Engaging with the Çatalhöyük Database
House at Çatalhöyük (HATCH) and Other Applications

Arkadiusz Marciniak, Jacek Marciniak, Patrycja Filipowicz, Katarzyna Harabasz, and Jędrzej Hordecki

Çatalhöyük East, TPC Area, B. 150, the context of deposition of a large anthropomorphic figurine.

Çatalhöyük remains one of the most intensively and meticulously excavated prehistoric sites in the history of archaeology. Two large excavation campaigns, carried out in the years 1961–1965 and 1993–2017, have generated considerable data. A particularly large body of data was accumulated over the twenty-five-year period of excavation and analyses carried out by the Çatalhöyük Research Project. One of the project's major undertakings was to make all data collected accessible for wider audiences (Hodder 2000).

The Çatalhöyük Research Project took place in a period of unprecedented development of IT technologies and their application in archaeology. In the 1990s, computers started to become a widely used tool for archaeology for the writing of texts, the processing of drawings and maps, as well as storing and organizing a wide range of data in the form of databases. They significantly enhanced different quantitative and statistical analyses of archaeological data. The Çatalhöyük project from its very beginnings was in the frontline of computer application in archaeology, and it became the first excavation available via the World Wide Web (www.catalhoyuk.com). As such, it is an exemplary case of how technological advances throughout the project have been incorporated into digital data-management and enhanced the multifaceted process of interpretation of several categories of data (Engel and Grossner 2015).

The Çatalhöyük Database

The first Microsoft Access 2000 databases at Çatalhöyük were designed for excavation, finds, and specialist data. Two servers acquired at the beginning of the project, one for use on site and the other for the Cambridge office, made it possible to install Microsoft SQL Server 2000 to use as a single backend database and to start re-centralizing the databases. The databases were under constant development (different teams kept changing personnel and methods through time), so the system of recording has been updated by different specialists. The main challenge was to build a synergy between field teams and increasing number of labs and specializations (Engel and Grossner 2015).

The most recent Çatalhöyük database is a relational database, running primarily on a MS SQL 2008 R2 Server. The database contains some one thousand relational tables that hold all excavation data since the beginning of the project. MS Access 2000–2016 applications are used as frontends (Graphical User Interfaces). Team data all refer to object finds and samples from excavation *units*. Units are grouped into *features*, *spaces*, and *buildings.* Database frontends include: Excavation, Finds, Botany, Animal, Human Remains, Clay/Figurines, Chipped Stone, Shell, and Ground Stone, among others. Using the central server also allows the user to connect the database with a website that can theoretically show all the data accessible by the MS Access applications. Images and other types (nontextual, geospatial) of data are integrated in the database: photographic catalogue (Extensis Portfolio), GIS geodatabase (ESRI ArcGIS) and digital drawings, diary (more subjective "soft" data), and 3D data (since 2009 and using different devices and technologies) (see Engel and Grossner 2015).

The main dataset was only accessible intra-site (offline) during each excavation season; once the field season was over, the database information was incorporated into the official Çatalhöyük website. However, the online version contains only basic information; access to the full content is restricted by password

