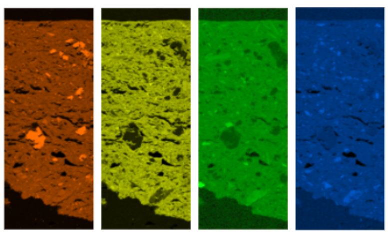


Figure 39: BeArchaeo raw data: electron microscopy image of the archaeological finding.

This process generates raw data (a magnification is shown in Figure 39, jpeg file format). Then scientists can decide which additional procedures they will conduct to understand the chemical components of the fragment. The task of data modelling and processing enriches raw data with metadata that reveals an interpretation of the asset at some level (e.g., measure the size of the fragment or assign the label "clay").



Site	Material	Na2O	MgO	Al2O3	SiO2	P2O5	SO3	K2O	CaO	TiO2	MnO	Fe2O3
Ueno 1	clay from Kibi	1.24	1.98	27.82	48.61	0.97	0.43	0.96	2.65	2.41	n.d.	13.60
		0.65	1.00	37.45	47.59	1.40	0.84	0.26	1.21	1.26	n.d.	9.21
		0.59	0.78	39.00	47.94	1.45	0.82	0.24	0.83	0.74	n.d.	8.41

a b

Figure 40: a) Elemental maps of a portion of the sample after Scanning Electron Microscope/Energy Dispersive Using X-Ray (SEM-EDX Analysis);
b) Excel sheet about the result of chemical components metadata of archaeological finding

Elemental maps of a portion of the sample, which are visible in Figure 40a, highlight that the coating is depleted in Al2O3; later, it may suggest an enrichment in iron compounds, which would indicate that a coating was present. Such information derives from the combination of different scientific tests and different expertise. The outcome of this step, namely the processed data, is the numeric values (as seen in Figure 40b) concerning the elemental composition determined after a set of analyses, enriched with metadata. For example, according to scientists, it is evident that the surface is made of a flaking brown layer (and a conservation treatment with Binder 17 - acrylic acid ester copolymer has been applied in the past), which may represent the residual of a surface coating. Also, elemental maps of a portion of the sample are added as metadata (actually, elemental maps have been acquired before, within the process of Energy Dispersive X-Ray Analysis (EDX) together with results from SEM analysis, but logically belong to this phase). The task of data modelling and processing enriches raw data with metadata that reveals a feature of the asset at some level (e.g., the possible presence of a surface coating).

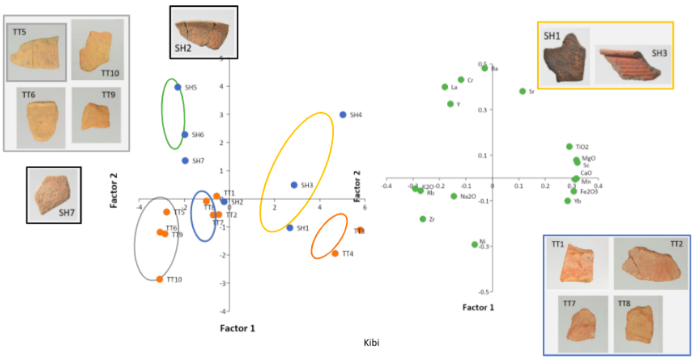


Figure 41: Data interpretation results of chemical analysis clusters show similar and different components.

Together with metadata-enriched processed data, these data undergo the data interpretation process, becoming a proper model of the cultural heritage domain (as seen in Figure 41). Based on the findings, the resulting statement is that the layer is formed by very thin and purified clay-bearing oxidised iron compounds. In general, such statements may include the scientific context underlying the data as well as the additional questions and interpretations. As the morphology of the fragment would arise more questions, some further samples of the same typology in a better conservation state would clarify this point.

The interpretation shows that this fragment may share the same provenance with other fragments, because of the cluster that appears after data elaboration, which tentatively sets it together with other fragments. We can therefore assume that the fragments included in the same cluster might have been produced in the same area. So, the current conclusion is that SH1 can originate in the Kibi area (interpretations reported through PowerPoint slides). The intermediate and the final data are stored in the repository, namely, a Google Drive shared folder, through the tasks of Data curation and preservation.

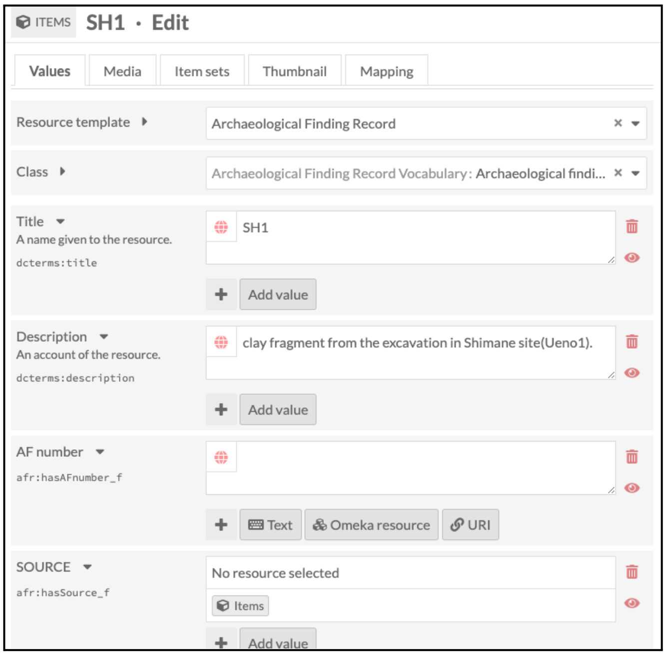


Figure 42: Screenshot from the back-end BeArchaeo resources website, concerning the Archaeological Finding SH1.

Moreover, the interpretations in the format of PowerPoint slides are also selected and stored, as part of the Data documentation and archiving task, into the BeArchaeo Archive, namely a MySQL database, underlying an Omeka-S installation, which also works as a centralised database for the coordination of digital data curation. The model will also be enriched with further metadata (e.g., the digital image also receives the identifier of the physical fragment). The database schema design and the organisation of the Google Drive folders are based on the proposed semantic model worked out after the conceptualization phase, to ease the problems of interoperability and connection between the archaeological and the archaeometric data.

On the other hand, the model and the relative knowledge about the domain at hand, can be used in the archives as a part of the data documentation and archiving (e.g., the digital image also receives the identifier of the physical fragment). In order to make the knowledge available to the archaeologists on the field, there is a BeArchaeo project website, based on the same Content Management System (CMS) called Omeka-S[65]. The knowledge construction and the registration of the archaeological findings and forms as templates are made possible through this web-publishing platform. This open-sourced tool has particularly served well for the import of semantic properties defined in an RDF file, the definition of customised vocabularies, and the construction of templates for the instantiation of filling forms (as seen in Figure 42).

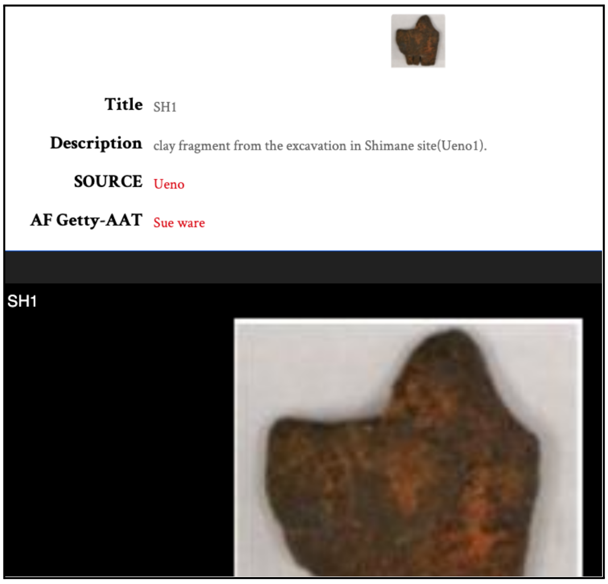


Figure 43: Screenshot from the BeArchaeo resources website, concerning the Archaeological Finding SH1, with the related media.

Related to these concerns and potential interpretations, the database design of the BeArchaeo project provides the information structure for all the digital curation phases of the project. In this case, it provides a repository while creating the archive of the archaeological findings with the related media. Media and metadata are stored in the BeArchaeo database as Archaeological Finding form, interfaced by Omeka-S CMS (Collection Management System) web platform in order to support the archaeologist's work in recording the excavation and interpretation activities. The implementation of MySQL, an open-source relational database management system, allows users to easily reach the digital image as it is a part of a virtual collection exhibited through the website (as seen in Figure 43). Considering it is a publication and dissemination tool as well, elements in red are links to other elements of the documentation (e.g., Ueno) or some external knowledge source (e.g., Getty AAT thesaurus).

65 https://bearchaeo.unito.it/omeka-s (last visited on 15 January 2022)

According to our knowledge, the holistic representation of the archaeometric processes with the archaeological investigations at large has not yet found its way through the Semantic Web endeavours. As Bearchaeo presents a conceptual model and ontology for supporting this transdisciplinary conception of archaeological investigations, we examine the contribution of archaeometry to the archaeologists' reflexive methodology in the context of an encompassing digital curation of the archaeological data.

Conclusion of the Chapter

In conclusion, the contribution of the BeArchaeo project is in the phase of dissemination which also will allow tracking of the archaeological findings' entire lifecycle through various institutions and collections. The archaeological finding can be a part of the Stratigraphic Unit, museum collection or place as well as the exhibition. The ontology also allows the use of rich media materials in different dissemination and interpretation techniques. The data publication of the ontology and the database is also employed in the definition of the contents of the exhibition that will present the outcomes of the BeArchaeo project to large audiences. In the next section, we will examine further the design of the BeArchaeo database concerning the Digital Data Curation model.

Chapter 5

Ontology modelling for BeArchaeo and Preliminary Evaluation

In this chapter, we show how the archaeological activities carried out in the BeArchaeo project have been supported by semantically encoded digital curation. We will describe the conceptualization of the archaeological, archaeometric, and cataloguing knowledge, which leads to the design of the database schema and its employment in a content management system for the monitoring of the activities of the BeArchaeo project. The conceptualization has led to the development of the BeArchaeo ontology, which comprises three modules, namely archaeological knowledge, archaeometric knowledge, and catalogue record knowledge. We address the major decisions for the ontology modelling process and then we provide an overview of the classes and properties of the BeArchaeo ontology. We will examine the implementation of the backend and the frontend of the project repository through CMS Omeka-S and the guidelines and recommendations for the usage of the platform when annotating the metadata through catalogue records that are inspired by the Italian Ministry of Culture Italian (ICCD) Standards.

