

Journal of Field Archaeology

ISSN: 0093-4690 (Print) 2042-4582 (Online) Journal homepage: www.tandfonline.com/journals/yjfa20

Towards a Living Archive: Making Multi Layered Research Data and Knowledge Generation Transparent

Dominik Lukas, Claudia Engel & Camilla Mazzucato

To cite this article: Dominik Lukas, Claudia Engel & Camilla Mazzucato (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

To link to this article: https://doi.org/10.1080/00934690.2018.1516110

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 08 Nov 2018.

Submit your article to this journal

Article views: 2178

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=yjfa20

JOURNAL OF FIELD ARCHAEOLOGY
2018, VOL. 43, NO. S1, S19–S30
https://doi.org/10.1080/00934690.2018.1516110

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Towards a Living Archive: Making Multi Layered Research Data and Knowledge Generation Transparent

Dominik Lukas, Claudia Engel and Camilla Mazzucato

Stanford University

ABSTRACT

Çatalhöyük was first discovered as one of the earliest urban settlements in the late 1950s and excavated by James Mellaart between 1961 and 1965. The 9000-year-old town in central Turkey rapidly became famous internationally due to the large size and dense occupation of the settlement, as well as the spectacular wall paintings and other art that were uncovered inside the houses. Since 1993, under the direction of Ian Hodder, research at Çatalhöyük has pioneered a reflexive approach to archaeological practice, in which information is permanently open to reinterpretation by both scholars and the public. This approach acknowledges the mutual dependence of knowledge and the underlying research, to which end the Çatalhöyük Project decided to make its records available via the Web and to invite public comment since its onset. After 25 years of excavation, the project's digital assets now amount to close to 5TB, including formal textual and numeric records, freetext documents, audiovisual materials, and a comprehensive collection of spatial data. The reflexive method, or 'documentation of the documentation process,' adds a separate layer of information that specifies how data have been gathered, and facilitates critique, understanding, and the evolution of knowledge.

In this paper we lay out our vision of an interactive archive that provides access to the multi layered information contained in this massive amount of data and how web technological advances have been incorporated into the digital data management at Çatalhöyük. Ultimately, the goal is to support an interdisciplinary process of assembling data into arguments on the basis of multiple lines of evidence. The 'Living Archive' will enable intuitive engagement across the entire variety of research, making use of the rich reflexive information stored with the data. The results of new analyses can in turn be reintegrated with the already existing data. The application will use open standards so that the knowledge gathered at Çatalhöyük can be linked with other projects that follow similar publication procedures based on the semantic web approach.

KEYWORDS

Archaeological Information Systems; Digital Field recording; Semantic Modeling; Reflexive Method; Neolithic

Introduction

The idea of developing specific software applications to facilitate collaboration in archaeological research has existed since the inception of the earliest archaeological information systems. In 1973 Pryor formulated a list of key features that should define a fieldwork recording system in archaeology. While focusing on the applicability of the system in a maximum of fieldwork scenarios and the possibility of constant modification and extension of the stored information, Pryor underscored the need for a system that had to be easy to use and adapt to the nature of most archaeological projects. Most importantly, the systems had to accommodate a variety of users—ranging from those generating and curating the data to those "involved in archaeological research, both locally and further afield" (Pryor 1973: 8).

Early attempts to ease the process of data processing in archaeology go back to the early 1950s. Starting in 1980s we see constantly improving and cheaper hardware systems, which gave rise to the development of computational fieldwork infrastructures. Early systems like SYSLAT (Py 1997), initially developed for the excavations in Lattes (France), but then further adapted as SYSAND in Jovols (Maggiolo-Schettini et al. 1995) and SYASS in Fratte di Salerno (Italy) (Santoriello and Scelza 2000), speak to the effort to develop standardized approaches to data recording.

It is in this technological context that research at the Neolithic site of Çatalhöyük began. The 9000-year-old settlement in central Turkey is among the earliest urban settlements ever discovered and opens a crucial window into an era that is arguably the most significant social shift in human history: the transition from hunter-gathering to sedentary farming societies. The site was first discovered in the late 1950s and excavated by James Mellaart between 1961 and 1965. Ian Hodder continued and directed the excavations between 1993 and summer 2017 with an international team of archaeologists. Methodologically, research at Çatalhöyük has pioneered a reflexive approach to archaeological practice, in which information is permanently open to reinterpretation by both scholars and the public. Since its outset the Çatalhöyük Research Project stood out as an early adopter of latest innovations in information technology and digital recording solutions, which is reflected in the archaeological information system in use on site including a relational database system that allows for a separation of frontend- and backend- architecture, as well as dedicated and highly specialized tools for data entry, storage, and data retrieval (Engel and Grossner 2014: 14).

CONTACT Dominik Lukas dominik.lukas@stanford.edu Department of Anthropology, Stanford University, Main Quad, Building 50, 450 Serra Mall, Stanford CA.

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Today, after 25 years of research, the system has grown out of the initial design, with numerous extensions and modifications, new functionalities, data models, and types of data that followed the implementation of new methods on site. The complexity of a field recording system tends to increase over time, which at first glance is reflected simply in the sheer size and the diversity of data recorded. The Çatalhöyük central database hosts textual and numeric records of the excavation and contains more than 1000 related tables and 8.3 million data rows. The complexity of the system is increased further by the integration of data gathered with the reflexive approach, including contextual information concerning the process of research.

With the end of fieldwork on site and the end of processing of the primary data insight, the Çatalhöyük Research Project is now focusing on the future of the information recorded during more than two decades of fieldwork. It is in this context that we ask: Is it possible to make research at Çatalhöyük transparent for future researchers, preserving data and providing access not only to content but also to the processes that led to the production of that content, and thus provide a more coherent reflection of archaeological knowledge at specific moments in time?

Data Complexity at Çatalhöyük

The process of data production

In 1964 James Mellaart observed that "at Çatal Höyük there is no shortcut to obtaining quick results and the only course for action is that of the gradual stripping off layer after layer" (1964: 39). These words well represent the nature of the site. Çatalhöyük is an extremely complex site whose archaeology and deep stratigraphy resist "short cuts." Similarly, the many years of intensive archaeological research conducted at the site have produced a diverse "stratification" of numerous datasets. The Çatalhöyük Research Project is one of the most comprehensive and complex archaeological projects in contemporary archaeology due to its long-term commitment, the amount of people and resources involved, and the methodologies applied (Farid 2015; Marciniak 2015). During the 25 years of excavation, the Çatalhöyük Research Project has produced a 'data deluge' (Faniel and Zimmerman 2011; Bevan 2015) that challenges storage practices, analysis, and visualization.

During the 2014 field season, Hodder (2015) counted 160 team members divided between field archaeologists and laboratory specialists encompassing 36 different specializations. Over the years, despite changing team members, new technologies, updated research agendas, and revised interpretations, the commitment to a reflexive and multivocal ethos has persisted. This reflexive theoretical framework remains the most central and distinctive feature of the Çatalhöyük Research Project (Farid 2015; Hodder 1997, 2000, 2003, 2008; Hodder et al. 2015).

Since the reopening of the site in 1993, Ian Hodder strongly committed to bringing "critical reflexivity" to the "laboratory and the excavation trench" and to situate archaeological knowledge within a "network of flow" (Hodder 2000: 48). The archaeological research at the site was envisioned as transparent and open, available to local communities and the general public and to the mediation of a multitude of different voices (multivocality). Furthermore, multivocality was to be achieved at the 'trowel's edge' where the interpretation of the record begins.

Central to the reflexive approach is the idea of the substantial faultiness of compartmentalized archaeological practices divided between an 'objective' recording phase in the field and a following interpretation phase. Interpretation commences from the very beginning of archaeological investigations within the "trench" and "at the trowel's edge" (Hodder 1997, 2000, 2003: 58). While archaeologists dig, they interpret their deposits in light of their own experience. It is therefore of crucial importance that they can access all the information they require and that knowledge is free to flow between team members. The presence at the site of the main specialists' labs, their involvement in all phases of excavation, and the complex feedback loop between specialists and lab members are part of the reflexive approach to archaeological research. Adopting a critically reflexive approach involves actively breaking down barriers and fostering collaboration and data integration at each stage of the research process (Berggren and Nilsson 2014). A reflexive framework furthermore entails the recognition of one's positionality and the way this affects knowledge creation throughout the entire course of research (Hodder 2003). Team members are encouraged to be transparent about their own interpretations and a key part of the research process involves finding time for and incorporating tools for discussions about the process of interpretation (Berggren and Hodder 2003; Hodder 2003).

