CHAPTER 2

Representing the Archaeological Process at Çatalhöyük in a Living Archive

Claudia Engel and Karl Grossner

Background

Research at the nine thousand year old Neolithic settlement of Çatalhöyük in central Turkey has pioneered the implementation of a reflexive approach to archaeological practice, known as post-processual archaeology, in which information is permanently open to re-interpretation by both scholars and the public (Hodder, 2000). According to Hodder, who has been directing the excavations on this site since 1993 'post processual approaches focus on interpretation, multivocality, meaning, agency, history' (Hodder, 1999: 12), resulting in 'a diversity of views […] espoused with no singular and unified perspective imposed on the discipline' (Hodder, 1999: 5).

Reflexive archaeological method acknowledges that the archaeological discipline itself, including the scholars, their methods, and tools, influences the resulting images of the past. Archaeology does not produce objective facts; rather, different researchers and communities will produce different interpretations, some of which may appear to fit the data better than others. In the Çatalhöyük project,

> …reflexivity is defined as an examination of the effects of archaeological assumptions and actions on the different communities involved in an archaeological process, including participants in the project as well as other archaeologists and non-archaeological communities […] The individual excavator is emphasized as playing an important role in forming the interpretations, and the goal is to record this subjective trait. The field methods are, therefore, aimed at documenting what may influence the archaeological interpretations (e.g. the preconceptions and assumptions of the excavating staff). (Berggren, 2014a)

Research at Çatalhöyük is emphasized as playing an important role in forming the interpretations. The ongoing series of decisions during the excavation—responding to circumstances as they arise—are based on subsequent assumptions and new questions arising from continuous interpretation and re-interpretation. Furthermore, it is considered imperative to bring transparency to those processes.

The implementation of the reflexive approach at Çatalhöyük has been outlined in detail by Hodder (1997, 1999, 2000) and includes the following: facilitated interaction between the excavators and the specialists, (e.g. collaborative decisions about research priorities); improved 'fast track' contextualization of finds through immediate availability of lab results for the excavators to help determine excavation strategies; video documentation of excavations; a central database—accessible to all team members—to integrate dispersed data collections; a 'digital diary' for excavators to reflect on their daily excavation process and contextualize the database records; and public access to the database. In 1993, when just over six hundred websites existed globally[1], Çatalhöyük became the first excavation to make its records available via the Web (www.catalhoyuk.com) and to invite public comment.

This chapter reports on how technological advances have been incorporated into the digital data management at Çatalhöyük with the ultimate goal to support an inter-disciplinary process of assembling data into arguments on the basis of multiple lines of evidence. We describe the database infrastructure at Çatalhöyük and how it currently supports reflective practice. We then lay out our vision of an interactive archive, components of a web application and a radically re-designed data store we are currently developing in collaboration with Hodder. This 'living archive' leverages recent technological innovations in geo-visualization and Linked Open Data (LOP) to support long-term, collaborative, multivocal knowledge creation.

The Current Çatalhöyük Research Repository

Since the beginning of the project under the directorship of Ian Hodder, excavated material has been meticulously digitally recorded. At Çatalhöyük technological innovations were constantly explored and applied for their potential to improve information