5.1 Conceptualization for BeArchaeo

The conceptualization of the BeArchaeo project is encoded in some Semantic Web language (e.g., OWL, RDF) and provides the backbone for the database schema design that will account for the description and encoding of the digital data produced by the project. The BeArchaeo ontology, presented here, addresses the archaeological knowledge, the archaeometric knowledge, and the design of the forms to be filled during the archaeological/archaeometric endeavour.

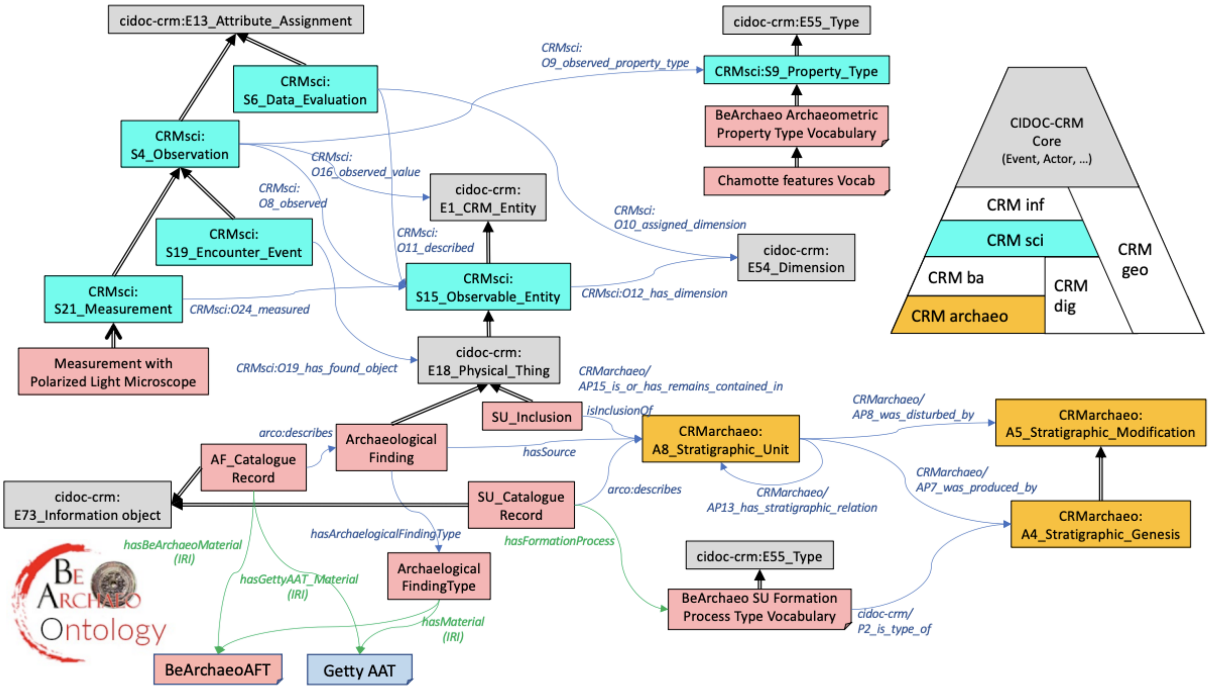


Figure44: Major relationships between BeArchaeo and CIDOC-CRM family. Colours are employed to distinguish the ontological modules

In BeArchaeo ontology, Figure 44 represents the conceptual model of the knowledge domain independently from the technologies used for organising and storing the information content itself. It shows major relationships between BeArchaeo and the CIDOC-CRM family. Colours are employed to distinguish the ontological modules. The main advantage of this representation is that it enables the domain experts to highlight the proper historical and cultural meanings of cultural data under which prospective data is to be disseminated. Once this conceptual ontology is defined, it is necessary to translate it into a machine-readable format. This bottom-level machine-readable ontology is expressed in a language that is processable by a computer system. Creating a new ontology from scratch is a time-consuming task and, therefore, it is better to exploit a general ontology from which customised cultural ontologies can be derived. In our case, to develop such an ontology we leverage the CIDOC Conceptual Reference Model (CIDOC-CRM). From a technical point of view, the model has been described as a number of sub ontologies concerning the stratigraphic unit (the correspondent CRMarchaeo class) and the archaeological finding record (subclass of *CIDOC-CRM/E18* Physical Thing in Figure 43), respectively, and the archaeometric processes as subclasses of the CRMsci reference model with the ontology expressed in OWL/RDF formats. The machine-processable ontology of BeArchaeo is then mapped onto the underlying data repositories which is a relational database for the BeArchaeo project.

5.2 Database structure

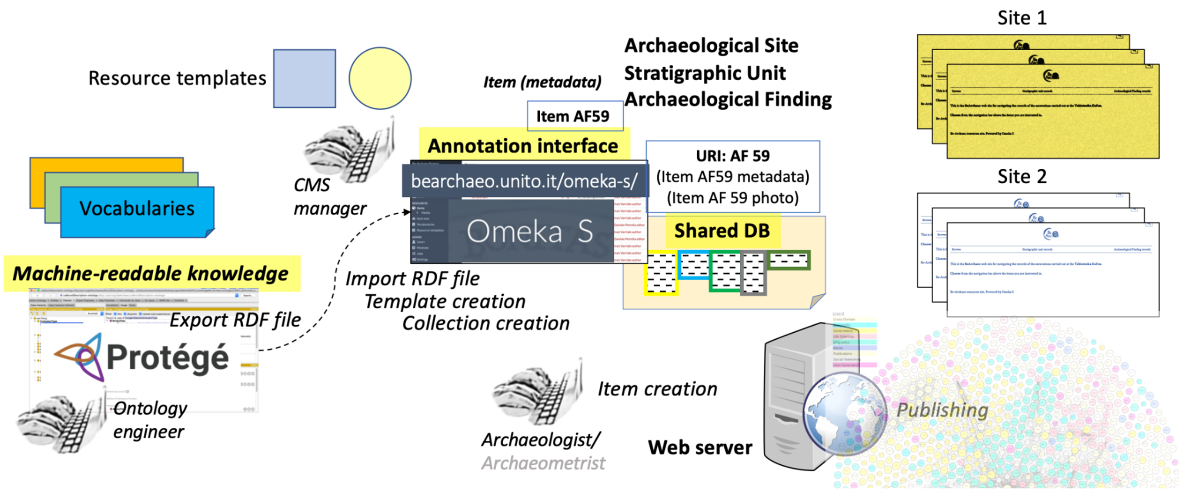


Figure 45: Visual representation of the database structure of BeArchaeo.

Figure 45 shows the visual representation of the database structure of BeArchaeo. We used the Protégé[66], which is an open-source ontology editor developed by Stanford Center for Biomedical Informatics Research. Protégé supports various ontology formats: OWL, RDFS, XML Schema, etc. The database structure is created to share data among all researchers to avoid isolation of the various databases and spreadsheets that lived on the team members' personal computers that were migrated to a centralised system. With the centralization of the data, the need arose for standardised vocabularies to be used by all teams. For this purpose, data from a relational database transforms into graph data described by an RDFS/OWL ontology, allowing the representation, retrieval and traversal of data according to semantic relationships. The conceptualisation part of the digital data curation model allows the digital repository as a "meta"-knowledge base which can be inspected concerning the interpretative values hidden in the data structures. The database schema design and the organisation of the google drive folders are based on the proposed semantic model worked out after the conceptualization phase, to ease the problems of interoperability and connection between the archaeological and the archaeometric data.

[66] http://protege.stanford.edu/ (last visited on 7 January 2022)

The complexity of the CIDOC-CRM is hidden from the end-user, and the domain-specific documentation structure is defined by semantic documentation templates that map parts of thesauri to a partial domain ontology which is built on top of CIDOC-CRM (Häyrinen, 2008). Although the documentation template does not define any user interface elements a user interface can be freely created for the template concerned. Therefore we introduce the record templates constructed from the survey forms in the following section.

5.3 Construction of the vocabularies for the survey forms

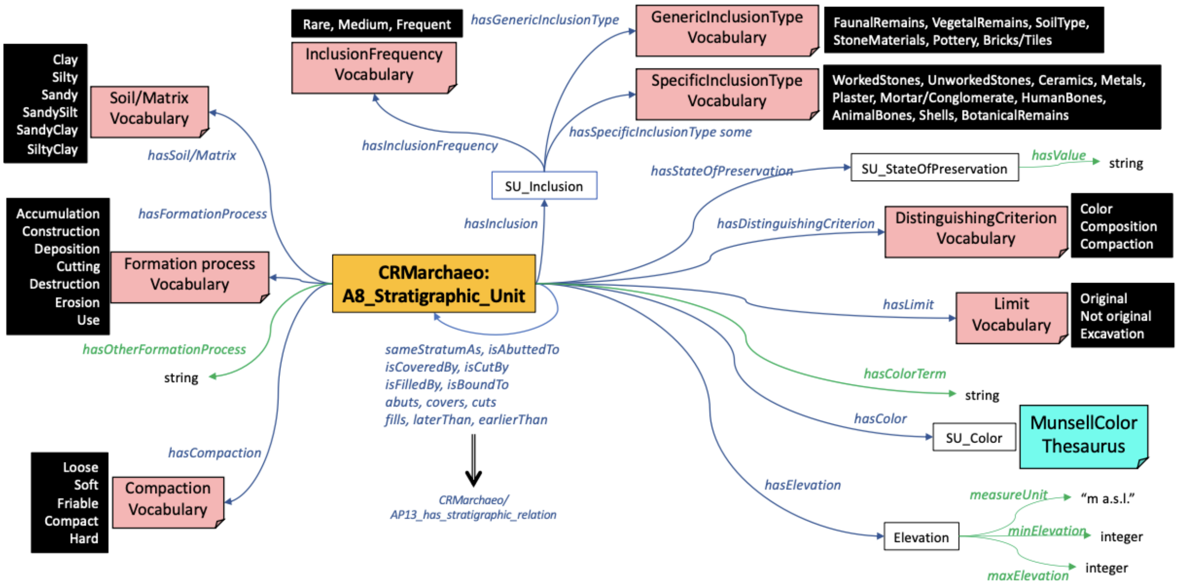


Figure 46: Modeling of the stratigraphic knowledge including references to thesauri and vocabularies (with the list of terms).