A series of strategies were put in place in the early 1990s to achieve an efficient reflexive workflow on site, which are summarized by Hodder in a 12-point plan (see Hodder 2000). In the information system, specific areas for data entry were created where excavators and laboratory teams had the opportunity to provide explicit accounts of their interpretation processes and where they could furnish details on alternative interpretations, uncertainties, and competing ideas. Some of these locations were free-text entries in the database, like the diaries or the unit discussion and interpretation fields in the excavation and human remains databases. Similarly, video recording on site of both priority tours and excavators' weekly summaries were employed as tools for boosting cooperation and sharing of information between different components of the project (e.g. Chrysanthi et al. 2016). More recently, digital sketches drawn on tablet computers aimed at the creation of a daily visual excavation diary.

Recent publications have assessed and re-evaluated Hodder's original 12-points strategy after years of implementation, providing an account of the revisions that were made to the workflows to better adapt to the different phases of the Çatalhöyük Research Project (see Berggren and Nilsson 2014; Berggren et al. 2015; Farid 2015). The practice of data creation through the years was constantly subjected to adjustment and fine-tuning of research goals, the application of new technologies, and the modification of interpretations (Berggren et al. 2015; Farid 2015). This practice has produced an extremely rich but complex dataset and a stratification of different typologies of data.

Together with the data generated with this reflexive approach, a core set of more standardized datasets were routinely collected and recorded by excavators and specialists. The implementation of a centralized database and the creation of standardized categories that could be more easily queried were part of this effort (see Farid and Hodder

2014). While most archaeological projects record fieldwork data in similar ways, the reflexive approach added another layer of information to the data in which the how and why of the recording of each set of data was stored. The evolution of the system therefore is framed by a steady discourse of research strategies and technological development, where neither aspect had priority over the other.

Furthermore, the tension between the need of a standardized recording system as a prerequisite for obtaining a highly accessible record and the aim of fostering fluidity and flexibility on site has been difficult to resolve (Berggren and Nilsson 2014; Farid 2015). Since 2009, the adoption of an intra-site GIS increased the need for more standardized datasets. The choice of the geographical information to be used was restrained by the existing technological infrastructure at the site and since the beginning the intra-site GIS relied heavily on the pre-existing relational database (Taylor et al. 2018). ESRI technology and a personal GeoDatabase were chosen because they effectively and efficiently integrate with the central relational MS-SQL database through a Microsoft Access database. From the outset, the intra-site GIS was conceived to support the reflexive theoretical framework at Çatalhöyük and was designed to act as a flexible spatial link between the work conducted in the field and the work carried out by the specialists, between the graphic and spatial information produced by excavators and the textual information stored in the central database (Berggren et al. 2015; Taylor et al. 2018). The intra-site GIS was developed through collaborative efforts, and while a core set of fields extracted from the central excavation database were permanently stored in the geodatabase, it was decided to integrate specialist information when needed for analytical purposes.

In 2011, when the Çatalhöyük Research Project shifted to a fully digital recording workflow, the GIS served as the primary recording tool used on tablets in the field by archaeologists (see Taylor et al. 2018 for the evaluation and a detailed description of the digital workflow at Çatalhöyük).

The development of digital field recording on site

The design of the current Çatalhöyük database infrastructure dates back to 2004, when in order to share data among all researchers the various databases and spreadsheets that lived in isolation on the team members' personal computers were migrated to a centralized system (Ridge and May 2004; Ridge et al. 2005; Ridge and Jones 2006; Jones 2012a).

This centralization required a change in the recording workflow. To avoid recording of artifacts through the filter of a specialist's eye, who might seek to fit it into a preconceived model, initial entries about basic measurements, descriptions, and classifications, were made by the excavator. Next, specialist teams began recording by entering a unit description and team-specific, basic information. Recording then branched off and became more detailed according to each specialist team's more extensive measurements, analyses, and interpretations.

Each team accessed the database through a distinctive interface for data entry, tailored and continuously adjusted to the particular needs of their respective specialism. Members of the various teams could record very different data about the same object, but were not exposed to interfaces onto the same base excavation data unrelated to their specialism (Ridge et al. 2005).

With the centralization of the data the need arose for standardized vocabularies to be used by all teams. Such sharing of common codes between teams could facilitate a 'common language' and create the foundation for more powerful cross discipline analysis (Ridge et al. 2005). The Finds register developed at that time would be instrumental in arriving at such a common language: "The concept of the finds register is to provide a tracking tool for objects from site (recorded in the excavation database with the excavators interpretation) through the finds office (with the finds officers interpretation) and on with the number allowing tracking to the various lab teams for their interpretation" (Jones 2012b: 13). Links between the Finds and Specialist Tables were created, allowing simultaneous viewing of corresponding records across specialist teams as well as tracing artifacts through the stages of their study during which the interpretation of the same object may differ.

A centralized database also afforded the ability to generate queries across specializations, teams, and areas as dynamically updated tables ('views') that respond to changes in the content of the tables the queries are based on. These queries allow users to group materially unrelated artifacts by interpretative category and run customized cross-table queries tailored to specialist requirements. Finally, centralization of the excavation and specialist data in a database server also allowed direct access to its content from other databases or applications, for example GIS.

In the years following this major restructuring, the database underwent continuous development following the changing recording requirements motivated by new team leadership, changing recording priorities, and new analysis practices. However, none of these represented any radical changes, so the current system still reflects fundamentally the design principles introduced in 2004.

Post-Excavation Research and Preservation

With the end of fieldwork by the Çatalhöyük Research Project in 2017, the project started the process of preparing the data for long-term preservation and reuse. As a consequence of the long duration of the project its digital archive faces a peculiar situation: While at the beginning of the Çatalhöyük project the question of digital preservation did not receive much attention—if any at all; that has changed dramatically today. A first attempt to publish data with long term open access stems from the year 2006, when the research database was opened for public access via the project's website. The publication of "Last House on the Hill" in 2012 (Tringham and Stefanović 2012) based on work of the BACH team (Berkeley Archaeologists Çatalhöyük) is a remarkable attempt to go beyond a traditional paper publication, making parts of the research data available online (c.f. http://lasthouseonthehill.org/). It is also a first attempt to map the BACH fieldwork data to the EH-CRM (an extension of the conceptual reference model CIDOC-CRM) (Ashley et al. 2011). However, even after decades of discussion about the appropriate measures for long-term storage and reuse of archaeological data (Faniel and Zimmerman 2011; Faniel et al. 2013; for a critical reflection see Atici et al. 2013), available solutions are not yet entirely satisfactory. Data are either handed over to agencies for long-term storage or—and unfortunately this is still the case—they are left unattended without any reuse strategy. If the data are accessible, their historical

contextualization is often difficult and the integration of different data types cumbersome. While Engel and Grossner (2014: 22) discuss a solution to allow for subsequent addition of data to a research archive with the use of 'collections' of data including annotations, it is generally not possible to add new data resulting from new analyses.

The static digital archive

Digital archives for research data, like the Archaeology Data Service (ADS) in the United Kingdom (University of York) or the Digital Archaeological Record in the United States (tDAR, University of Arizona), cater to the need to preserve primary data collections of archaeological projects for the future, storing the datasets on university servers to keep the information accessible (Richards 2009; McManamon et al. 2010; Kansa 2012: 505; Faniel et al. 2013). The primary purpose of these archives is preservation and easy retrieval of singular data objects and the mentioned archives have opened vast amounts of formerly closed data for future research. Data are stored in formats meant to be readable for a foreseeable future. Most importantly, complex relational databases are often stored as a set of individual spreadsheets, which contain the information of each table, but the relationality of the data is mostly lost or only recorded in the documentation of the project in the form of Entity-Relationship diagrams (e.g. concerning storage of databases at ADS: Guidelines for Depositors - Version 3.0 September 2015 - http://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#Databases,%20Spreadsheets%20and%20Statistics). As a result, while data is technically available, its reuse is difficult. Another problem, that is particularly relevant for the Çatalhöyük Research Project is that any information generated during the continuous interpretation and reinterpretation of data—the data gathered via the reflexive method on site—is not part of a standardized process for long-term preservation.