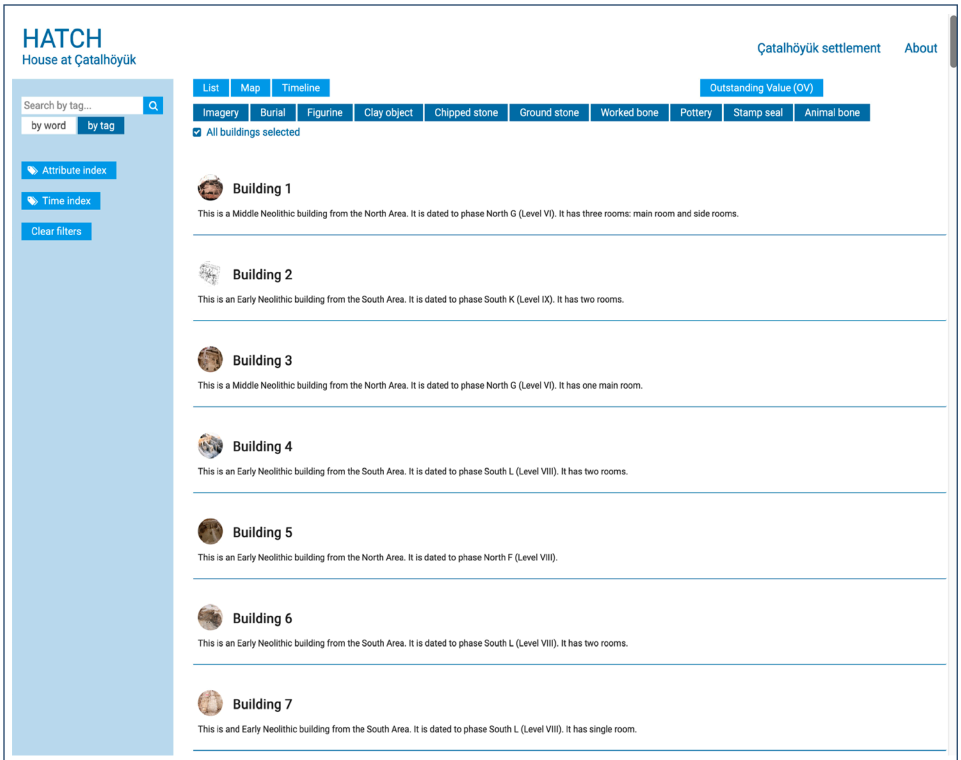

Figure 1. HATCH: House at Çatalhöyük main page.

and cannot be accessed by everyone. For example, the website version of the Human Remains Database, as compared to the version available at the site or with password, does not contain detailed information about the interred individual (e.g., sex, age, specific osteological description and georeferenced data). It only contains generic descriptions of selected individuals, contextual information (unit, feature, space, building), and graphics in some instances.

As indicated above, the project at Çatalhöyük produced a vast amount of data. The challenge we now face is how to make them available given rapid technological change, how to integrate all kinds of data in one place, and how to make specialist data more comprehensible to the public. Data sharing remains one of the most demanding tasks facing archaeologists today. In comprises a wide range of technical, conceptual, and incentive problems, particularly in terms of pooling and integrating complex and unstandardized data. The problem of interoperability and integrated search, browse, and analysis tools remains of major challenges. One of the most developed data sharing platform in Near Eastern archaeology is Open Context developed by the Alexandria Archive Institute (https://alexandriaarchive.org/what-we-do/open-context). It is a highly flexible database that enables researchers to publish structured data along with textual narratives and media (images, maps, drawings, videos) on the web (Whitcher Kansa, Kansa, and Schultz 2007).

Different forms of engagement with the Çatalhöyük database not only enhance its visibility and accessibility, but can serve as an exemplary case for archaeology at large.

Engaging with the Çatalhöyük Database: HATCH - House at Çatalhöyük

The most recent attempt to present Çatalhöyük data in a considerably different way is HATCH (House at Çatalhöyük; http://hatch.e-archaeology.org; fig. 1). It is a new open-access repository aimed at presenting a wide range of data related to the settlement in a strictly integrated format.

HATCH has the form of MIODEC—Multidimensional Interactive Open Digital Educational Collection—which is an innovative software for creating and maintaining digital collections. It allows data to be presented in a multiscalar and interactive form, interlinking information of different characters (e.g., types of objects, relations among them) with different forms (e.g., text, photographs, graphics). HATCH implements a model-driven architecture where a domain model is constructed as a wordnet-based ontology. This approach allows integrating, in the form of one database, data of different characters: general (accessible for general public), expert (produced by archaeologists), and application (data structures organizing the archaeological site). Such a multifaceted method is the core of a semantic search engine designed for a nonexpert public.