The goal of the Digital Data Curation for Bearchaeo is to support the effective exchange of knowledge and interoperability and keep a consistent global database including the several collections involved. In Figure 46, we address the overall ontological approach as well as the modelling of the stratigraphic knowledge considering a Stratigraphic Unit. Going clockwise, a stratigraphic unit has inclusions (i.e., entities that are contained in stratum), which are of some type, that can be generic or specific, and has a frequency of occurrence in the unit, qualitatively valued as rare, medium, or frequent. Inclusions have types that are taken from partially overlapping vocabularies, based on the practical experience of the archaeologists (these may change and should be aligned with the types included in the thesauri for the archaeological findings). Some informal properties noted as free text, are the state of preservation of the unit and the measurements taken during the excavation, with a particular concern for Elevation. The distinguishing criterion determines how this unit has been identified: the terms that concern this attribute are three (Colour, Composition and Compaction) and the other three properties possibly specify the actual values for such attributes (namely 6-valued soil/matrix term for composition, the 5-valued term for compaction, and a free string for colour). Colour, in the relationship with archaeometrists (specifically, the soil scientists) has been augmented with the encoding provided by the well-known Munsell colour system, in use in pedological studies. Finally, the formation process concerns a specialisation of the processes that are responsible for the creation and modification of the unit, with a frequent term vocabulary, which can be further augmented with free text insertion. The properties in the centre of the figure specify the stratigraphic relations, in spatial and temporal terms.

5.4 Implementation of forms and repository through Omeka-S platform

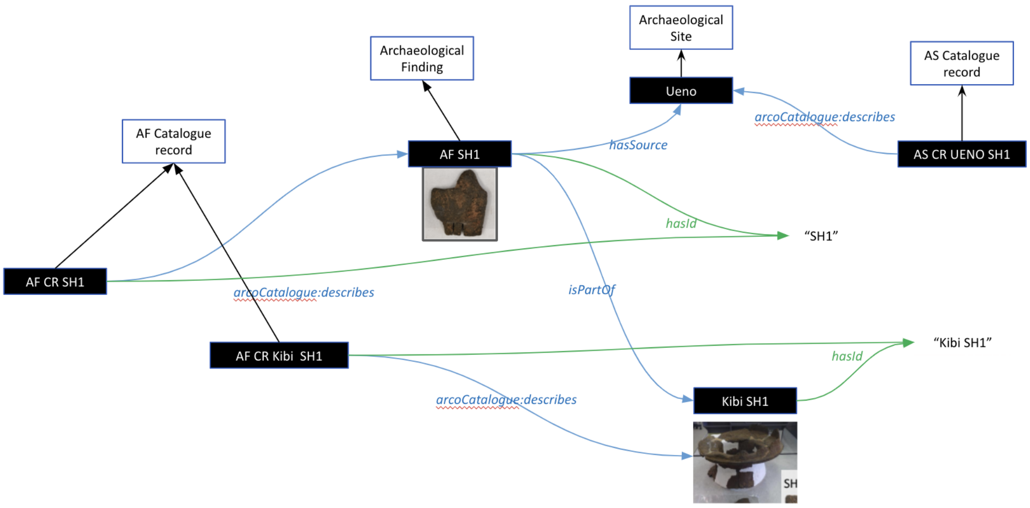


Figure 47: A diagram to show the conceptual relation between the archaeological finding(SH1), and the archaeological site where it was found in.

As seen in Figure 47, first we create the entry for Archaeological finding SH1 in the relational database (Omeka-s site). The Archaeological Finding itself is described by, based on Knowledge Graph for the Italian cultural heritage ArCo[67], (*arcoCatalogue:describes*) an Archaeological Finding Catalogue Record, filled through the forms in the website back end. The items (black boxes) belong to classes (blue-bordered rectangles). Blue curved arrows represent relations between items (called properties in Omeka-s, that point to other items) and green arrows represent values in the forms (also properties in Omeka-s, but pointing to

[67] http://wit.istc.cnr.it/arco/?lang=en (last visited on 28 November 2021)

string or numeric values). The reference model for cataloguing Italian archaeological heritage was published in 1984 by the Istituto Centrale per il Catalogo e la Documentazione, an institute of the Ministry of Cultural Heritage. It refers to a limited number of forms: Stratigraphic Trench (SAS), Stratigraphic Unit (US), Archaeological Finding (RA), "Wall Covering" Stratigraphic Unit (USR), Paleo-Anthropological Remains and Archaeological Monuments (MA) (D'Andrea, 2006).

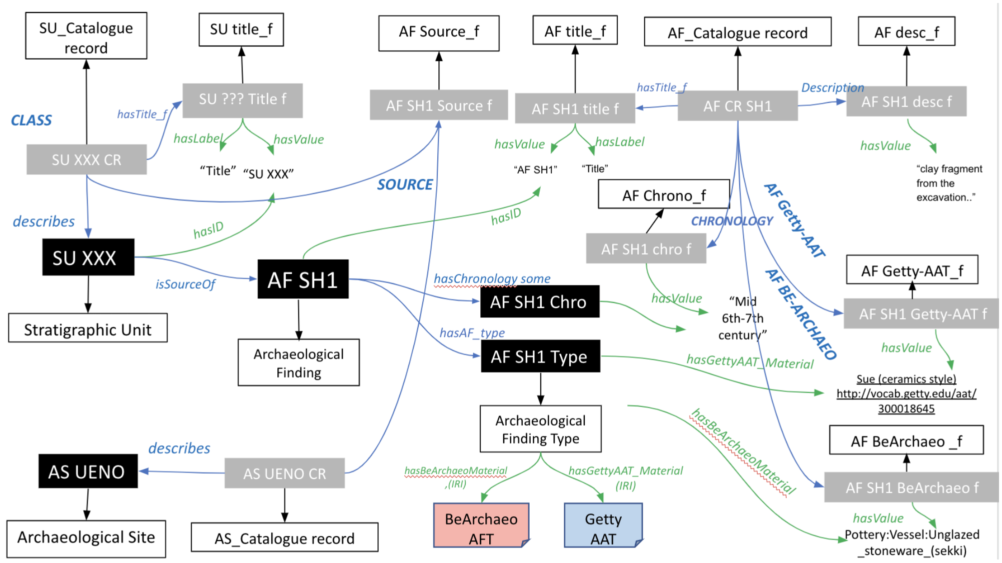


Figure 48: A diagram to show the conceptual relation between the archaeological finding(SH1), and the archaeological site where it was found in.

For example, in Figure 48, archaeological finding SH1 is described by the catalogue record "AF CR SH1" (Archaeological Finding Catalogue Record SH1); SH1 is part of the archaeological finding Kibi SH1, described by the catalogue record AF CR Kibi SH1. SH1 was found in (property *hasSource*) Ueno archaeological site, described by the catalogue record AS CR UENO SH1. Strings denoting ids are connected to the items as well as to their catalogue records.

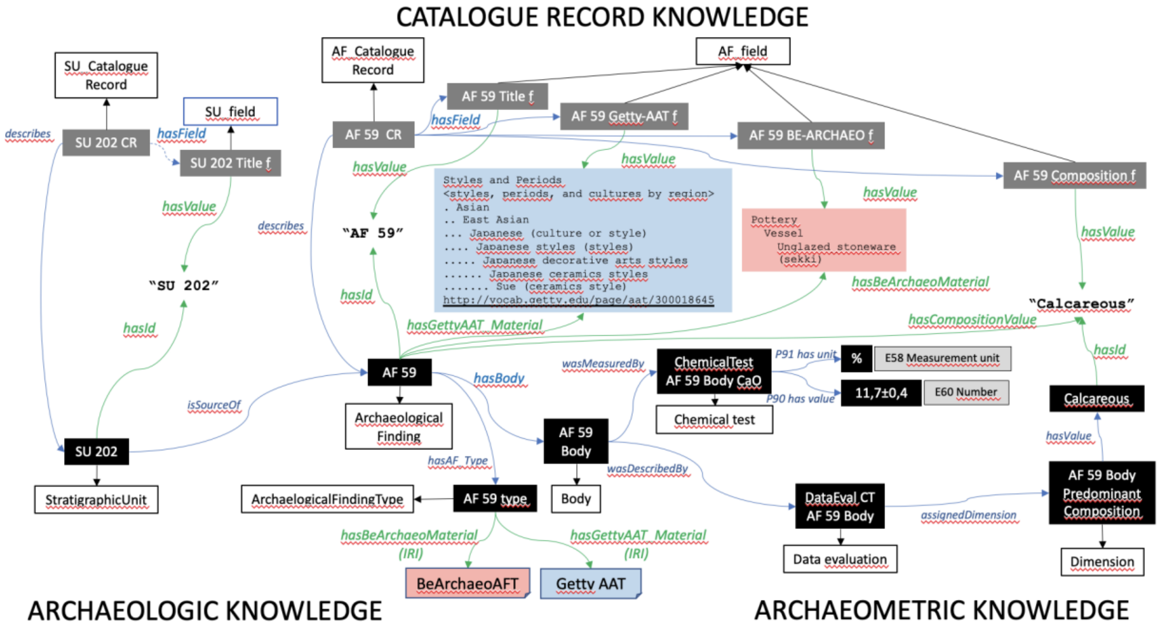


Figure 49: Modeling of the archaeological finding “AF 59”, exemplifying archaeological and archaeometric knowledge, respectively, and the corresponding fields in the archaeological finding record.

Figure 49 shows the modelling of the archaeological finding “AF 59”, exemplifying archaeological and archaeometric knowledge, respectively, and the corresponding fields in the archaeological finding record. The rectangles in grey or black are the individuals; the white rectangles are the classes; object properties are depicted as blue lines, while datatype properties are depicted as green lines; the three elements in Courier font are the strings that are actually written in the final form interface.