(Re-) connecting interpretation

The difficulty of transparently connecting research data with their scientific interpretation is not a problem restricted to the field of archaeology. The requirement to make research reproducible has become widespread in the sciences and platforms like CodaLab (https://worksheets.codalab.org/) are increasingly common. Research publications are interwoven with the actual data and the code, which is used to process them and create the scientific output, and the steps from the experiment design to the conclusion are traceable with technological means. These approaches are also known as 'GitHub for the sciences' (von Mühlen 2014)—applications allowing for the versioning and computer code documentation of systems for research. As an example in archaeology, OpenContext made use of GitHub to further transparent publication of software and data (Kansa 2012: 510). Documenting research design and the flow of interpretation is much less common in the humanities. An initial attempt from within archaeological research is the Patina project (Patina of Notes). As part of the development of the XEROS system (Jewell et al. 2012), Patina was designed to include the provenance of interpretation—built from capturing several interpretive events, edition of entities and by applying non-destructive removal of states of knowledge (https://tina.ecs.soton.ac.uk/djangopon/forum/). The system no longer seems in use and does not appear active since over 3 years.

Parallel to the development of computerized field recording systems in the 1970s and 1980s, research was carried out not only concerning data capture following different fieldwork approaches, but also to explore and model the research processes themselves. These expert systems and knowledge bases, which implement archaeological procedures of scientific inference and allow for computer-supported research, required a representation of archaeological knowledge in a machine-interpretable way.

An early attempt to find a solution to this problem used the logicist approach (Roux and Courty 2013; Roux and Aussenac 2013), which held that archaeological theories can be represented as computational structures with two constituents: a database/set of declarative propositions, relating to objects/phenomena in the outer world, and an inference tree made up of formulas (cf. Dallas 2016). These formulas were of the form like "(IF) p (THEN) q," involving declarative propositions. These formulas were to express the logico-discursive structure of archaeological theories from an epistemological viewpoint. That would allow for the transparency of implicit assumptions, which had not already been formalized and recorded with the information system used (Orlandi 2004). Formulas also described association rules like "If pottery(X) and type(X,bf) then chronology(X, 1570)" (Puyol-Gruart 1999: 21), to describe the theoretical assumption that the co-occurrence of a specific pottery type in an archaeological context was an indication for the entire context to be of a certain dating.

Approaches to capture theoretical assumptions and interpretations were embedded in a consistent, unified classification and interpretation (Orlandi 2004). The Arkeotek Project (Gardin and Roux 2004) is the most recent example for the interoperability of knowledge bases and the idea of accompanying scholarly publications with logicist diagrams representing the ebb and flow of arguments was implemented in one of the project's recent workshops (Roux and Courty 2013). The logicist approach even went so far as to propose an artificial language for archaeology (e.g. Gardin 2002) and to describe archaeological research formally—having its philosophical roots in the work of the Vienna circle and later approaches by Toulmin (1958) and Grize (1974).

Digital workflows have become so integral to the research process that they now shape our knowledge creation processes (Chrysanthi et al. 2012; Kintigh et al. 2015). Following Dallas' (2016) 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: 322). Researchers at Çatalhöyük have pointed into the same direction and reflected on the complex parameters that influence the evolution of ideas within research teams on long-term projects (Hodder 2015; Mickel 2016).

While the idea to capture knowledge and scientific meaning of data had been at the center of expert systems in the beginning of computer applications in archaeology, a different approach to achieve similar objectives that has gained much momentum today is the use of ontologies and the semantic web.

Semantics and the linked open data approach

While traditional long-term archives rely on the possible standardization of data types, 'Linked Open Data'

approaches, like the OpenContext (Kansa 2012: 510) platform, in contrast, respond to the problem of identifying the informational structure of the datasets by assigning commonly understood attributes and allowing to link between similar information. In the 'Linked Open Data' approach a specific type of information—like a dataset describing a faunal bone—can be identified in the data cloud and can be accessed with a specific web-address. It is possible to link similar information residing in different archives by using the same publication approach. However, while this improves data interaction, the role of the data in a particular research context remains difficult to capture.

Attempts to use semantic web technologies to code archaeological knowledge in machine interpretable way have existed since the 2000s (Isaksen 2011: 17), notably the work based on the CIDOC CRM (Dörr 2003; LeBoeuf et al. 2017) semantic model. This reference model—a higher-level ontology (Hohmann and Fichtner 2015: 117)—was designed to create a general framework for integrating different sets of data. While this model was initially developed to deal with objects in museums, its usefulness to deal with data deriving from archaeological excavations was almost immediately acknowledged. In the last decade, this ontology has seen widespread use and approaches as to its systematic implementation were discussed (e.g. Cripps et al. 2004). Domain ontologies like the English Heritage CRM have been developed to further specify general archaeological concepts defined in the CIDOC-CRM (Doerr 2003), so archaeological practice could be organized alongside the specifications of these models and the data collections could be interlinkable. In the context of a continuing discussion concerning the need to model inference in archaeology (see Kintigh et al. 2015) further extensions to the CIDOC-CRM model, CRMDig and CRMInf (Doerr and Theodoridou 2011; Stead and Doerr 2015) cater to a growing interest in capturing scientific argumentation.

In this respect, the mentioned 'Linked Open Data' has to be understood as a merely pragmatic approach to the coding of knowledge, which has also gained popularity because of the difficulty to implement ontological conceptualizations. In contrast to complex ontologies like the CIDOC-CRM, other ontologies have proven to be more easily applicable (such as Open Annotation) (Isaksen et al. 2014: 198). OpenContext follows a similar pragmatic approach: ontologies are implemented depending on the specific context of the information (e.g. Encyclopedia of Life and Uberon for faunal remains) but are not integrated with a central reference model like the CIDOC-CRM (Faniel et al. 2013: 302). A first implementation for the Çatalhöyük data had been developed along similar guidelines (Engel and Grossner 2015). However, the Çatalhöyük digital material requires that the meta-information deriving from the application of the reflexive method explicitly indicates meaning. We need to extend on these existing models and implement a web-based platform to integrate these models and to capture that information.

'The Living Archive' Approach to Scientific Knowledge

Theoretical archaeology and the history of computer applications in archaeology show that comprehensive solutions to develop either a unified fieldwork database or a final expert system to envelop the entirety of archaeological research are futile. Neither a unified fieldwork database exists nor any definitive expert system is currently available. In fact, because of theoretical differences in archaeological projects worldwide, the production of such an all-encompassing system does not seem possible. This is true as well for any approaches to model an all-encompassing—valid in breadth and detail—ontology for archaeology. In consequence, the digital archive of Çatalhöyük has to be understood in its entirety as a knowledge base with an incompletely expressed ontology. The archive does effectively contain knowledge from the years of research on site, but it has never been exposed in a transparent and complete manner. In this sense, the knowledge is hidden in the relational table-structure, the application logic and a pool of singular data objects and therefore the archive requires a consistent formalization of its information structure. If archaeology as a field of research cannot be adequately formalized as whole, the knowledge of archaeology as it is stored in databases, lies in the approach to model distinct archaeological information systems, underscoring the differences of projects and the explicit definition of each semantic model in comparison to other models.

The objective of the 'Living Archive' is the combination of the mentioned approaches in semantic modeling and developing a consistent semantic-web platform. Research at Çatalhöyük as an interdisciplinary project based on reflexivity and multivocality needs to conceive a technical formalization, which reflects these aspects. In consideration of this complexity, distinct parts of the stored knowledge need to be modeled as domain and task ontologies (Hohmann and Fichtner 2015: 117). These parts hereby reflect disciplinary boundaries like faunal remains and groundstone, as well as changing research strategies, which impacted the way research has been recorded. In contrast to the aforementioned Arkeotek project, the 'Living Archives' primary attempt is the identification of assumptions, which relate to reasoning backgrounds, in the digital archive itself whereas Arkeotek reconnected the database to actual written research literature (articles, books). This aspect is only of secondary importance for the 'Living Archive.'

Çatalhöyük archive transformation

To transform the digital archive from a system primarily based on a central relational database to an ontology driven application, it is necessary to understand the implications of how the data are embedded in the existing environment and how this impacts the data themselves. Each part of the archive has to be examined as to the semantic value of these implications. We will discuss each major component of the Çatalhöyük archives in turn below.