[1] http://stuff.mit.edu/people/mkgray/net/web-growth-summary.html

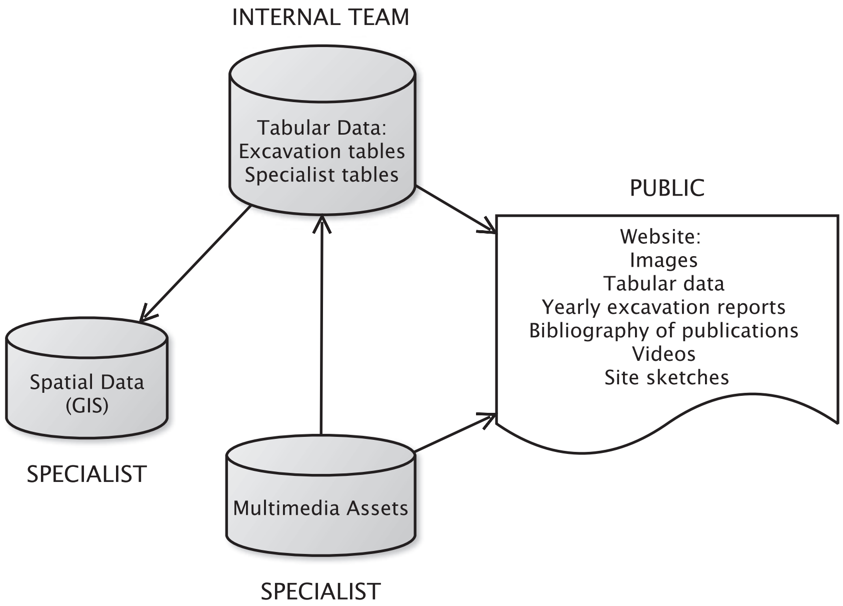

Figure 1. *Database infrastructure.*

flows between the trenches, labs, and beyond and to provide new ways of documentation and capturing the archaeological process. Past innovations include the implementation of a local computer network on site, the daily video documentation of the excavation progress or laboratory work, and the use of hypertext and virtual reality (e.g. Quicktime-VR).

The current infrastructure of the Çatalhöyük repository comprises three major databases (Figure 1). All tabular data are housed in a standard relational database (Microsoft SQL Server). Multimedia assets are stored in an instance of a proprietary image database system (Extensis Portfolio). Both run as services that can be accessed remotely. Spatial (GIS) data are held in a stand-alone file geodatabase (ESRI ArcGIS) without the possibility for remote access[2].

Programmed automated procedures, scripts, and database 'views' (dynamic tables created by queries) function as conduit between the different databases. This infrastructure allows to serve images and a subset of tabular data to the public project website, to augment the GIS footprints for excavation units, features, buildings, spaces, areas, and occasionally, special finds (termed 'X-finds') with information from the central database, and to make photographs and images available to be viewed together with records from the central database tables.

The Çatalhöyük team includes a GIS specialist, responsible for entering and querying spatial data as well as producing cartography, and a Multimedia specialist who manages upload and retrieval of multimedia assets.

[2] Efforts are currently under way to migrate the spatial data to an ArcSDE server to allow for remote access.

Data entry into the tables of the central database is performed by the Excavation and Specialist Teams through dozens of team-specific forms. The forms are highly customized desktop interfaces built with Microsoft Access and Visual Basic (hereafter, Access). The tables can be queried within a separate generic Access interface that exposes many of the over six hundred tables of the system. A limited web browsing capability is open to the public via the project website.

Centralization of excavation and specialist team data

The central relational database system is a pillar of the Çatalhöyük project. It has served as one of the fundamental key resources for archaeological research and analysis by hundreds of project members for the last decade. Its content consists of formal textual and numeric records in a set of 'excavation' tables and additional sets of 'specialist' tables maintained by each of the thirteen current specialist teams. Excavation and specialist tables are joined by virtue of the single context recording method: excavation units have unique IDs that are the central organizing principle for all data; all data relate to units.

The design of the current system began in 2004, when the process of artefact recording at Çatalhöyük underwent major revision. A multitude of stand-alone databases and spreadsheets that lived on the team members' desktop computers were migrated to a centralized system that would allow sharing of data among all researchers working at the excavation (Ridge & May, 2004; Jones, 2012a). The new system was partially implemented and users were trained on site in 2005 (Ridge et al., 2005), then built out further over the winter months and throughout the 2006 season (Ridge & Jones, 2006).

In order to make basic data accessible to all team members while incorporating different recording methods for particular specialisms over the life of the project, the design followed a core-specialist paradigm, resulting in what has been termed a 'defragmented recording model' (Ridge et al., 2005).