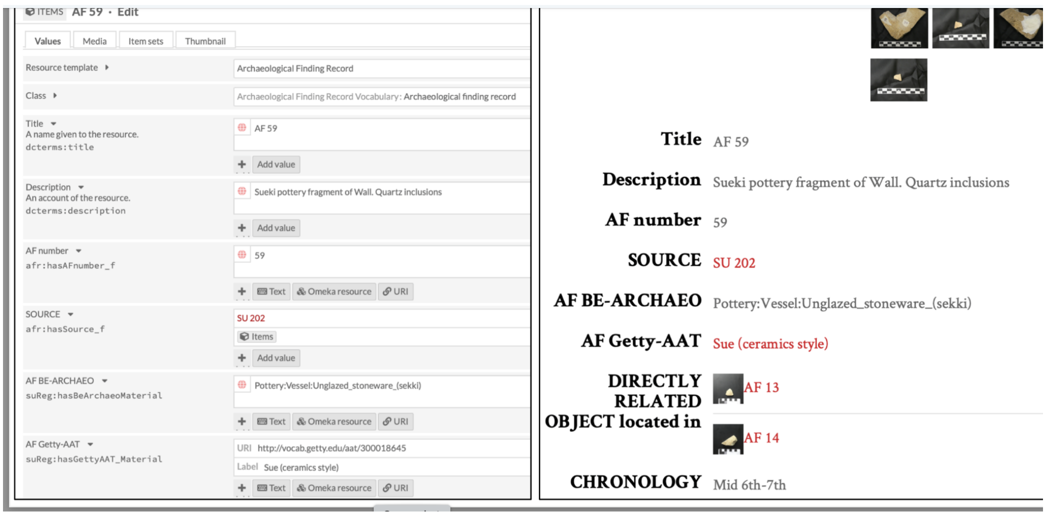


Figure 50: An exempt Omeka-built website of the BeArchaeo project for AF59. [68]

[68] https://bearchaeo.di.unito.it/omeka-s/s/bearchaeo-resources-site/page/welcome (last visited on 15 January 2022)

Figure 50 is a screenshot from the BeArchaeo resources website, concerning Archaeological finding no. 59, with the related fields and media as seen on the left, the back-end; on the right, the front-end of the application software. Elements in red are links to other elements of the documentation (e.g., Stratigraphic Unit 202) or to some external knowledge source (e.g., Getty AAT thesaurus). Also, it can become a publication and dissemination outcome (e.g., the digital image is part of a virtual collection exhibited through a website). In this case, Omeka software is used to publish the collection online. It uses RDF to describe items and collections and it relies on Linked Data for publishing the collections. Although it is not a repository software, it can be used with the repository platforms such as Fedora[69] or Dspace[70]. In the Omeka back-end, collection administrators and collaborators can contribute, annotate, organise, and manage the content using various plug-ins and tools while in the front end users also can do almost most of these things using a plug-in. MyOmeka plug-in is designed to make the process of creating an exhibit or selecting material to do. It allows users to store the files on the external cloud platforms rather than the local server disk. Using Omeka-S software, many cultural institutions have been disseminating their collections and curated content over the years[71]. Ontologies and thesauri were instantiated in RDF/XML and HTML, including a graph representation and published online in Link Open Data (LOD) form.

5.5 Evaluation of the approach with feedback from the scientific team

For the evaluation of the model, there have been organised two half-day workshops that hosted eighteen researchers from diverse backgrounds and scientific fields. Half of the participants were part of the BeArchaeo team, while the other half consisted of researchers working in archaeology and other related cultural heritage domains. There were international participants from Italy, Portugal, Brazil, Ukraine and Turkey, who work in a multidisciplinary team with four archaeologists (with different period/location backgrounds), two museologists, one information scientist, one 3D modelling designer, one dating expert and nine specialists from fields of archaeometry with backgrounds in chemistry, biology, physics and Earth sciences, respectively. The workshop had two objectives; first, it aimed at presenting the major theoretical and contextual background of the BeArchaeo database and the digital data

[69] https://duraspace.org/fedora/ (last visited on 19 November 2021)
[70] https://duraspace.org/dspace/ (last visited on 19 November 2021)
[71] https://omeka.org/classic/directory/ (last visited on 5 November 2021)

curation schema, secondly the audience was encouraged to employ the back-end interface provided on a development site (where experimental annotations and software modules are tested before being implemented in the production site). Also, the participants were asked to give their feedback on the annotation schema of the catalogue records while the workshop moderator was carrying on form-filling activities, starting on exemplary findings and moving to novel archaeometric cases, to suggest individual encodings on the BeArchaeo online publishing platform.

A first general statement was that the semantic approach to the database led interdisciplinary teams to appraise the core of the encoding process and mediate between the various habits and practices related to established national or disciplinary procedures. Going cross-countries, in the team of archaeologists, some supported the requirement of some national authorities for mandatory entries (encoded through object and datatype properties), while others have pointed out that other national authorities are less committed. The agreed solution was to leave semantic properties to be optionally valued while developing specific interfaces for the national contexts (currently, we have a European interface in the English language (based on the Italian Ministry of Culture forms that we mentioned earlier in this thesis) and a Japanese interface (in development). In terms of the cross-disciplinarity aspect, the archaeometric areas that were not engaged in the current development of the archaeometric knowledge, for example, the biologists, were able to catch the tenets of the semantic encoding; in practice, the workshop could trigger the process for the extension of the archaeometric encoding as well as identify the entities, namely the stratigraphic units for biologists, that can pivot the form filling process in synergy with the archaeological recordings.

The issue of having some mandatory property also emerged for the archaeometric investigation. In particular, it seems that the property concerning "the acquisition details" should be mandatory, as it has been often stressed that instrumental details and sample treatment are very relevant information to be linked to scientific data. Considering the feedback, we decided to act mostly on the interface of the filling forms, by providing a message that illustrates the importance of the acquisition details and the necessity of inserting such information in the individual entries of the archaeometric investigations. All researchers acknowledged that being educated about the digital data curation schema underlying the semantic encoding was very helpful in understanding the form filling process, especially in the relationship between the archaeological annotations and interpretations and the archaeometric investigations and interpretations.

While the current model is very inclusive in terms of the media and data to be included in the representation for proper documentation of the outcomes of the on-the-field and the in-the-lab activities, respectively, there is an ongoing discussion in the archaeological disciplines on how to be effective in the report of selected information in the repository and how to deal with the interdisciplinary knowledge, to include and link the different cues that come from the different approaches. For example, an archaeologist pointed out that the representation must include the Harris matrix to support the identification of the stratigraphic units; however, going back to the national diversity issues above, some others noticed that the Harris matrix is not generally adopted in the Japanese archaeological studies. Indeed, during the realisation of the BeArchaeo project, several interesting issues also arose from the different excavation techniques that pertain to the two schools of archaeology. Most of the archaeological knowledge available relies on concepts and terms, such as trenches, sections, and rooms, that have slightly different definitions according to the two traditions (e.g., in terms of depth of a trench accepted as a default); so, the ontological model should be adequately updated to include such differences and promote more fruitful collaboration for the international teams. However, the current representation has been evaluated as particularly valuable in supporting the construction of new knowledge through the many interpretations of the data that are linked to archaeological entities, together with the acquisition and processing phases that report on the setting and tools employed. In particular, some archaeologists reported that the organised repository could effectively support the comparison of the interpretations as they emerge while information grows from data production and modelling during the ongoing project activities. This is particularly appreciated in the context of reflexivity in archaeology.

A missing feature of the current semantic model is the encoding of the sampling procedures, which are well described in the CRMsci model, as prominent in scientific investigations. It is customary to produce samples from some archaeological findings, to perform some individual measurements that are then compared to provide some parameter evaluation for the whole archaeological finding (this happens, e.g., for archaeomagnetism researchers). However, our efforts in the conceptualization process have given priority to the representation of objects that are composed of a number of fragments retrieved individually and subsequently analysed to discover that they were part of a single object. Both fragments and composed objects have the status of entities in the representation, with archaeological data and archaeometric investigations attached to them.

Conclusion of the Chapter

In conclusion, for the evolution of the BeArchaeo project, the current representation of composed objects can be immediately adapted to the sampling issue, when limited to cases where the samples have the status of recorded items and not simply samples taken for measurements and then considered only a support of the interpretation process. Further developments are needed in the future to address this specific feature to provide a consistent representation of the archaeometric investigations. The publication of the ontology and the availability of a widespread CMS can be easily replicated in further projects. In future, ontology[72] and the derived database will also be employed in the definition of the contents of the exhibition that will present the outcomes of the BeArchaeo project to large audiences in collaboration with museum institutions in Europe and Japan.

[72] The BeArchaeo ontology is publicly available at /purl.org/beArchaeo (last visited on 5 March 2022)

Chapter 6

Overall Discussion and Future Directions

In this chapter, we discuss the findings and lessons learned from the analysis and the feedback from the workshops carried out to train the specialists and discuss the potential issues with them. As Bearchaeo is a transdisciplinary project, we worked with several scientists from diverse backgrounds. We discuss the issues that came out from the workshops as well as the direct feedback from the users of the database. Later, we provide some insights and share some lessons learned from the technical requirements of the project. We also comment on the system design and the user experience aspect of the database with potential redesign suggestions.

In the last section, we summarise the overall discussion of our comprehensive approach to the conceptualization of the archaeological and archaeometric domains at the base of a transdisciplinary approach to archaeological investigations. Together with the evaluation of the digital data curation workflow, it concludes with the examination of its benefits for the semantic organisation of the archaeological data in support of the coordination of all the tasks, from the excavation planning to the final exhibition of the results.

6.1 Analysis of the findings

There is no doubt today that the trajectories along the Semantic Web are evolving and having an impact on the richness and heterogeneity of the cultural heritage domain as a result of interdisciplinary research combining the two areas of information science and archaeology. Moreover, the frequent sceptical attitude of the communities of heritage professionals toward the promises of the new technological era can only be overcome with specific contributions by providing effective digital curation models for the cultural heritage domain. This research aimed to contribute to the current discussion from a holistic perspective where the evaluation of the digital data curation is coupled with more transdisciplinary contributions concerning archaeological domain aspects and issues. A transversal and interdisciplinary research path, going from the representation to the dissemination of archaeological data, information and knowledge, has included the high-level digital data curation model on which relevant and interrelated tools and technologies have been identified.

We have started from the general definitions of the digital curation process and provided some background information and related terminology by exploring the digital technologies for representation, creation, visualisation and maintenance of the cultural heritage data. Beyond being a fundamental aspect in the context of the Semantic Web, the area of domain knowledge representation is the one in which notable results have been obtained in the last few years, with the standardisation of a domain ontology (including the CIDOC CRM). Our discussion specifically took into consideration the in-depth analysis, experimentation and discussion of the semantic encoding of the archaeological data model, which was based on the conceptualisation with the Semantic Web languages (mainly RDFS and OWL), actually the major methodological approach. The contribution of domain experts in defining the appropriate conceptualisation of the archaeological data model is fundamental (Binding *et al.*, 2008), but there are still a few frameworks that can reflect the archaeological activity in all its complexity and reflexivity, together with full documentation of the archaeological knowledge. Nevertheless, the digital curation approach is fundamental to making it possible for scientists to evaluate, understand and effectively use the data.

The analysis of both the current trends in the development of the Semantic Web and the existing digital curation frameworks and projects in cultural heritage and archaeology represented the starting point of investigation in this research. On one hand, the technological tools together with the new approaches in the deployment of Semantic Web applications emerged as the backbone for the project. On the other hand, the comprehensive review of the most relevant projects that adopted these approaches in the context of archaeology made it possible to precisely identify the key aspects on which to focus, according to the aforementioned interdisciplinary perspective.