RELATIONAL DATABASES

Ontology as it is conceived today, is based on the principles of the object-oriented data model. In consequence the task of transformation of relationally stored data into a system exposing data ontologically is in principle a schema conversion (Cripps and May 2004: 60). It is even possible to process the data model with semi-automated procedures (e.g. Chujai et al. 2014 or Telnarova 2010). Semi-automated processing is faster, but less detailed so it runs the risk of losing track of the very meaning of

the data for the research process we are after. With out aim of retrieving a maximum of information on the semantic value of the data, these approaches would then make the intended gain of transparency about the meaning of the different implications difficult to assess. In consequence, the relational model of Çatalhöyük data is transformed in a conservative reverse engineering approach (see Fong 2015: 159), documenting each relationship (primary, secondary keys) in the data model, which allows assessing the implicit scientific value of the data within the overall knowledge structure.

FRONTEND APPLICATIONS

In the Çatalhöyük information system a certain number of rules for data entry, processing, and visualization are implemented in the application logic of the frontend programs, through which excavators and researchers interact with the database. These rules constrain the information stored within the relational tables, but are not made explicit by the reverse engineering of the relational databases. The frontend application itself needs to be reverse engineered as well (cf. Fong 2015: 181).

As the teams' research goals, interpretations, and practices evolve, software requirements change—usually during the excavation season, leaving only a short time for implementation. Over the course of the project these seasonal adjustments have left traces not only in the feature set and the functionality of the system, but also in the source code of the application. In a recent paper, Lukas and Engel (2017) used Natural Language Processing and qualitative methods in an attempt to determine if and how the evolution of technologies and research practices is reflected in the history of the computer code behind the Çatalhöyük frontend applications. Preliminary results indicate that step changes between different historical versions of the code can indeed be identified and tied to transformative moments in research practices (FIGURE 1). The rules retrieved from the process of reverse engineering of the application then become part of the overall ontology.

FILE STORAGE

Additional files are stored in a heterogeneous folder structure containing a variety of different data and file types (hard disks, collections of CD-ROMs and DVDs, cloud-storage). These data objects are not directly linked with a relational database. Information about the distinct data sources, which are largely isolated in the folder structure of the file storage systems is used to determine how these data can be cross referenced with the database. Spreadsheets, File Maker, or MS Access standalone databases incorporate implicit scientific knowledge in similar ways like the central database. Images, audio, or CAD drawings do not only contain pixels in a Cartesian coordinate system, sound in temporal sequences or vector lines with spatial coordinates, but also standardized metadata that refer to archaeological concepts defined within the Çatalhöyük Research Project.

Semantic model of Çatalhöyük research

The English Heritage CRM (an extension of the CIDOC-CRM) used by the BACH team (Tringham and Stefanovic 2012) was based on the definition of the reference model in 2011 (Ashley et al. 2011). The proposed ontology deliberately reflected a simplified model of knowledge, stressing the project's intention to have an easily understandable structure. We revisit this approach to account for the complexity of the Çatalhöyük data as described above, and to allow for a more comprehensive integration of information. In order to show the set of assumptions necessary even for the recording of a 'simple' context we will discuss the modeling of the specific approach to the single context archaeology applied at Çatalhöyük.

In the English Heritage CRM (Ontological Model of Centre for Archaeology Information Domain V.9) the central entity of the single context excavation methodology is defined as a 'context.' This entity is further described as "a cut is defined by a three-dimensional plane, a deposit or structure is described by a three-dimensional volume" (CRM-EH 2014). Contexts are further defined by their property relations like 'falls-within' relating to places ('Group' and 'SiteSubDivision'), 'is identified

Assigned Topic over Changes by Year

2004	2005	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Navigation	Navigation	Navigation		Navigation				Navigation		Navigation	Navigation
		Validation	Validation	Validation	Validation	Validation					Validation
Human Remains	Human Remains	Human Remains	Human Remains	Human Remains						Human Remains	Human Remains
		User Interaction	User Interaction	User Interaction	User Interaction	User Interaction	User Interaction	User Interaction	User Interaction	User Interaction	User Interaction
Space/Larger Contexts		Space/Larger Contexts	Space/Larger Contexts	Space/Larger Contexts	Space/Larger Contexts	Space/Larger Contexts				Space/Larger Contexts	Space/Larger Contexts
				Interpretation	Interpretation	Interpretation		Interpretation		Interpretation	Interpretation
Building		Building		Building	Building					Building	Building
		Feature	Feature	Feature	Feature	Feature	Feature		Feature		Feature
		Image Processing	Image Processing	Image Processing		Image Processing	Image Processing		Image Processing		Image Processing
		Image Visualization	Image Visualization	Image Visualization							Image Visualization
Context Type/Search		Context Type/Search	Context Type/Search	Context Type/Search							

Figure 1 Example from the application of Natural Language Processing on programing code. 11 Topics were extracted from the code base for the Excavation database frontend and their occurrence mapped by year. A mention of a topic for a particular year indicates that the code base for this year was modified for this particular topic (Lukas and Engel 2017).

by' relating to spatial coordinates ('ContextDepiction') or 'occupied' relating to physical stuff ('ContextStuff').

Although in principle applying the same excavation method the Çatalhöyük Research Project has modified the original approach in some crucial aspects. Figure 2 represents a small part of the entire model. Here, Units are the basic recording element, which ideally represents a single identifiable depositional event, although in practice a unit may also represent multiple depositional events if the process of fieldwork requires it. Units can be divided into a number of general categories (e.g. cut, fill, layer) that in turn form the basis of a nested hierarchical system of lower level categories (layers are subdivided in construction-layers, midden-layers, etc.) (Hodder et al. 2007: 13). Units have a spatial connotation and fall within Spaces. Considering the required compatibility of the Çatalhöyük model, these Spaces could prove equivalent to EH-CRM-'SiteSubDivisions,' which themselves fall within larger spatial entities like Buildings and Areas. At Çatalhöyük Units are also related with Features. As there is no Feature in the EH-CRM, it is necessary to understand if there are alternative concepts in the model, which could contain the respective data.

On the one hand, Features at Çatalhöyük are groups of related Units, which need to be described as a whole, although not all units need to belong to features. In conclusion, a Feature—defined as a group of Units—could be understood as a specific archaeological entity to which a historic functionality is ascribed. On the other hand, Features divides the site into smaller areas. As such Features contain Units, but are themselves contained in Spaces.

Both possibilities to conceive of Features seem at first glance reflected in the EH-CRM. Here, a feature could be defined as a 'Group' (of units) related by context with a 'falls within,' or it could be understood as 'SiteSubDivision,' which again is related to 'context' by a 'falls within.'

This solution is problematic for several reasons. A Feature can be in a specific spatial relation with other features, while the units constituting the feature do not necessarily share these relations. In EH-CRM only a 'context' can border with another 'context,' like a Unit can border with another unit in Çatalhöyük, but a Feature—modeled either as a 'Group' of Units or as a 'SiteSubDivision'—cannot border with another Feature. Moreover, a Çatalhöyük Feature has a set of spatial relations that can go even beyond the relationships ascribed to a unit.

Modeling Çatalhöyük Features as 'context' would require a context to potentially have a 'falls within' relationship with another context, as Çatalhöyük Units—already modeled as 'contexts'—would fall into Features and therefore other contexts. In a proper single context excavation methodology, the concept of a context being composed of other contexts is not possible—it would no longer be a single context. Defining Çatalhöyük Units as 'context' and Features as 'Group' would also not comply with EH-CRM, according to which groups do not relate spatially with each other. Moreover, EH-CRM has no concept that allows for a group of contexts having its own relation to a 'SiteSubDivision.' The Çatalhöyük Feature relates to a space in ways that goes beyond the relations that units have within the space. Again, a Feature cannot be defined as a group of contexts. In consequence, with its definition of units and features, and features being a special type of context, Çatalhöyük deviates from the standard single context method used in contract archaeology in the UK. Further contradictions arise on the conceptualization of spatial coordinates as ContextDepictions. (Spatial coordinates have only recently been modeled consistently through the development of CRMgeo (Doerr et al. 2013) and prior approaches to spatial representation in the CIDOC-CRM—as one attempt is visible in the EH-CRM—need now to be seen as deprecated.