HATCH is a tool to serve explicitly the didactic goals of the lay public as well as members of the archaeological community who are not experts in the Near Eastern Neolithic. To standardize information, data were subjectively selected and described, and, as compared to the original database, their quantity was significantly reduced. Nevertheless, the solutions adopted in HATCH will facilitate an in-depth understanding of different aspects of this large settlement. Users start interacting with the tool by asking a range of simple questions and can further explore the repository by asking more advanced questions and delving into the previously unthinkable aspects of the settlement history.

HATCH was created by archaeologists—members of the Çatalhöyük Research Project team and IT specialists from Adam Mickiewicz University in Poznań, Poland—in collaboration with other members of the Çatalhöyük project. The data have been harvested from a wide range of available publications (monographs, excavation reports, journal articles, etc.). Since the analysis and publication of different sets of data from the site are pending, some aspects of the currently stored data are subject to future revisions.

The main type of object presented within HATCH is the building (fig. 2). Buildings are used to aggregate a wide range of other categories of data including burials, imagery, figurines, clay objects, chipped stones, ground stones, worked bones, pottery, stamp seals, and animal bones. The auxiliary entities are special deposits and finds. Each category of object is described according to a set of predefined attributes. These are divided into two categories: general and specific. The former comprises chronology, year of discovery, size, and textual descriptions. The latter includes the number of rooms, house construction elements, decorative motifs, among others. Textual data are accompanied by photographs, drawings, 3D reconstruction (if available), and plans of subsequent phases of occupation as distinguished by the excavators. Currently, the database contains information on 143 buildings dated back to the Early, Middle, Late Neolithic, Chalcolithic, Hellenistic, and Seljuk periods.

HATCH also allows a comprehensive presentation of burials found at the settlement. They are categorized in the form of five burial types as defined by the deposition of an individual in the burial pit. These include primary, primary disturbed, secondary, tertiary, and indeterminate interments. The age of the deceased is described by eight categories. Elaborated burials, characterized by a special treatment of body and/or with rich grave goods, were assigned to the special treatment category. Moreover, a presentation of each individual is supplemented by a comprehensive description of its osteological characteristics and archaeological context. Textual data are accompanied by photographs and drawings (if available) of the skeletons and their burial context.

HATCH presents the data in the form of three intertwined modes: typological, spatial, and chronological. The typological mode presents information in the form of a list of objects displayed according to the specified filters. Data in the spatial mode are displayed in the form of GIS objects placed in the scaled cartographic map (based upon Open Street Map). The chronological mode involves presentation of both relative and absolute chronologies. When presenting relative chronology, the objects are tagged by one or more mutually exclusive formats of relative chronology of the settlement: Hodder Phase, Mellaart Phase, and TP Phase (fig. 3; see the chronological scheme in table 1 of *NEA* 83.2 [p. 77]). As in a majority of cases, existing chronologies are not yet reliably translated into absolute chronology, buildings from corresponding occupational areas and stages of the settlement development are only presented in the form of relative chronology. Both relative and absolute chronologies are currently only available for strata featured by TP Phases.