This review demonstrated that even if the more theoretical principles and proposals of digital curation can be the basis of the emerging field today, the analysis of real-life applications is of primary relevance since it dramatically improves the evaluation of the new approaches. For this reason, a key direction and contribution of our research have been the joint discussion of theoretical and methodological aspects, together with the discussion of more practical issues at the application level. These discussions have greatly benefited from the application to a real case study and the development of the Digital Data Curation Model for archaeological projects. Archaeological investigations are challenging because of their multidisciplinarity in the contemporary era and the intense contribution of the archaeometric disciplines. There are many individual approaches to digital curation in the archaeological field nowadays.

A central aspect of this PhD research has been the issue of archaeological knowledge representation, in terms of the approaches for modelling domain data with the contribution of diverse disciplines. We used BeArchaeo as a digital-data-curation born project that carries out archaeological excavation and the related archaeometric analyses of the Tobiotsuka Kofun, located in Soja city in Okayama Prefecture of Japan. We have presented the Digital Data Curation Model alongside a transdisciplinary ontology-based approach to the encoding of archaeological and archaeometric knowledge. In particular, we have set up a procedure for addressing the transdisciplinary endeavour and we developed a prototype ontology of the interconnected archaeological and archaeometric domains, respectively. These issues are particularly relevant for the digital data curation of an archaeological investigation when devising how the knowledge is linked to the form interfaces, for collecting the data as the excavation goes on, to be continued in the analysis labs, and eventually with the design of the exhibition. We have identified the major entities that are required for a reflexive methodology of archaeology, especially in its relationship with archaeometric knowledge. The experimentation of our approach on the archaeological data coming from the case study made it possible to define an evaluation setting and get insights from scientists.

The conceptual model of BeArcheo is the outcome of several modelling sketches and subsequent discussions carried out by the members of the archaeological and archaeometric teams, representing the several disciplines involved. The conceptualization was developed in support of a digital data curation framework that served the needs of an ongoing archaeological investigation. The conceptual model and the ontology of the archaeometric knowledge served the design and implementation of the interface forms for both archaeological and archaeometric filling, to enable researchers operating on the field and afterwards in the labs to load their results into the database. Also, it was evident that ontology is the base for a CMS-based web platform for supporting the archaeologist’s work in recording the excavation and interpretation activities.

The results of the research following this approach are particularly valuable for scientists who work in the field of archaeology and aim to design and/or take part in the increasing number of projects where the semantic integration of archaeological data contributes from the excavation process to the organisation of the exhibition. As far as we know, BeArchaeo is the first born-semantic project that assumes a joint archaeological/archaeometric perspective from the start. The transdisciplinary, multicultural, and multilingual character of BeArchaeo raises a high demand for the interoperability of knowledge and data. The encoding of the archaeological knowledge in an ontology that is compliant with CRMarchaeo, and so to CIDOC-CRM, and the implementation of a CMS-based solution for a concrete project, namely BeArchaeo, can have a deep impact on fostering projects that adhere to the Semantic Web paradigm and address the issues with data sharing effectively. The vocabularies are encoded as custom vocabularies into an installation of the semantics-based Content Management System Omeka-S.

To improve the ontology interoperability, we connected the ontology with other resources for the cataloguing of cultural heritage assets by replacing a number of customised vocabularies with domain ontologies (e.g., for chronology and formation processes). The catalogue record module is connected to the archaeological and archaeometric knowledge with the possibility to perform inferences and consistency checking of the interpretations in the future. The forms have been deployed as "Resource Templates", with the fast prototyping of user interfaces for both the back-end of the system, accessible by the archaeologists and the archaeometrists and the front-end, where supervisors and stakeholders check the development of the archive and the related findings. Also, reflecting the multicultural and multilingual specifications of the BeArchaeo project, knowledge interoperability between Japanese-speaking and English-speaking researchers as well as data terminology have been addressed by providing also in Japanese language resource templates for the Archaeological Finding and Stratigraphic Unit records, respectively. Also, we have uploaded rich media materials such as photos and 3D models acquired from photogrammetry and laser scanning, that are being used for interpretation and will be the basis for the exhibition.

6.2 Legacy of the research and future work

A preliminary achievement of this research has been the design and implementation of a semantic database for the encoding and storing of the digital data concerning an archaeological excavation and addressing the metadata belonging to several disciplines though concerning the same cultural heritage object (a stratigraphic unit, archaeological site or an archaeological finding). The BeArchaeo project, having notable inspiration from previous pioneering works, systematised this transdisciplinary approach to digital data curation and provided a viable system for the global supervision of digital data that are generated during the very long lifetime of an archaeological project. The convergence of many disciplines on single objects as well as the usage of the same data in many interfaces provided us with a stimulating challenge for the future of cultural heritage management and communication.

In the near future, we aim the continuation of the encoding of further archaeometric aspects and the strict connection with the archaeological interpretations, to implement some form of automatic reasoning in the data collection. The multi-disciplinary, multicultural, and multilingual characters of BeArchaeo raise a high demand for the interoperability of knowledge and data. The alignment with CIDOC-CRM is pursued at the disciplinary level, by aligning the archaeological and the archaeometric descriptions through the CRMarchaeo and CRMsci models, where possible. As the project database is growing in the collection of data, the user interfaces will be improved for engaging a higher number of diverse researchers and promote the usage of the conceptual model in other archaeological and archaeometric projects.

Omeka-S was chosen because of its practicality about the user experience as frontend, which has been an immediate solution for monitoring the project's initial database schema (given some previous experience with the tool). In future, the current predefined is planned to be replaced by a customised interface, while continuing to serve as a backend to the database monitoring. We are also working on a novel repository (currently a Google drive folder) for the media supporting the archaeometric analyses and interpretations. In particular, we are currently in the phase of analysing the requests about the possible uses of the data in the future, to devise the best repository solution. In future, we are going to evaluate the contribution of the centralised semantics-enhanced digital data curation in its impact on the final exhibition. In particular, we are going to address a deep analysis of the potential disciplinary targets, to propose advanced specific user interfaces to the centralised data concerning cultural heritage interpretation and for exhibition purposes.

The BeArchaeo archaeological team was a proper representative of the “archaeological community”: the Japanese archaeologists are strictly linked to the Japanese Research Institute for the Dynamics of Civilization, the Portuguese archaeologists are part of the Centro de Arqueologia de Universidade de Lisboa, and the Italian archaeologists are set within the International Research Institute for Archaeology and Ethology. The realisation of an overall approach, together with the adherence to well-known standards and an implemented workflow from the excavation design to the exhibition, can greatly contribute to the replication of the method across other projects and easily be adopted in further initiatives. Lastly, the scope of this thesis was the scientific investigations and representation of digital data curation in archaeology rather than scholarly publishing or data curation or aggregation which I would like to explore further after this thesis.

BIBLIOGRAPHY

Abbott, D. (2008) *What is digital curation?* Digital Curation Centre: Edinburgh, UK. Retrieved October 28, 2019, from http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation

Akers, K. and Doty, J. (2013) Disciplinary Differences in Faculty Research Data Management Practices and Perspectives. *International Journal of Digital Curation,* 8(2), pp. 5-26.

Akers, K. and Green, J. A. (2014) Towards a Symbiotic Relationship Between Academic Libraries and Disciplinary Data Repositories: A Dryad and University of Michigan Case Study. *International Journal of Digital Curation*, 9(1), pp. 119-131.

Akmon, D., Zimmerman, A., Daniels, M. and Hedstrom, M. (2011) The Application of Archival Concepts to a Data-Intensive Environment: Working with Scientists to Understand Data Management and Preservation Needs. *Archival Science,* 11(3), pp. 329-348.

Aliaga D. G., Bertino E., and Valtolina S. (2011) DECHO—a framework for the digital exploration of cultural heritage objects. *Journal on Computing and Cultural Heritage (JOCCH)* 3, 3, Article 12 (March 2011), 26 pages. DOI:https://doi.org/10.1145/1921614.1921619

Allard, S. (2014) *Evaluating a Complex Project: DataONE. In: J. Ray, ed. Research Data Management: Practical Strategies for Information Professionals.* West Lafayette: Purdue University Press, pp. 255-274.

Anderson, J.Q., Rainie, L., Batorski, M.M., and Hadden, D. (2010) *The impact of the Internet on institutions in the future.*

ANDS. (2011) *Australian National Data Service: Governance Structure.* Accessed September 3, 2021, from http://ands.org.au/governance.html

Ashley, K. (2010) IDCC10 Keynote [Video]. Proceedings of the 6th International Digital Curation Conference. Accessed September 3, 2021, from http://vimeo.com/17614413.

Ashley, M., Tringham, R., Perlingieri, C. (2011) 'Last House on the Hill: Digitally Remediating Data and Media for Preservation and Access'. *Journal on Computing and Cultural Heritage (JOCCH).* 4, 13:1–13:26.

Artioli, G. (2010) *Scientific Methods and Cultural Heritage: An introduction to the application of materials science to archaeometry and conservation science.* Oxford Scholarship Online.

Beagrie, N. (2000) The JISC digital preservation focus and the digital preservation coalition. New Review of Academic Librarianship, 6, 257-267.

Benardou, A., Constantopoulos, P., Dallas, C., & Gavrilis, D. (2010). A conceptual model for scholarly research activity. iConference 2010 http://hdl.handle.net/2142/14945

Berggren, Å., Hodder, I. (2003) Social Practice, Method, and Some Problems of Field Archaeology. *American Antiquity* 68, no. 3 (July 1): pp. 421–434.

Berggren, Å., Dell'Unto, N., Forte, M., Haddow, S., Hodder, I., Issavi, J., Lercari, N., Mazzuccato, C., Mickel, A., and Taylor, J.S. (2015) Revisiting reflexive archaeology at Çatalhöyük: integrating digital and 3D technologies at the trowel's edge. *Antiquity*, 89 (344), pp. 433–448.

Berners-Lee, T. B., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American*, 284(5), 34-43.

Binding, C., May, K., and Tudhope, D. (2008) *Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC-CRM.* In Christensen Dalsgaard, B., editor, Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries, volume 5173 of Lecture Notes in Computer Science, pages 280–290, Berlin Heidelberg. Springer.

Bitgood, S. (1990) Toward an Objective Description of the Visitor Immersion Experience. *Visitor Behaviour*, vol. 5, no.2, pp.28.

Bordoni, L., Mele, F. and Sorgente, A. (eds.) (2016) *Artificial Intelligence for Cultural Heritage*, Newcastle, Cambridge Scholars Publishing.