A more recent example of a semantic model for archaeology is the CIDOC-CRM extension CRMarchaeo (Ronzino 2017). A context is now understood as a 'stratigraphical unit'—a concept terminologically closer to the Unit used at

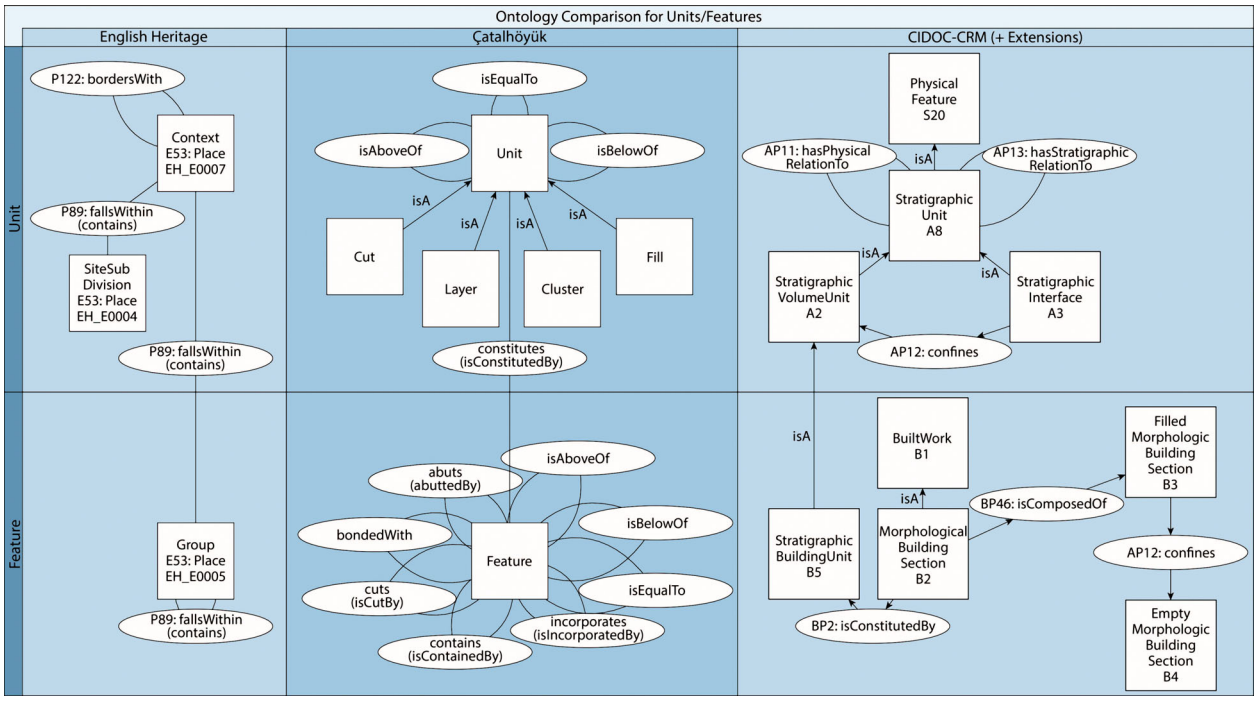


Figure 2 Mapping of Çatalhöyük concepts for different contexts (unit, feature) to the English Heritage (EH-CRM) and CIDOC Conceptual Reference Models (CIDOC-CRM, including presently available extensions).

Çatalhöyük. The stratigraphical unit is conceived as a physical feature and can be defined in detail by its subclasses 'stratigraphic volume unit' and 'stratigraphical interface'. This approach corresponds to the separation between 'fill' and 'cut' at Çatalhöyük. Still, the Feature problem remains: CRMarchaeo does not add an entity, which would allow for grouping contexts and capturing features at Çatalhöyük.

Another extension of the CIDOC-CRM, the CRMba (Ronzino et al. 2016)—a model developed for the description of buildings in archaeology—introduces concepts to further define aggregates of stratigraphical units, which convey functional aspects of these units as they are constitutive for a building in its entirety. The introduction of the 'stratigraphic building unit' allows for the grouping of contexts into 'morphological building sections,' which on their end constitute built ensembles of building sections, defining the actual 'built work' like a building. The 'morphological building sections' allow also for the attribution of properties to ascribe spatial relationships to other building sections. In consequence, the spatial description of Çatalhöyük features therefore can at least partially be achieved by the use of explicit properties 'is bound with,' 'is adjacent to,' 'cuts' and 'is filled by.' Thus, while CIDOC-CRM and its extensions can serve as semantic model to large extent of Çatalhöyük data, problems remain.

All the aforementioned extensions to the CIDOC-CRM have now been aggregated in the Ariadne Reference Model (Dörr et al. 2015) and as such allow for a detailed modeling of archaeological data. As methodological intricacies and specific levels of detail achieved at Çatalhöyük cannot be fully mapped to these models, we have begun the development of a specific semantic model of Çatalhöyük Research Data. It makes use of compatible concepts and extends the model where necessary. The most direct benefit of our approach is the opportunity to define entities as being equivalent to other entities in existing ontologies, allowing for further integration of the Çatalhöyük data where a direct overlap of types of information is possible.

Secondly, it is possible to turn to and integrate additional existing disciplinary and interdisciplinary ontologies, which reflect the spectrum of interdisciplinary research at Çatalhöyük. One example is the Foundational Model of Anatomy (FMA) describing anatomical concepts and allowing to link aspects of human remains found on site. In order to support a maximum of interconnectivity following the Linked Open Data approach, mapping to further models like Encyclopedia of Life and Uberon, as they are used within OpenContext (Faniel et al. 2013 302), is planned. This will also allow to tap into other Linked Open Data platforms such as Pelagios (Simon et al. 2016).

25 years of data recording and paradigm shifts

During 25 years most parts of the relational database underwent various modifications, often going along with a significant shift in the research approach and the way data was structured and stored before and after these changes, which led to inconsistencies in the datasets over time. Often only a few data fields, like a primary identifier, standard measurements, and the location of objects in the depot were migrated to a new data model, whereas other fields were substituted with completely new categories following the shifted paradigm.

Evidently, when a data query extends over a 'paradigm shift,' differences in the research methods before and after the shift have to be considered. Consequently, considering the entire data set for the whole runtime of the research project, all methodological frameworks that were applied during that time have to be compared, harmonized and formalized to make research reproducible. For example, in 2012, the faunal remains lab implemented a new recording methodology, which resulted in the addition of a set of columns to some tables ('structural changes') and the modification of tables containing standardized vocabularies ('changes of content'). The vocabularies (lists of values, code tables, etc.) were modified in one of two possible ways: either by adding new values and preserving already existing values, or by substituting values with a completely new set.

The set of tables for faunal remains data at Çatalhöyük is highly normalized and most values in the data tables are based on standardized vocabularies. Even an extensive modification of the recording system does not necessarily result in the changes of the table structure, but rather in changing values of the vocabularies. In 2012 structural changes were applied to two tables: 'Faunal Unit Description'—containing information on the entire Unit from which faunal remains were retrieved—and 'Basic Faunal Data'—containing basic data about the unique faunal bones retrieved in a unit. Modifications included the removal of numerical columns from the 'Basic Faunal Data,' intended for the recording of measurements besides the standard method of measuring the bones in predefined zones, and the removal of two columns from 'Faunal Unit Description,' containing data on the quantification of specimens whose species could not be determined.

These modifications can be captured by the addition of extensions to the standard semantic reference model. The removed measurements are defined as measurement classes related to the respective bone class (e.g. Femur) by means of an object property 'hasExtensionInZone.' Similarly, the amount of bones from unspecified species in 'Faunal Unit Description' can be mapped to a data property 'containsUndeterminedBones' containing the additional values.

In contrast to structural modifications, changes to content are made either through the addition of new values by retaining all previous values, or by the complete transformation of information. During the changes brought to the faunal remains database in 2012, both types of modifications were implemented in a set of standardized vocabularies. For example, the original vocabulary for different types of heat infliction on bones contained the values: 'not burned,' 'calcined,' 'charred: burned and/or carbonized,' 'see comments,' and 'heat-altered.' These values were then extended to contain the values 'Calcined,' 'Partly calcined,' 'Calcined + carbonized,' 'Carbonized,' 'Partly carbonized,' 'Carbonized + burnt,' 'Burnt,' 'Partly burnt,' and 'Slightly burnt.' The values in the first list could be retained, as they were not contradictory to the new system of recording implemented. The semantic value of 'calcined' was mapped to 'Calcined,' 'charred: burned and/or carbonized' to 'Burnt.' The value 'heat-altered' was deprecated as it had been substituted by more specific values describing the exact heat alteration. In addition, two values were eliminated as they only stated implicit information: 'not burned' was eliminated and substituted by 'NULL' values from 2012 forward, while 'see comments' only pointed to a description field that contained

information in the case this field actually was ever selected or not.