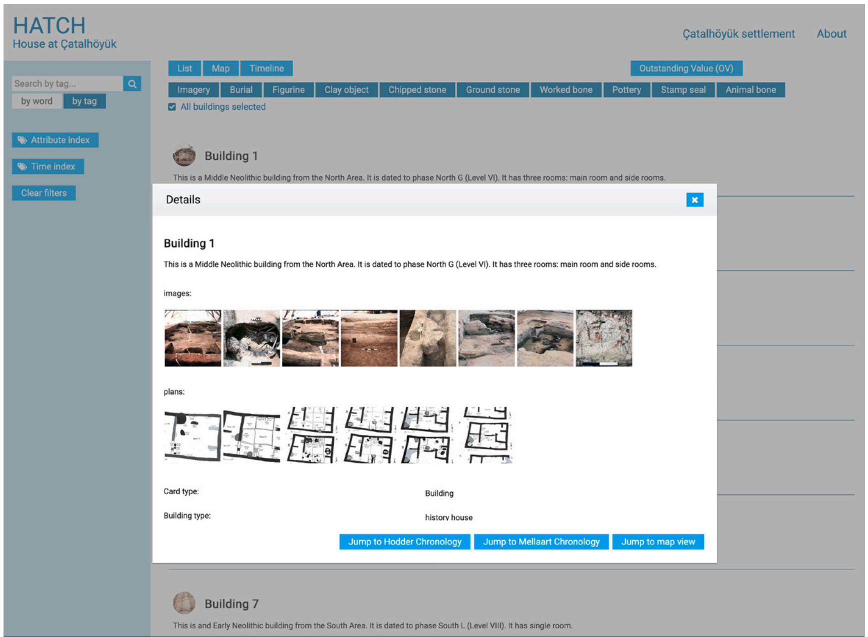

Figure 2. HATCH. Example of a Building card: Building 1 and different phases of its occupation.

Figure 3. HATCH: Presentation of the Çatalhöyük house.

Data can be displayed in a range of different formats thanks to the advanced search engine. It operates in two modes: word search and tag search. The former facilitates access to the stored data by attributes of a textual type, while the latter makes it possible to search by tags attached to the objects as attribute values immersed in the wordnet-based ontology. In both cases, the user can select thematically linked tags. The search engine makes it possible automatically to prompt tags that remain semantically linked to the word used in a search by referring to relevant connections from the wordnet based ontology.

Exploring the Çatalhöyük Database: The Çatalhöyük Living Archive and Seshat

Çatalhöyük Living Archive

As HATCH was designed for the lay public and members of the archaeological community at large, only an insignificant amount of the large Çatalhöyük data was used to explore the settlement history. This approach differs considerably from the

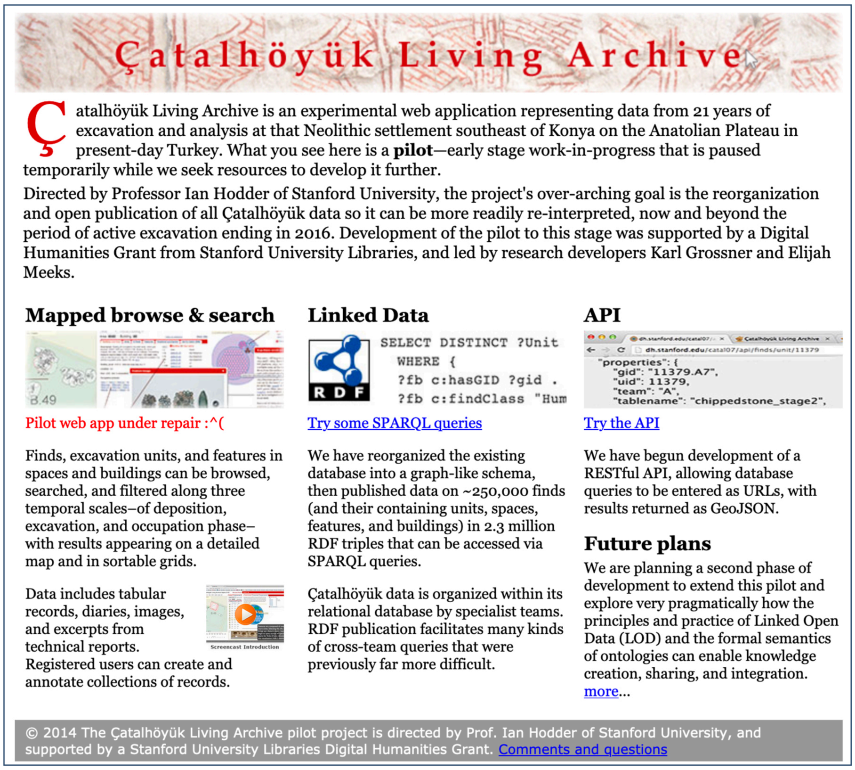

Figure 4. Çatalhöyük Living Archive; main page of the first prototype.