Bouchenaki, M. (2003) The interdependency of the tangible and intangible cultural heritage. Proceedings of the 14th ICOMOS General Assembly and International Digital Cultural Heritage and Linked Data 14 Liber Quarterly Volume 30 2020 Symposium, 1–5. Retrieved March 17, 2021, from https://www.icomos.org/victoriafalls2003/papers/2%20-%20Allocution%20Bouchenaki.pdf

Brase, J. (2004) Using Digital Library Techniques – Registration of Scientific Primary Data. 488-494. 10.1007/978-3-540-30230-8_44.

Brizard, T., Derde, W. and Silberman, N. (2007) *Basic Guidelines for Cultural Heritage Professionals in the Use of Information Technologies. How can ICT support cultural heritage?* Available at http://www.enamecenter.org/files/documents/Know-how%20book%20on%20Cultural%20Heritage%20and%20ICT.pdf [accessed: 13 September 2021].

Buranarach M, Buranasing W, Rungcharoensuksri S, Sarawasee P, Ngootip T, and Chansanam W. (2022) Metadata Integration Framework for Data Integration of Socio-Cultural Anthropology Digital Repositories: A Case Study of Princess Maha Chakri Sirindhorn Anthropology Centre. Informatics. 2022; 9(2):38. https://doi.org/10.3390/informatics9020038

Capurro, C., Nollet, D. and Pletinckx, D. (2014) “Reconstruction of the interior of the Saint Salvator abbey of Ename around 1290”, *VAR*, vol. 5, no. 11, November 2014.

Capurro, C., Nollet, D. and Pletinckx D. (2015) “Tangible Interfaces for Digital Museum Applications” *2015 Digital Heritage*, 2, pp. 271–76.

Carboni, N. and de Luca, L. (2016) Towards a conceptual foundation for documenting tangible and intangible elements of a cultural object. *Digital Applications in Archaeology and Cultural Heritage*, 3(4), 108–116. https://doi.org/10.1016/j.daach.2016.11.001.

Chisholm, H., ed. (1911). "Crozier". Encyclopædia Britannica. Vol. 7 (11th ed.). Cambridge University Press. p. 520.

Coburn, E. (2010) LIDO - Lightweight Information Describing Objects Version 1.0. Retrieved from http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html

Codd, E. F. (June 1970). "A Relational Model of Data for Large Shared Data Banks" (PDF). Communications of the ACM. 13 (6): 377–387.

Collins, E. (2012). The National Data Centres. In: G. Pryor, ed. Research Data Management. London: Facet, pp. 151-172

Constantopoulos, P. and Dallas, C. (2010) Aspects of a Digital Curation Agenda for Digital Heritage. Athens, IEEE.

Costopoulos, A. (2016). Digital Archeology Is Here (and Has Been for a While). Frontiers in Digital Humanities. 3. 10.3389/fdigh.2016.00004.

Constantopoulos, P., Dallas, C., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., Kotidis, Y. and Papatheodorou, C. (2009) DCC&U: An Extended Digital Curation Lifecycle Model. *International Journal of Digital Curation*. 4. 10.2218/ijdc.v4i1.76.

Corti, L., Van Den Eynden, V., Bishop, L. and Woolard, M. (2014) Managing and Sharing Research Data: A Guide to Good Practice, Los Angeles: Sage.

Crane, D. and Chadwick, I. (2017) Practical strategies for Research Data Management. http://www.open.ac.uk/blogs/the_orb/wp-content/uploads/2017/10/slides-RDM_PracticalStrategiesForRDM_2017-11.pdf

Cripps, P., Greenhalgh, A., Fellows, D., May, K. and Robinson, D. (2004) Ontological Modelling of the work of the Centre for Archaeology.

Cripps, P. and May, K. (2004) To OO or not to OO? – Revelations from Defining an Ontological Model for an Archaeological Information System. CAA 2004 presentation. (Accessed September 12, 2021): https://www.slideshare.net/pauljcripps/to-oo-or-not-to-oo-revelations-from-defining-an-ontology-for-an-archaeological-information-system-2578773

CRM-EH (2014). CRMEH - English Heritage Extensions to CIDOC CRM for the Archaeological Domain. Retrieved from http://www. heritagedata.org/crmeh/crmeh_current.rdf

Crofts, N., Doerr, M., Gill, T., Stead, S., and Stiff, M. (2009). Definition of the CIDOC Conceptual Reference Model Version 5.0.1. Technical report, ICOM/CIDOC CRM Special Interest Group, Accessible at: http://cidoc.ics.forth.gr/docs/ cidoc_crm_version_5.0.1_Nov09.pdf.

Crowston, K. and Qin, J. (2011) A Capability Maturity Model for Scientific Data Management: Evidence from the Literature. Proceedings of the American Society for Information Science and Technology. 48. 10.1002/meet.2011.14504801036.

Curdt, C., Hoffmeister, D., Waldhoff, G., Jekel, C. and Bareth, G. (2012) Development of a metadata management system for an interdisciplinary research project. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. I-4. 7-12.

Çatalhöyük Database [WWW Document] http://www.catalhoyuk.com/research/database

Dallas, C. (2007) An Agency-Oriented Approach to Digital Curation Theory and Practice. Toronto, Archives and Museum Informatics.

Dallas, C., Constantopoulos, P., Androutsopoulos, I., Angelis, S., Deligiannakis, A., Gavrilis, D., Kotidis, Y. and Papatheodorou, C. (2009) DCC&U: An Extended Digital Curation Lifecycle Model. International Journal of Digital Curation. 4.

Dallas, C. (2015a). Curating Archaeological Knowledge in the Digital Continuum: from Practice to Infrastructure. Open Archaeology. 1. 10.1515/opar-2015-0011.

Dallas, C. (2015b). Digital curation beyond the “wild frontier”: a pragmatic approach. Arch Sci 16, 421–457. https://doi.org/10.1007/s10502-015-9252-6

Damiano, R. and Lombardo, V. (2015). Labyrinth 3D. Cultural archetypes for exploring media archives. *Digital Creativity* 27(3), pp. *234-255.* [Online]. Available at:

https://www.researchgate.net/publication/306528983_Labyrinth_3D_Cultural_archetypes_for_exploring_media_archives (Accessed: 20 March 2021)

Demetrescu, E., and Ferdani D. 2021. "From Field Archaeology to Virtual Reconstruction: A Five Steps Method Using the Extended Matrix" Applied Sciences 11, no. 11: 5206. https://doi.org/10.3390/app11115206

Donnelly, M. and Jones, S. (2009) DCC data management plan content checklist. Digital Curation Centre. Accessed September 13, 2021, from http://www.era.lib.ed.ac.uk/handle/1842/3343.

Doneus, M. and Neubauer, W. (2005) Laser scanners for 3D, documentation of stratigraphic excavations. In: International Workshop on Recording, Modeling and Visualization of Cultural Heritage, Ascona, May 22-27, Taylor & Francis.

Dowdy, B., Raeford R. (2014) Electronic Resources Workow: Design, Analysis and Technologies for an Overdue Solution. Serials Review 40, 3(July 2014), 175–187.

D'Agnano, F., Balletti, C., Guerra, F. and Vernier, P. (2015) "Tooteko: A Case Study of Augmented Reality for an Accessible Cultural Heritage. Digitization, 3D printing and sensors for an audio-tactile experience", *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40 (5), p. 207. Available at http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W4/207/2015/isprsarchives-XL-5-W4-207-2015.pdf [accessed: 10 August 2021].

D'Andrea A. (2006) Documentazione Archeologica,Standard e Trattamento Informatico. BUDAPEST:archeolingua, ISBN: 963-8046-76-7

Edwards, P. N. (2004) 'A Vast Machine': Standards as Social Technology. *Science,* 304(5672) 7 May, pp. 827-828.

Edwards, P. N. (2011) Science Friction: Data, Metadata, and Collaboration. Social Studies of Science, 41(5), pp. 667-690.

Fai, S., Graham, K., Duckworth, T., Wood, N. and Attar, R. (2011) "Building information modelling and heritage documentation" in *Proceedings of the 23rd International Symposium,* *International Scientific Committee for Documentation of Cultural Heritage (CIPA), Prague, Czech Republic*, pp. 12–16.

Faniel I., Kansa E. and Kansa S. W. (2013) The challenges of digging data: a study of context in archaeological data reuse, in Proceedings of 13th ACM/IEEE-CS joint conference on digital libraries, Indianapolis, IN, New York: ACM, 22–25 July 2013, pp. 295–304.

Forte, M. (1994) Scientific visualization and archaeological landscape: the case study of a terramara, Italy, in “Archaeological and Geographical Information Systems”, (ed. by G.Lock e Z.Stancic), London, 1995, 231-238

Forte, M., Dell'Unto, N., Jonsson, K. and Lercari, N. (2015). Interpretation Process at Çatalhöyük using 3D. In I. Hodder, and A. Marciniak (Eds.), Assembling Çatalhöyük (Vol. 1). (Themes in Contemporary Archaeology; Vol. 1). Maney Publishing.

FORTH-ICS. (2012). 3M - Mapping Memory Manager. Retrieved from http://139.91.183.3/3M

Forum on Information Standards in Heritage (FISH) (2012). MIDAS Heritage. The UK Historic Environment Data Standard v1.1. Retrieved from https://historicengland.org.uk/images- books/publications/midas-heritage

Gardener, H. (1985) *Frames of Mind: The Theory of Multiple Intelligences.* New York: Basic Books

Gardin, J. (1990) *Mathematics and Information Science in Archaeology. A Flexible Framework*, volume 3 of Studies in Modern Archaeology, chapter The Structure of Archaeological Theories, pages 7–25. Holos Verlag, Bonn.

Gardin, J. and Roux, V. (2004) The Arkeotek Project: a European Network of Knowledge Bases in the Archaeology of Techniques. *Archeologia e Calcolatori,* 15:25–40.

Gavrilis, D., Dallas, C., Angelis, S.: A curation-oriented thematic aggregator. In: Aalberg, T., Papatheodorou, C., Dobreva, M., Tsakonas, G., Farrugia, C.J. (eds.) (2013) TPDL 2013. LNCS, vol. 8092, pp. 132–137. Springer, Heidelberg. doi:10.1007/978-3-642-40501-3_13

Gay, E., Galor, K., Cooper, D. B., Willis, A., Kimia, B. B., Karumuri, S., Taubin, G., Doutre, W., Sanders D., and Liu, S. (2010) "REVEAL Intermediate Report", Proceedings of CVPR Workshop on Applications of Computer Vision in Archaeology (ACVA'10).