Different from traditional approaches that strive to harmonize the data values, 'The Living Archive' conserves the history of content modification and describes it by means of provenance modeling. It allows users to reconnect the data with the method of their creation—supporting a reflexive approach. By making theoretical assumptions underlying the structures of data explicit and identifying discontinuities and contradictions in the datasets, the stored knowledge of the Çatalhöyük Research Project will not only be available as a static archive for the long-term, but the modalities of data generation will remain transparent.

Development of the Çatalhöyük Living Archive

Because 'The Living Archive' application is being developed while the primary database and the digital archive are still in use and data is still continuously added, we created a software architecture that allows for as much dynamical integration of the original data store with the semantic database already in development as possible. While it is most common to ingest a final set of data through mapping into a triple store according to a semantic model—mostly defined within the OWL-language, the need for more dynamism with our approach forced us to experiment with dynamic mapping procedures. In our case this is provided via OnTop (Calvanese et al. 2015), which resides in between the relational database (MS SQL Server) and the object oriented triple store (rdf4j) (FIGURE 3). OnTop allows for the dynamic integration of the actual relational data into Protégé (Musen 2015), a software for the creation of ontologies, which is also used in the development of 'The Living Archive.' The JAVA-based triple store rdf4j is run on an Apache Tomcat Server. On the one hand this allows for direct querying of the data with the use of SPARQL (REST-API), on the other hand the API provided by rdf4j can be directly used within 'The Living Archive' application.

The user environment is developed in PHP by making use of the Symfony framework. Semantic data is integrated into the application via the Nemrod-interface, which allows for direct usage of RDF-data retrieved from the rdf4j triple-store. Data can be handled within the Symfony frameworks with the same procedures as with relational databases within

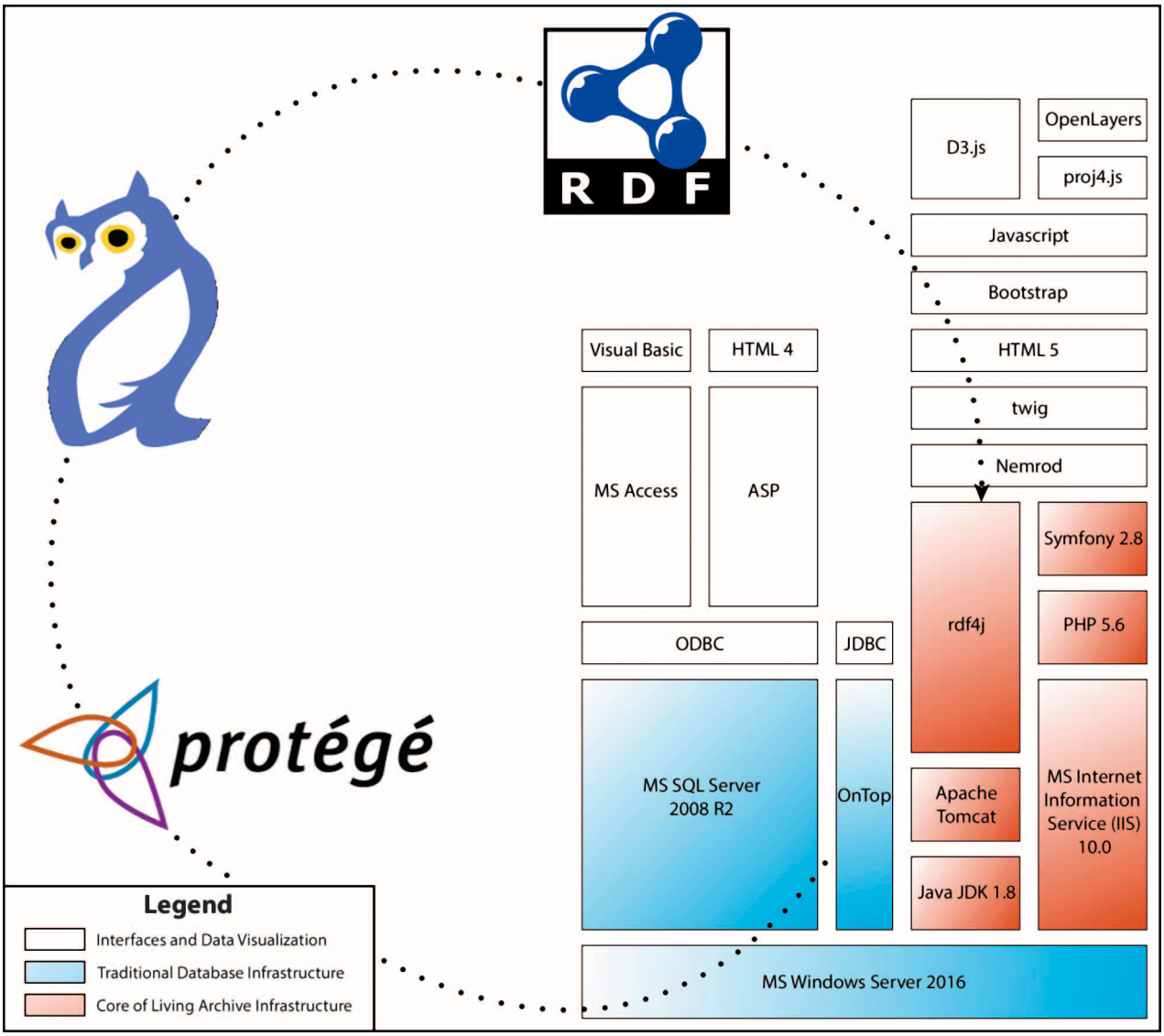


Figure 3 Transformation of data from the original relational database into semantic triples. Protégé (with the plugin OnTop) is used to develop the ontology (owl) and to export the semantic triples in rdf/xml-format for ingestion into the rdf4j triple store. The RDF logo was originally designed by Bill Schwappacher. The OWL logo and the RDF logo are both denoted to the W3C (WorldWideWeb Consortium). Contents from the W3C are licensed under the W3C-Document license https://www.w3.org/Consortium/Legal/2002/copyright-documents-20021231

standard web-applications. Data output is provided using PHP as preprocessor and twig as templating language. Javascript libraries are used for a limited data display. Navigation of the Living Archive follows a standardized API approach, so that data can always be addressed with a clear logic for URLs. The address of a specific information is always human readable 'down the graph' (e.g. the construction layer 10540 is found at /Entity/ConstructionLayer/Unit-10540). By reducing the usage of overarching Javascript code to a minimum, data can be accessed transparently and visualizations can make use of libraries like D3.js, OpenLayers or ESRIs ArcGIS API for JavaScript.

Conclusions

While the Çatalhöyük information system reflects the type of field recording systems available since the mid 1990s it also afforded the inclusion of a rather unique reflexive archaeological methodology. Thus, the archive today contains a large amount of 'data about the data'; and we suggest an approach to tap into this wealth of information.

Linked Open Data is an approach by which definition advances transparency in the publication of research data. Yet, the underlying assumptions and theoretical implications are generally not made visible. Assuming that knowledge contained in the Çatalhöyük Research Project research data is subjective, we propose that research repositories that can be considered a "meta"-knowledge base can be inspected with regard to the interpretative values hidden in the data structures. By allowing for contradictory data to show up in the publication, different approaches of research during archaeological practice allow for the understanding of how data was created and to some extent why a resulting publication supported a specific interpretation.

If we are to advance the exposure of scientific knowledge on the web, we must examine data models to formalize implicit semantics and make underlying assumptions explicit. While the Linked Open Data approach already focuses on making the informational structure transparent as opposed to the traditional publication of data in its own right, in publishing the knowledge of a complex long-term project like Çatalhöyük we need to also make the reflexive layer transparent. The formalization of research and its publication evolve in a process of continuous interaction with the data, a process we try to capture in 'The Living Archive.' It will allow for extending research at Çatalhöyük far beyond the 25 years of archaeological fieldwork on site.