Figure 5. Seshat main page.

attempt to make all data acquired and collected throughout the Çatalhöyük Research Project accessible and usable well beyond its duration. To meet this goal, the newest recently developed technologies were adopted not only to ensure the availability of the project data but also to enhance considerably the ways in which they can be organized, stored, and analyzed. The Microsoft Access database built for the Çatalhöyük project has certainly not allowed the required fluidity as the large number of diverse data created obstacles for integration (Berggren and Nilsson 2014; Lucas, Engel, and Mazzucato 2018). Therefore, attempts were made to move the project data to a web-based system and a long-term "living archive." This approach, following completion of the project, will make frozen data accessible for future analysis, interpretation, and annotation by its publication as Linked Open Data (LOD) (Hodder et al. 2014). Hence, the living archive would not only be the repository of the already collected data but would also let researchers add the results of queries so that they are available to future researchers.

The Çatalhöyük Living Archive had been initially developed within a Stanford University digital humanities project (fig. 4). Its first step involved the transformation of a significant portion of the original data from a relational model to experimental graph representations. The Resource Description Framework (RDF), and its formalization in the form of the Web Ontology Language (OWL), was the model used for tagging data. Its core structure of statement is in a "triple" form (subject, predicate and object), which makes the data sematics readable and hence linkable with data in other web pages. The graph data model of RDF/OWL enables also an easier representation of displays of finds, features, and samples excavated from spaces, buildings, and areas. Since 2019 the initial prototype being developed has been further supported by a Digital Art History Grant from the Getty Foundation. Using the API of the triple store rdf4j allows for direct use of data within the Living Archive application and at the same time enables direct querying of the data with the use of SPARQL (REST-API), allowing for external applications to connect to the archive in a Linked-Open-Data approach.

All data in the archive will be organized along its semantic value—its position on a knowledge graph—and its geospatial location on site, by implementing the latest web-mapping functionalities. All data can be queried and visualized following the respective best practices established for each data type to date. Moreover, two timelines (deposition and excavation time scales), and interactive visualizations will be joined with a traditional search form, enabling researchers to navigate the data store visually in novel ways. Data include tabular records, diaries, images, and excerpts from technical reports. Registered users can create and annotate collections of records. Furthermore, the user will be able to manipulate the semantic structure in order to experiment with different interpretations of the data. In this way, the system will facilitate reflexivity and multivocality, important aspects of the Çatalhöyük Research Project (Grossner et al. 2012).

Çatalhöyük in Seshat: The Global History Databank

A completely different form of engagement with the Çatalhöyük database is the Seshat Global History Databank (http://seshatdatabank.info). Data from the Çatalhöyük excavations got integrated with Seshat in the same way as many sites from Southwest Asia. They can be used for testing different hypotheses about the Neolithic of this region and beyond. The sharing and incorporating of data into different, larger scale Big Data projects (like Seshat) can be a model for other sites.

The Seshat Global History Databank is a massive database of information on past societies (fig. 5). It was founded by historians in 2011 (Turchin et al. 2015). It focuses on "Big Questions" and identifying long term trends. It was designed as a general tool for testing different theories about social evolution and historical dynamics. The long-term goal is to incorporate structured data on all relevant variables for all past human societies. The Seshat databank is very comprehensive. It comprises social, economic, ritual, warfare, and other variables, which are extracted from web sites, dedicated databases, academic publications, and human experts. These are recorded in the Natural Geographic Areas (NGAs) that are divided into specific historical societies, or polities. Seshat defines a polity as an independent political unit e.g. villages (local communities) through simple and complex chiefdoms to states and empires.