Gladney, H. M. (2007) *Preserving digital information*. Berlin: Springer.

Gold, A. (2010) *Data Curation and Libraries: Short-Term Developments, Long-Term Prospects*. San Luis Obispo: California Polytechnic State University.

Greenberg, J. (2016) Big Data, Little Data, No Data: Scholarship in the Networked World: by Christine L. Borgman. Cambridge, MA: MIT Press,

Gregory, A. (2011) "The Data Documentation Initiative (DDI): An Introduction for National Statistical Institutes" https://businessdocbox.com/Logistics/78970208-The-data-documentation-initiative-ddi-an-introduction-for-national-statistical-institutes.html

Gruber, T. (1993) “A Translation Approach to Portable Ontology Specifications”. *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220.

Hacıgüzeller, P. (2019) “Archaeology, Digital Cartography and the Question of Progress: The Case of Çatalhöyük (Turkey).” In ReMapping Archaeology: Critical Perspectives, Alternative Mappings, edited by M. Gillings, P. Hacıgüzeller, and G. Lock, 267–280. London: Routledge.

Hacıgüzeller, P., Taylor, J. 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), 1709-1730. https://doi.org/10.1515/opar-2020-0220

Harvey, R. (2010) *Digital Curation: A How to Do It Manual.* New York: Neal Schuman.

Havemann, S., Settgast, V., Berndt, R., Eide, Ø. and Fellner, D. W. (2009) The Arrigo Showcase Reloaded—towards a sustainable link between 3D and semantics. *Journal on Computing and Cultural Heritage (JOCCH)*, 2(1), 1-13.

Hedden, H. (2010) Taxonomies and controlled vocabularies best practices for metadata. Journal of Digital Asset Management 6, 279–284. https://doi.org/10.1057/dam.2010.29

Higgins S. (2008) The DCC curation lifecycle model. *International Journal of Digital Curation* 3, 453. https://doi.org/10.1145/1378889.1378998

Hodder, I. (1997) ‘‘Always momentary, fluid and flexible’: towards a reflexive excavation methodology, Antiquity, 71 (273), pp. 691–700. DOI: 10.1017/S0003598X00085410.

Hodder, I. (1999) *The archaeological process: An introduction.* Oxford, England: Blackwell.

Huggett, J. (2015a) ‘A manifesto for an introspective digital archaeology’, *Open Archaeology*, 1 (1), pp. 86–95. DOI: 10.1515/opar-2015-0002.

Huggett, J. (2015b) ‘Digital haystacks: open data and the transformation of archaeological knowledge, in Wilson, A.T. and Edwards, B. (eds) Open source archaeology: ethics and practice. De Gruyter Open, pp. 6–29. DOI: 10.1515/9783110440171-003.

Huggett, J. (2020) Is Big Digital Data Different? Towards a New Archaeological Paradigm, Journal of Field Archaeology, 45:sup1, S8-S17, DOI: 10.1080/00934690.2020.1713281

Hugh, D. 2012. A New Introduction to the London Charter. Paradata and Transparency in Virtual Heritage. Digital Research in the Arts and Humanities Series. Ashgate.

Hyvönen, E. (2009) *Semantic portals for cultural heritage.* In Handbook on ontologies (pp. 757-778). Springer, Berlin, Heidelberg.

Ioannides, M., Magnenat-Thalmann, N., Fink, E., Zarnic, R., Yen, A. Y. and Quak, E. (eds.) (2014). *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 5th International Conference, EuroMed 2014, Limassol, Cyprus, November 3–8, 2014, Proceedings* (vol. 8740). Springer

Isaac, A. and Summers, E. (2008). SKOS Simple Knowledge Organization System Primer. W3C Working Draft, World Wide Web Consortium, Accessible at: http://www.w3.org/TR/skos-primer.

Isaksen, L. (2008). The TRANSLATION Framework for Archaeological Excavation Data. Accessible at: http://eprints.soton.ac.uk/63320/.

Isaksen, L., Martinez, K., Gibbins, N., Earl, G., and Keay, S. (2009). Linking Archaeological Data. In Proceedings of the 37th Computer Applications and Quantitative Methods in Archaeology. In preparation.

Jeffrey, S. (2015) 'Challenging heritage visualisation: beauty, aura and democratisation', *Open Archaeology,* 1 (1), pp. 144–152. DOI: https://doi.org/10.1515/opar-2015-0008.

Jokilehto, J. (2017) Questions of authenticity concerning different cultures and preservation. 10.30618/978-989-658-467-2_02.

Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M. and Gergatsoulis, M. (2007). Integrating Dublin Core metadata for cultural heritage collections using ontologies. In International Conference on Dublin Core and metadata applications (pp. 128-139).

Kansa, E. C. and Kansa, S. W. (2013) We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication, Journal of Eastern Mediterranean Archaeology and Heritage Studies 1(1) (2013), 88–97.

Kansa, E.C. (2015) Click here to save archaeology. Available at: https://ekansa-pubs.github.io/click-here-to-save-archaeology/ (Accessed: 20 August 2021).

Kansa, S., Atici, L., Kansa, E., and Meadow, R. (2020). Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data. Advances in Archaeological Practice, 8(1), 40-52. doi:10.1017/aap.2019.36

Kansa, E. C. and Kansa, S. W., 2021. Digital Data and Data Literacy in Archaeology Now and in the New Decade, Advances in Archaeological Practice 9(1).

Karatas, T. and Lombardo, V. (2020) A Multiple Perspective Account of Digital Curation for Cultural Heritage: Tasks, Disciplines and Institutions, in Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020, T. Kuflik, I. Torre, R. Burke and C. Gena, eds, ACM, 2020, pp. 325–332.

Kress, G. (2010) *Multimodality: A Social Semiotic Approach to Contemporary Communication.* New York: Routledge

Lampe, K-H, Riede, K. and Doerr, M. (2008) Research Between Natural and Cultural History Information: Benefits and IT Requirements for Transdisciplinarity, ACM Journal on Computing and Cultural Heritage 1(1).

Laursen, D., Christiansen, K. F. and Olsen, L. L. (2012) “Management of Metadata for Digital Heritage Collections”, *Microform and Digitization Review*, 41 (3–4), pp. 151–58

Lee, C. A., Tibbo H. (2007) Digital Curation and Trusted Repositories: Steps Toward Success. *Journal of Digital Information*, 8.

Lee, C. A., Tibbo H. (2011) “Where’s the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills.” *ARCHIVARIA* 72: 123–168.

Lemonnier, P. (2012) Mundane Objects: Materiality and Non-verbal Communication (1st ed.). Routledge. https://doi.org/10.4324/9781315424255

Lercari, N., Shiferaw, E., Forte M. and Kopper, R. (2017) Immersive Visualization and Curation of Archaeological Heritage Data: Çatalhöyük and the Dig@IT App. *Journal of Archaeological Method and Theory.* DOI: 10.1007/s10816-017-9340-4.

Levine, M. (2014) Copyright, Open Data, and the Availability-Usability Gap: Challenges, Opportunities, and Approaches for Libraries. In: J. Ray, ed. Research Data Management: Practical Strategies for Information Professionals. West Lafayette: Purdue University Press, pp. 129-147

Lombardo, V. (2005) The Virtual Electronic Poem (VEP) Project. In: *ICMC-International Computer Music Conference Proceedings on Free Sound.* Catalonia College of Music, 4-5 September. Finland: SuviSoft Oy, pp. 451-454

Lombardo, V., Valle, A., Nunnari, F., Giordana, F. and Arghinenti, A. (2006). Archeology of multimedia. 269-278. 10.1145/1180639.1180706.

Lombardo, V., Valle, A., Fitch, J., Tazelaar, K., Weinzierl, S. and Borczyk, W. (2009). A Virtual-Reality Reconstruction of Poème Électronique Based on Philological Research. Computer Music Journal. 33. 24-47. 10.1162/comj.2009.33.2.24.

Lombardo, V., Damiano, R., Karatas, T. and Mattutino, C., 2020. Linking Ontological Classes and Archaeological Forms, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, J.Z. Pan, V.A.M. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne and L. Kagal, eds, Lecture Notes in Computer Science, Vol. 12507, Springer, 2020, pp. 700–715. doi:10.1007/978-3-030-62466-8_43 .

Lombardo V., Karatas T., Gulmini M., Guidorzi L. and Angelici D. 2022. Transdisciplinary approach to archaeological investigations in a Semantic Web perspective. In: SEMANTIC WEB. in press, pp. 1–30. URL: http://www. semantic-web-journal.net/content/transdisciplinary-approach-archaeological-investigations-semantic-web-perspective-1.

Lord, P. and Macdonald, A. (2003) e-Science curation report - Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision. http://www.cs.york.ac.uk/ftpdir/pub/leo/york-msc-2007/information/vsrcuration/science-dc-report.pdf. Accessed 20 Apr 2014

Lucas, G. (2012). "The Trouble with Theory". Understanding the Archaeological Record. Cambridge: Cambridge University Press. pp. 1–17. ISBN 978-1-107-01026-0.

Lukas D., Engel C. and Mazzucato, C. (2018) Towards a Living Archive: Making Multi-Layered Research Data and Knowledge Generation Transparent, Journal of Field Archaeology, 43:sup1, S19-S30, DOI: 10.1080/00934690.2018.1516110

Lyon, L. (2012) The Informatics Transform: Re-Engineering Libraries for the Data Decade. International Journal of Digital Curation. 7. 126-138. 10.2218/ijdc.v7i1.220.

Mantegari, G. (2010). Cultural heritage on the semantic web: from representation to fruition [PHD]. Università degli Studi di Milano-Bicocca.

McDavid, C. (2002) Archaeologies that hurt; descendants that matter: a pragmatic approach to collaboration in the public interpretation of African-American archaeology, World Archaeology. 34, 303–314.

Meghini, C., Dekkers, M., Gradmann, S., Clayphan, R., Molendijk, J., Purday, J. and Charles, V. (2016) Europeana Data Model Primer. (A. Isaac, Ed.). Retrieved from https://pro.europeana.eu/resources/standardization-tools/edm-documentation

Miles, A. and Bechhofer, S. (2009). Simple Knowledge Organization System Reference. Retrieved from https://www.w3.org/2004/02/skos/

Moser, S., Glazier, D., Phillips, J.E., el Nemr, L.N., *et al.* (2002) Transforming archaeology through practice: Strategies for collaborative archaeology and the Community Archaeology Project at Quseir, Egypt, World Archaeology 34, 220–248.

Niccolucci, F., Felicetti, A., Hermon, S. and Nys, K. (2009) Managing Full-text Excavation Data with Semantic Tools. In VAST (pp. 125-132).