Acknowledgments

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Geolocation Information: 37.667N, 32.828E

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on Contributors

Dominik Lukas (M.A. 2008, University of Leipzig) is the Database Developer of the Çatalhöyük Research Project at Stanford University. He studied prehistoric archaeology and philosophy (University of Leipzig, Université-Marc-Bloch/Strasbourg). In 2008 he completed his master's thesis on the archaeological concept of the late iron age "Oppidum." From 2008 to 2012 he worked at the German Archaeological Institute (DAI) and then at the Excellence Cluster Topoi in Berlin as coordinator for research databases.

Claudia Engel (Ph.D. 1990, University of Göttingen) is an Academic Technology Specialist and Lecturer at the Department of Anthropology, Stanford University, USA. She also is an affiliate of the Center for Interdisciplinary Digital Research at the Stanford University Libraries. From 2012–2014 she was the Database Developer for the Çatalhöyük Research Project.

Camilla Mazzucato (MSc 2006, University College London) is a Ph.D. Candidate in Anthropology at Stanford University and a Ph.D. fellow at Koç University (Istanbul). She completed her BA and MA in Archaeology at Bologna University and her MSc in GIS and Spatial Analysis at UCL. Mazzucato has been part of the Çatalhöyük Research Project since 2009 as the main GIS specialist. She additionally spent two years (2011–2012) working as a researcher at the Institute of Social and Cultural Anthropology of Oxford University.

ORCID

Dominik Lukas http://orcid.org/0000-0001-9195-8481
Claudia Engel http://orcid.org/0000-0002-5234-3924

References

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

Atici, L., S. W. Kansa, J. Lev-Tov, and E. C. Kansa. 2013. "Other People's Data: A Demonstration of the Imperative of Publishing Primary Data." *Journal of Archaeological Method and Theory* 20 (4): 663–681.

Berggren, Å., and I. Hodder. 2003. "Social Practice, Method, and Some Problems of Field Archaeology." *American Antiquity* 68 (3): 421–434.

Berggren, Å., and B. Nilsson. 2014. "Going Back, Looking Forward: Reflexive Archaeology or Reflexive Method?" In *Integrating Çatalhöyük. Themes from the 2000–2008 Seasons*. Vol.10, edited by I. Hodder, 57–70. Los Angeles: Cotsen Institute of Archaeology Press.

Berggren, Å., N. Dell'Unto, M. Forte, S. Haddow, I. Hodder, J. Issavi, and J. S. Taylor. 2015. "Revisiting Reflexive Archaeology At Çatalhöyük: Integrating Digital And 3D Technologies At The Trowel's Edge." *Antiquity* 89 (344): 433–448.

Bevan, A. 2015. "The Data Deluge." *Antiquity* 89 (348): 1473–1484.

Calvanese, D., B. Cogrel, S. Komla-Ebr, D. Lanti, M. Rezk, and G. Xiao. 2015. "How to Stay On Top of Your Data: Databases, Ontologies and More." In *The Semantic Web: ESWC 2015 Satellite Events. ESWC 2015 Satellite Events Portorož, Slovenia, May 31 – June 4, 2015 Revised Selected Papers*, edited by F. Gandon, C. Gueret, S. Villata, J. Breslin, C. Faron-Zucker, and A. Zimmermann, 20–25. New York: Springer.

Chrysanthi, A., P. Murrieta Flores, and C. Papadopoulos. 2012. "Archaeological Computing: Towards Prosthesis or Amputation?" In *Thinking beyond the Tool Archaeological Computing and the Interpretive Process. BAR International Series 2344*, edited by A. Chrysanthi, P. Murrieta Flores, and C. Papadopoulos, 23–44. Oxford: Archeopress.

Chrysanthi, A., Å. Berggren, R. Davies, G. P. Earl, and J. Knibbe. 2016. "The Camera "at the Trowel's Edge": Personal Video Recording in Archaeological Research." *Journal of Archaeological Method and Theory* 23 (1): 238–270.

Chujai, P., N. Kerdprasop, and K. Kerdprasop. 2014. "On Transforming the ER Model to Ontology Using Protégé OWL Tool." *International Journal of Computer Theory and Engineering* 6 (6): 484–489.

Cripps, P., A. Greenhalgh, D. Fellows, K. May, and D. Robinson. 2004. Ontological Modelling of the Work of the Centre for Archaeology CIDOC CRM Mappings, Specializations and Data Examples. CIDOC CRM Technical Paper.

Cripps, P., and K. May. 2010. "To OO or not to OO? Revelations from Ontological Modelling of an Archaeological Information System." In *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of*

CAA2004, Prato 13–17 April 2004, edited by F. Nicolucci and S. Hermon, 59–63. Budapest: Archaeolingua.

CRM-EH 2014. CRMEH - English Heritage Extensions to CIDOC CRM for the Archaeological Domain. Retrieved from http://www.heritagedata.org/crmeh/crmeh_current.rdf. [Accessed 20 August 2018]

Dallas, C. 2016. "Jean-Claude Gardin on Archaeological Data, Representation and Knowledge: Implications for Digital Archaeology." *Journal of Archaeological Method and Theory* 23 (1): 305–330.

Doerr, M. 2003. "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata." *AI Magazine* 24 (3): 75.

Doerr, M., and M. Theodoridou. 2011. "CRMdig: A Generic Digital Provenance Model for Scientific Observation." *TaPP* 11: 20–21.

Doerr, M., G. Hiebel, and Ø. Eide. 2013. "CRMgeo: Linking the CIDOC CRM to GeoSPARQL Through a Spatiotemporal Refinement." Institute of Computer Science FORTH, Tech. Rep. GR70013.

Doerr, M., M. Theodoridou, E. Aspöck, and A. Masur. 2015. "Mapping Archaeological Databases to CIDOC-CRM." In *CAA 2015 KEEP THE REVOLUTION GOING. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, 443–451.

Engel, C., and K. Grossner. 2014. "Representing the Archaeological Process at Çatalhöyük in a Living Archive." In *Assembling Çatalhöyük*, edited by I. Hodder and A. Marciniak, 13–23. Leeds: Maney Publishing.

Faniel, I. M., and A. Zimmerman. 2011. "Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse." *International Journal of Digital Curation* 6 (1): 58–69.

Faniel, I., E. Kansa, S. Kansa, J. Barrera-Gomez, E. Yakel, S. W. Kansa, and I. Faniel. 2013. "The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse." In *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries. Indianapolis, Indiana, USA – July 22 – 26*, 2013, 295–304. New York: ACM.

Farid, S. 2015. "Proportional Representations. Multiple Voices in Archaeological Interpretation at Çatalhöyük." In *Material Evidence. Learning from Archaeological Practice*, edited by R. Chapman and A. Wylie, 59–78. London and New York: Routledge.

Farid, S., and I. Hodder. 2014. "Excavation, Recording and Sampling Methodologies." In *Çatalhöyük Excavations. The 200-2008 Seasons*, edited by I. Hodder, 91–130. Los Angeles: Cotsen Institute of Archaeology Press.

Fong, J. S. P. 2015. *Information Systems Reengineering, Integration and Normalization*. New York: Springer International Publishing.

Gardin, J.-C. 2002. "Les Modeles Logico-Discursifs en Archeologie." *Archeologia E Calcolatori* 13: 19–30.

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

Grize, J.-B. 1974. "Argumentation, Schématisation et Logique Naturelle." *Revue Européenne des Sciences Sociales* 32: 183–200.

Hodder, I. 1997. "'Always Momentary, Fluid and Flexible': Towards a Reflexive Excavation Methodology." *Antiquity* 71 (273): 691–700.

Hodder, I. 2000. "Where is Çatalhöyük? Developing a Reflexive Method in Archaeology." In *Towards Reflexive Method in Archaeology: The Example at Çatalhöyük*, edited by I. Hodder, 3–15. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Hodder, I. 2003. "Archaeological Reflexivity and the "Local" Voice." *Anthropological Quarterly* 76 (1): 55–69.

Hodder, I. 2008. "Multivocality and Social Archaeology." In *Evaluating Multiple Narratives. Beyond Nationalist, Colonialist, Imperialist Archaeologies*, edited by J. Habu, C. Fawcett, and J. M. Matsunaga, 196–200. New York, NY: Springer.

Hodder, I. 2015. "Assembling Science at Çatalhöyük. Interdisciplinarity in Theory and Practice." In *Assembling Çatalhöyük*, edited by I. Hodder and A. Marciniak, 7–12. Leeds: Maney Publishing.