In the first phase of its development, the Seshat databank was based on a WIKI engine and required manual data collection. The second phase involved a development of a sophisticated software platform based on the RDF technology. The Seshat data are now stored in a Triplestore database. It is a purpose-built database for the storage and retrieval of triples through semantic queries. This format enables knowledge to be represented in a machine-readable way. This solution allows a move away from manual data harvesting, therefore to speed up the data-gathering (and later querying) process by building web crawlers that go over large quantities of digital text. It structures the data-gathering process and keeps track of the quality of information. The Seshat database is currently managed through the DaCura Linked Data platform, developed at Trinity College Dublin (http://dacura.scss.tcd.ie). DaCura provides support for dataset capturing, curation, and publication.

As Seshat was designed to address historically specific processes, its relevance to archaeology was not directly evident. Hence, it was necessary to construct an Archaeological Seshat in order to address a range of hitherto unexplored issues and make it possible to harvest a wide range of archaeological data. The Archaeological Seshat was designed and constructed by a group of archaeologists from Adam Mickiewicz University—the members of the Çatalhöyük project—in cooperation with scholars from Europe and the United States.

Archaeological Seshat offers a new means of unprecedented heuristic potential of exploring the Çatalhöyük database. It involves (1) an automated harvesting of different types of data from the site database and integrating them with a wide range of data originating from publications, press releases, lectures, etc., (2) expanding the Seshat designed perspective by moving from polity level to micro-level narrations of smaller research entities (site, house, features, excavation areas; family, households, society, individuals), (3) comparing different variables with corresponding developments at other sites and regions in a systematic and comprehensive way. The future work will involve investigating the nature of the trajectory of development from the beginning of Neolithic way of life to more complex, sedentary farming societies across the Near East, including Anatolia (including the Konya Plain with Çatalhöyük), through the Balkans and Central Europe. The collected data are going to be used to test explicitly formulated hypotheses by means of appropriately selected mathematical models.

Concluding Remarks

The Çatalhöyük Research Project was always in the vanguard of computer application in archaeology. As it was conducted for as long as a quarter of a century, it inevitably involved a never-ending catching up with the newly emerging technologies and software. The rich and highly complex datasets it produced are set to explore new technological advances even after its formal completion. They go beyond the hitherto dominant mode of presenting archaeological finds in the form of integrated excavation and specialist datasets and providing open access services. Innovative modes of engagements with the Çatalhöyük database presented above facilitated a significant enhancement of interpretive and heuristic potential of the project-generated data. They not only yield attractive forms of exploration of different categories of data by both the lay public and archaeological community but also offer new ways of exploring and investigating the hitherto undeveloped dimensions of existence of farming groups at the dawn of settled life in the Near East.

The amount of data produced by the project makes it highly complicated to understand and to use effectively, even by experienced scholars, not to mention the lay public. Moreover, the anticipated rapid development of IT solutions in the future will make access to the database, developed during the duration of the project, significantly restricted or even impossible. These two circumstances triggered different modes of engagements with the Çatalhöyük database and its ultimate transformation into new formats. They are characterized by different levels of advancement and detail. They are directed to different end users, with their specific needs and expectations. All three of them are based on model-driven ontologies and adopt different semantic web solutions. The major advantage of HATCH is the presentation of the most important aspects of Çatalhöyük in a comprehensive and easily assessable way. It ideally serves the needs of the academic community and the general public at large. Seshat, on the other hand, makes it possible to transform the Çatalhöyük data into different formats, putting up the settlement in relation to other sites and regions facilitating the study of long durée processes. The Çatalhöyük Living Archive is being developed to enable researchers as well as nonexperts to query research at Çatalhöyük, making the research intuitively understandable and allowing new data to be produced from within the existing database. In this way the vast amounts of data are being made sustainable and long lasting for future scholarship.