Niccolucci, F., Hermon S., and Doerr. M. (2015) “The Formal Logical Foundations of Archaeological Ontologies.” In Mathematics and Archaeology, edited by Juan A. Barceló and Igor Bogdanovic, 86–99. Boca Raton, FL: CRC Press Taylor and Francis Group.

Opitz, R. S. and Johnson, T. D. (2016) Interpretation at the controller’s edge: Designing graphical user interfaces for the digital publication of the excavations at Gabii (Italy). Open Archaeology, 2(1).

O’Reilly, T. (2007) What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Communications & Strategies, 1(65), 17–37.

Parry, R. (2007) *Recoding the Museum: Digital Heritage and the Technologies of Change* London: Routledge

Patel, M., Coles, S., Giaretta, D., Rankin, S. and McIlwrath, B. (2009) The Role of OAIS Representation Information in the Digital Curation of Crystallography Data, IEEE eScience 2009, Oxford, UK.

Pennock, M. (2007) Digital curation: A life-cycle approach to managing and preserving usable digital information. Library & Archives. 1.

Peterson, A.J. (1991) Evaluation of Hypermedia and Interactivity in the Museum: A Constructivist Approach to Instructional Design. In: Bearman, D., (ed.) *Archives and Museums Informatics Technical Report No. 14 - Hypermedia & Interactivity in Museums: Proceedings of an International Conference.* Pennsylvania, 14- 16 October. Pittsburgh: Archives and Museum Informatics, pp.74-80.

Pletinckx D. (2007) *Virtex: a multisensory approach for exhibiting valuable objects*. [Online]. Available at: http://media.digitalheritage.se/2010/07/Virtex_TII.pdf (Accessed: 20 March 2019).

Poole, A. (2016) The Conceptual Landscape of Digital Curation. Journal of Documentation. 72. 10.1108/JD-10-2015-0123

Post, C., Chassanoff, A., Lee, C., Rabkin, A., Zhang, Y., Skinner, K. and Meister, S. (2019) Digital Curation at Work: Modeling Workflows for Digital Archival Materials. 39-48.

Pothen, P. (2006) Digital curation and preservation: Defining the research agenda for the next decade. Ariadne, 46. Accessed September 2, 2019, from http://www.ariadne.ac.uk/issue46/warwick-2005-rpt/.

Pouchard, L. (2015) Revisiting the Data Lifecycle with Big Data Curation. International Journal of Digital Curation. 10. 10.2218/ijdc.v10i2.342.

Poyart, E., Snyder, L., Friedman, S. and Faloutsos, P. (2011) VSim: real-time visualization of 3D digital humanities content for education and collaboration. In Proceedings of the 12th International Conference on Virtual Reality, Archaeology and Cultural Heritage (pp. 129-135).

Pryor, G. (Ed.) (2012) Managing Research Data. Facet. doi:10.29085/9781856048910

Prud'hommeaux, E. and Seaborne, A. (2008). SPARQL Query Language for RDF. W3C Recommendation, World Wide Web Consortium, Accessible at: http://www.w3.org/TR/rdf-sparql-query/.

Powell, A., Nilsson, M., Naeve, A., Johnston, P. and Baker, T. (2007) DCMI Abstract Model. Retrieved from http://dublincore.org/

Power C., Lewis A., Petrie H., Green K., Richards J., Eramian M. and Rijke M. D. (2017) Improving archaeologists' online archive experiences through user-centred design. *Journal on Computing and Cultural Heritage (JOCCH)*, 10(1), 1-20.

Ray, J. (2009) "Sharks, Digital Curation, and the Education of Information Professionals," Museum Management and Curatorship, vol. 24, no. 4, pp. 357–68

Rendgren, S. (2018) "What Do We Mean By 'Data'?" idalab blog, June 20, 2018. https://idalab.de/blog/data-science/what-do-we-mean-by-data.

Research Information Network and British Library. (2009) Patterns of information use and exchange: case studies of researchers in the life sciences. Accessed September 12, 2021, from http://www.rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf.

Richards, J.D. (2006a) "The Archaeology Data Service (ADS)", in Conservation Bulletin 51 *English Heritage*, 27.

Richards, J.D. (2006b) "Electronic publication in archaeology", in T.L.Evans and P.Daly, (eds.) *Digital Archaeology: bridging method and theory,* Routledge, 2006, 213-25.

Richards, J.D. and C. Hardman. (2008) Stepping back from the trench edge. An archaeological perspective on the development of standards for recording and publication. pp. 101-112. In: M. Greengrass and L. Hughes (eds.). The Virtual Representation of the Past. Ashgate, London.

Richards, J.D., Jeffrey S., Waller S., Ciravegna F., Chapman S. and Zhang Z. (2011) The Archaeology Data Service and the Archaeotools project: faceted classification and natural language processing. Pp. 31-56. In: S. Whitcher Kansa, E.C. Kansa and E. Watrall (eds.). Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication. Cotsen Institute for Archaeology Press, Los Angeles.

Ridge, M., and May, R. (2004) “Database and IT Developments [online].In: Çatalhöyük 2004 Archive Report.”http://www.catalhoyuk.com/archive_reports/2004/ar04_32.html

Ridge, M., May, R. and Mackie D. (2005) Çatalhöyük IT Team ArchiveReport 2005 [online]. In Çatalhöyük 2005 Archive Report. http://www.catalhoyuk.com/sites/default/files/media/pdf/Archive_Report_2005.pdf

Ridge, M., and Jones, S. (2006) IT Team [online]. In Çatalhöyük 2006Archive Report. http://www.catalhoyuk.com/sites/default/files/media/pdf/Archive_Report_2006.pdf]

Riley, J. (2014) Metadata Services. In: J. Ray, ed. Research Data Management: Practical Strategies for Information Professionals. West Lafayette: Purdue University Press, pp. 149-165.

Ross, S. (2000) Changing trains at Wigan: Digital preservation and the future of scholarship. National Preservation Office: London. Accessed September 10, 2019, from http://www.bl.uk/blpac/pdf/wigan.pdf.

Schloen, J. D. and Schloen, S. R. (2012) OCHRE: An online cultural and historical research environment.

Seifert C., Bailer W., Orgel T., Gantner L., Kern R., Ziak, H. and Granitzer, M. (2017) Ubiquitous access to digital cultural heritage. Journal on Computing and Cultural Heritage (JOCCH), 10(1), 1-27.

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) The semantic web revisited. IEEE intelligent systems, 21(3), 96-101.

Shi C., Li Y., Zhang J., Sun Y. and Yu P. S. (2016) "A survey of heterogeneous information network analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 1, pp. 17-37

Simon, N. (2010) *The Participatory Museum*. California: Museum 2.0

Snyder, L. M. (2014) VSim: scholarly annotations in real-time 3D environments. In DH-CASE II: Collaborative Annotations on Shared Environments: metadata, tools and techniques in the Digital Humanities (pp. 1-8).

Soergel, D. (2009). Digital Libraries and the Semantic Web. Keynote presentation at the 2009 International Conference for Digital Libraries and the Semantic Web. Accessible at: http://icsd-conference.org/sites/default/ files/SoergelICSD2009Keynote.ppt.

Taylor, J.M (2001) The UK e-science programme [Powerpoint presentation], e-Science London Meeting. http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/research/esci/ jtaylor.pdf. (Accessed 20 Sep 2021)

Terras, M. (2015) “Cultural Heritage Information: Artefacts and Digitization Technologies”, in *Cultural Heritage Information: Access and Management*, London, Facet Publishing, pp. 63–88.

The National Archives. (2011) Understanding digital continuity, version 1.2. Accessed September 20, 2021, from http://www.nationalarchives.gov.uk/documents/information-management/understanding-digital-continuity.pdf.

The Task Force on Archiving of Digital Information. (1996) Preserving digital information. In report to the Commission on Preservation and Access and the Research Libraries Group. Accessed September 19, 2021, from http://www.clir.org/pubs/reports/pub63watersgarrett.pdf.

Tibbo, H. (2015) Digital Curation Education and Training: From Digitization to Graduate Curricula to MOOCs. International Journal of Digital Curation. 10. 10.2218/ijdc.v10i1.352.

Tillett, D. B. (2005) What is FRBR? A conceptual model for the bibliographic universe. The Australian Library Journal,54 (1), 24–30. https://doi.org/10.1080/00049670.2005.10721710

Torres, D. (2017) *The Use Of Digital Technologies In The Conservation, Analysis And Dissemination Of Cultural Heritage.* AC/E Digital Culture Annual Report. Spain: Acción Cultural Española. [Online]. Available at: https://www.accioncultural.es/en/report2017 (Accessed: 5 April 2021)

Tringham, R. and Stevanovic, M. (ed.) (2012) Last House on the hill: BACH area reports from Catalhoyuk, Turkey. Los Angeles (CA): CotsenInstitute of Archaeology Press.

Tung, B. (2013) Excavations in the North Area, 2013. Çatalhöyük 2013 Archive Report Web publication.

Wagner G.A. (1998) Age determination of young rocks and artifacts. Physical and chemical clocks in Quaternary geology and archaeology. Springer, Berlin.

Wallis J., Borgman C., Mayernik, M. and Pepe, A. (2008) Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research. International Journal of Digital Curation, 1(3), pp. 114-126.

Wallis, J. C., Rolando, E., Borgman, C. L. (2013) If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS One, 8(7), pp. 1-17.

Wang, Z. (2013) Co-Curation: New Strategies, Roles, Services, and Opportunities for Libraries in the Post-Web Era and the Digital Media Context. Libri, 63(2), pp. 71-86.

Whitehead, C. (1989). Faceted Classification In The Art And Architecture Thesaurus. *Art Documentation: Journal of the Art Libraries Society of North America*, *8*(4), 175–177. http://www.jstor.org/stable/27948127

Weinberger, D. (2011) *Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room.* New York: Basic Books

Witt, M., Carlson, J., Scott Brandt, D. and Cragin, M. (2009) Constructing data curation profiles. International Journal of Digital Curation, 4(3), 93-103. Accessed September 7, 2021, from http://www.ijdc.net/index.php/ijdc/article/viewFile/137/165.

Wright, H. (2011) Seeing Triple: Archaeology, Field Drawing and the Semantic Web, PhD thesis.

Yakel, E. (2007) Digital curation. OCLC Systems & Services. 23. 335-340. 10.1108/10650750710831466.

Yakel, E. et al. (2011) 'Digital curation for digital natives', Journal of Education for Library and Information Science, 52 (1), 23-31.

APPENDIX

Construction Catalogue Records and Templates in BeArchaeo Database