Hodder, I., C. Cessford, and S. Farid. 2007. "Introduction to Methods and Approach." In *Excavating Çatalhöyük: South, North and KOPAL Area Reports from the 1995-1999 Seasons*, edited by I. Hodder, 3–24. Cambridge: McDonald Institute for Archaeological Research, Cambridge University.

Hohmann, G., and M. Fichtner. 2015. "Digitales Kulturerbe : Bewahrung und Zugänglichkeit in der Wissenschaftlichen Praxis." [Digital Cultural Heritage: Preservation and Accessibility in Scientific Practice] In *Chancen und Herausforderungen in der Praktischen Anwendung von Ontologien für das Kulturerbe* [Chances and Challenges in Implementing and Applying Ontologies for Cultural Heritage], edited by R. H. Robertson-von Trotha and Caroline Y. Schneider, 115–128. Karlsruhe: KIT Scientific Publishing.

Isaksen, L. 2011. *Archaeology and the Semantic Web*. PhD Thesis at the University of Southampton. Faculty of Physical and Applied Sciences. School of Electronics and Computer Science.

Isaksen, L., R. Simon, E. T. E. Barker, and P. de Soto Cañamares. 2014. "Pelagios and the Emerging Graph of Ancient World Data." In *Proceedings of the 2014 ACM conference on Web science - WebSci '14*, edited by L. Fussell and W. Lutters, 197–201. New York, NY: ACM Press.

Jones, S. 2012a. The Çatalhöyük Database System. Unpublished internal report.

Jones, S. 2012b. The Çatalhöyük Team Systems. Unpublished internal report.

Jewell, M. O., E. Costanza, T. Frankland, G. Earl, and L. Moreau. 2012. "The Xeros Data Model: Tracking Interpretations of Archaeological Finds." In *Provenance and Annotation of Data and Processes 4th International Provenance and Annotation Workshop, IPAW 2012, Santa Barbara, CA, USA, June 19-21, 2012, Revised Selected Papers*, edited by P. Groth and J. Frew, 139–151. New York: Springer.

Kansa, E. 2012. "Openness and Archaeology's Information Ecosystem." *World Archaeology* 44 (4): 498–520.

Kintigh, K. W., J. H. Altschul, A. P. Kinzig, W. F. Limp, W. K. Michener, J. A. Sabloff, and C. A. Lynch. 2015. "Cultural Dynamics, Deep Time, and Data Planning Cyberinfrastructure Investments for Archaeology." *Advances in Archaeological Practice* 3 (1): 1–15.

Le Boeuf, P., M. Doerr, C. E. Ore, and S. Stead, S. 2017. "Definition of the CIDOC Conceptual Reference Model." Retrieved from http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.2, 1.

Lukas, D. and C. Engel. 2017. "Changing Technologies, Changing Practices: The Transformation of the Çatalhöyük Research Database." Presentation at the SAA 82nd Annual Meeting in Vancouver 2017.

Maggiolo-Schettini, A., Seccacini, C. D., Serratore, R., Pierobon-Benoit, R., and Soricelli, G. 1995. "SYSAND: a System for the Archaeological Excavations of Anderitum (Javols, Lozère, France) - CAA Online Proceedings." In *CAA94. Computer Applications and Quantitative Methods in Archaeology 1994 (BAR International Series 600)*, edited by J. Huggett and N. Ryan, 229–234. Oxford: Tempus Reparatum.

Marciniak, A. 2015. "Introduction. Assembling the Archaeological Process at Çatalhöyük." In *Assembling Çatalhöyük*, edited by I. Hodder and A. Marciniak, 1–5. Leeds: Maney Publishing.

McManamon, F., K. Kintigh, and A. Brin. 2010. "Digital Antiquity and the Digital Archaeological Record (tDAR): Broadening Access and Ensuring Long-Term Preservation for Digital Archaeological Data." *CSA Newsletter* 23 (2). http://csanet.org/newsletter/fall10/nlf1002.html. *Accessed 22 June 2014*.

Mellaart, J. 1964. "Excavating Çatal Höyük 1963. Third Preliminary Report." *Anatolian Studies* (14): 39–119.

Mickel, A. 2016. "Tracing Teams, Texts, and Topics: Applying Social Network Analysis to Understand Archaeological Knowledge Production at Çatalhöyük." *Journal of Archaeological Method and Theory* 23 (4): 1095–1126.

von Mühlen, M. 2014. "We Need a Github of Science." http://marciovm.com/i-want-a-github-of-science/

Musen, M. A. 2015. "The Protégé Project: A Look Back and a Look Forward." *AI matters* 1 (4): 4–12.

Orlandi, T. 2004. "Archeologia Teorica e Informatica Archeologica. Un Rapporto Difficile." *Archeologia e Calcolatori* XV: 41–50.

Pryor, R. A. 1973. "A General Scheme for Local Inventories." In *Computer Applications in Archaeology 1*, edited by J. D. Wilcock, 7–16. Stafford: Science and Archaeology no. 9. George Street Press.

Puyol-Gruart, J. 1999. "Computer Science, Artificial Intelligence and Archaeology - CAA Online Proceedings." In *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998 (BAR International Series 757)*, edited by J. A. Barceló, I. Briz, and A. Vila, 19–28. Oxford: Archaeopress.

Py, M. 1997. *SYSLAT 3.1, Système d'Information Archéologique, Manuel de Référence* [SYSLAT 3.1, Archaeological Information System, Reference Manual], edited by Guy Barruol et al., Lattes: Lattes : Ed. de l'Association pour la Recherche Archéologique en Languedoc Oriental.

Richards, J. D. 2009. "From Anarchy To Good Practice: The Evolution Of Standards In Archaeological Computing." *Archeologia E Calcolatori* 20: 27–35.

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

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

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

Ronzino, P., F. Niccolucci, A. Felicetti, and M. Doerr. 2016. "CRMba a CRM Extension For The Documentation Of Standing Buildings." *International Journal on Digital Libraries* 17 (1): 71–78.

Ronzino, P. 2017. "Harmonizing the CRMba and CRMarchaeo Models." *International Journal on Digital Libraries* 18 (4): 253–226.

Roux, V., and N. Aussenac. 2013. "Knowledge Basis and Query Tools for a Better Cumulativity in the Field of Archaeology: The Arkeotek Project." In *Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2010*, edited by F. Contreras, M. Farjas, and F. Melero, 267–272. Oxford: Archeopress.

Roux, V., and M.-A. Courty. 2013. "Introduction to Discontinuities and Continuities: Theories, Methods and Proxies for a Historical and Sociological Approach to Evolution of Past Societies." *Journal of Archaeological Method and Theory 20* (2), 187-193.

Santoriello, A., and F. Scelza. 2000. "Un Sistema Informativo Archeologico: l'applicazione del Syslat a Fratte di Salerno." [An archaeological information system: implementing Syslat at Fratte di Salerno] *Archeologia E Calcolatori* 11: 69–92.

Simon, R., L. Isaksen, E. Barker, and P. de Soto Cañamares. 2016. "The Pleiades Gazetteer and the Pelagios Project." In *Placing Names: Enriching and Integrating Gazetteers*, edited by M. L. Berman, R. Mostern, and H. Southall, 97–109. Bloomington, IN: Indiana University Press.

Stead, S., and M. Doerr. 2015. "CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to Support Argumentation." http://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdfhttp://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdf [Accessed 23 May 2016].

Taylor, J., J. Issavi, Å. Berggren, D. Lukas, C. Mazzucato, B. Tung, and N. Dell'Unto. 2018. "'The Rise of the Machine': the Impact Of Digital Tablet Recording In The Field at Çatalhöyük." *Internet Archaeology* 47 (47).

Telnarova, Z. 2010. "Relational Database as a Source of Ontology Creation." In *Proceedings of the International Multiconference on Computer Science and Information Technology*, edited by M. Ganzha and M. Paprzycki, 135–139. Piscatway, NJ: Institute of Electrical and Electronics Engineers.

Toulmin, S. E. 1958. *The Uses of Argument*. Cambridge: Cambridge University Press.

Tringham, R., and M. Stefanović. 2012. *Last House On The Hill: BACH Area Reports From Çatalhöyük, Turkey*. Monumenta Archaeologica 27. Los Angeles: Cotsen Institute of Archaeology Press.