Acknowledgment

All visual materials contained on the Hatch website are the property of Çatalhöyük Research Project. Hatch has been produced with the support of the European Commission (through the Culture 2007–2013 program) in the framework of the NEARCH Project New scenarios for a community-involved archaeology (Nr 3085/Kultura/2014/2). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

Berggren, Åsa, and Bjorn Nilsson. 2014. Going Back, Looking Forward: Reflexive Archaeology or Reflexive Method? Pp. 57–70 in *Integrating Çatalhöyük: Themes from the 2000–2008 Seasons*, vol. 10, ed. Ian Hodder. Los Angeles: Cotsen Institute of Archaeology.

Engel, Claudia, and Karl Grossner. 2015. Representing the Archaeological Process at Çatalhöyük in a Living Archive. Pp. 13–23 in *Assembling Çatalhöyük*, ed. Ian Hodder and Arkadiusz Marciniak. European Association of Archaeologists Contemporary Themes in Archaeology. Leeds: Maney.

Grossner, Karl, Ian Hodder, Elijah Meeks, Claudia Engel, and Allison Mickel. 2012. A Living Archive for Çatalhöyük. In *Proceedings of the 2012 Computer Applications and Quantitative Methods in Archaeology.* PLACE: PUBLISHER. *{AQ: please confirm date and title; I was not able to find this on the CAA website list}*

Hodder, Ian. 2000. Where Is Çatalhöyük? Developing a Reflexive Method in Archaeology. Pp. 3–15 in *Towards Reflexive Method in Archaeology: The Example at Çatalhöyük*, ed. Ian Hodder. Cambridge: McDonald Institute for Archaeological Research.

Witcher Kansa, Sarah, Eric C. Kansa, and Jason M. Schultz. 2007. An Open Context for Near Eastern Archaeology. *Near Eastern Archaeology* 70: 188–94.

Lukas, Dominik, Claudia Engel, and Camilla Mazzucato. 2018. Towards a Living Archive: Making Multi-Layered Research Data and Knowledge Generation Transparent. *Journal of Field Archaeology* 43: S19–S30, DOI: 10.1080/00934690.2018.1516110.

Turchin, Peter, Rob Brennan, Thomas Currie, Kevin Feeney, Pieter Francois, Daniel Hoyer, Joseph Manning, Arkadiusz Marciniak, Daniel Mullins, Alessio Palmisano, Peter Peregrine, Edward A. L. Turner, and Harvey Whitehouse. 2015. Seshat: The Global History Databank. *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* 6: 77–107, https://doi.org/10.21237/C7clio6127917.

Arkadiusz Marciniak is professor of archaeology at Adam Mickiewicz University in Poznań. His expertise is in the development of early farming communities in western Asia and central Europe and their progression to complex societies. He has been directing the excavation of the Late Neolithic settlement at Çatalhöyük within the Çatalhöyük Research Project.

Jacek Marciniak is IT professor at Adam Mickiewicz University in Poznań. He carries out research in the areas of Artificial intelligence and Natural Language Processing. The primary focus of his work is on ontological modelling based on wordnet based ontologies and applications of ontology-based approach as knowledge model in semantic search engines, Intelligent Tutoring Systems and interactive digital collections.

Patrycja Filipowicz is assistant professor of archaeology at Adam Mickiewicz University in Poznań. Since 2005 she has been a member of the Polish research team at the Neolithic site of Çatalhöyük.

Katarzyna Harabasz is an archaeologist and bioarcheologist. She is assistant professor of archaeology at University of Szczecin. She is a specialist in Near Eastern archeology. Since 2012, she has been participating in excavations of the Neolithic settlement in Çatalhöyük in Turkey. She is currently investigating Neolithic human remains from Çatalhöyük.

Jędrzej Hordecki received his PhD degree in archaeology in 2020 at the Faculty of Archaeology, Adam Mickiewicz University in Poznań. He has been a member of the Çatalhöyük Research Project since 2012. His research interests comprise the archaeology of central Anatolia in the Iron Age and Hellenistic period.