Rather than recording basic data for artefacts through the filter of a specialist's eye (who might seek to fit it into a preconceived model), initial entries about, for example, basic measurements, general descriptions, and simple classifications, reflect the excavator's view as closely as possible. These entries are available for any specialist. All teams begin recording by entering a unit description, where they can detail their thoughts on a unit. They also record team-specific, basic information about the research object, such as its measurements, weight, and condition. A core set of tables holds this inventory level data ('core data'), accessible and comprehensible to non-specialists. This practice is reflected in the database with table name suffixes, such as *_basic* or *_level_one*.

Recording then branches off into detailed information: data resulting from more extensive measurement, analysis, and the further interpretive classification that occurs when supported by the characteristics of the artefact (Figure 2). For instance the Faunal team can determine if an object is cranial or post-cranial and whether it worked or not. Naming conventions for the tables that hold information of these 'interpretive layers' in the central database were introduced to help make their contents more apparent for users' designing queries.

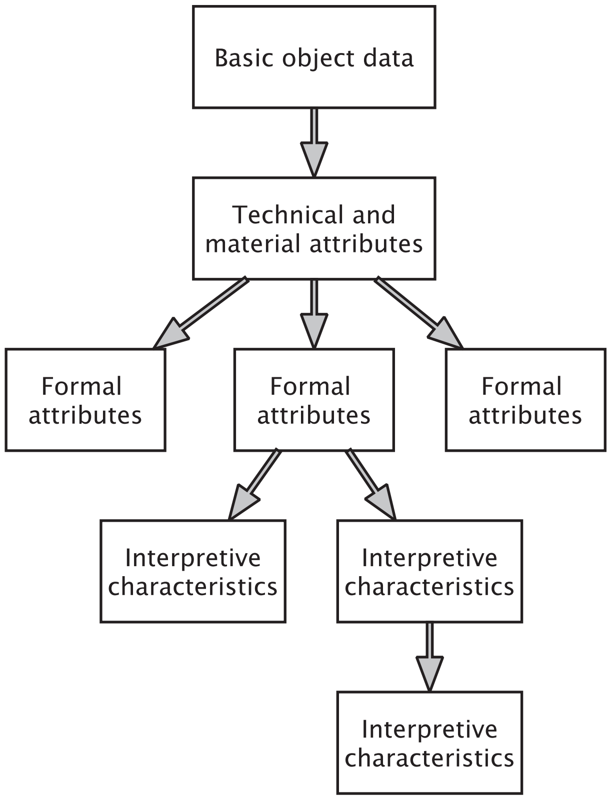

Figure 2. *Diagram showing the defragmented recording model (following figure 125 in Ridge et al., 2005: 260).*

As described earlier, each team logs into the database through their own, distinctive interface for data entry, which is tailored—and continuously adjusted—to the particular needs of their respective specialism. Members of the various teams can record completely different data about the same object, but are not exposed to interfaces onto the same base excavation data that are 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 codes and lists of values (LOV; e.g. for material or colour) 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).

Establishing a fixed set of artefact terminologies for the Finds register was foremost in the data-cleaning strategy when the centralized system was implemented (Cassidy, 2006). X-finds are recorded on a daily basis when they arrive at the Finds office. As the description of the artefacts had been free text, entries and spellings varied widely and abbreviations were used freely. 'For example, "potstand" was frequently entered as "potterystand" or "potterystant" […] obsidian became "obs.", pottery became "pot", […] comments such as "weird insecty thing" and "clay blob" were also fairly common' (Cassidy, 2006: 336). The Finds entry form now contains drop-down menus for material group, material subgroup, and object type, referencing LOV tables in the central database. There are twenty-five terms for material group (e.g. Human Bone, Metal, Phytolith, Plaster), thirty-eight for material subgroup (e.g. Wood, Textile, Daub, Mudbrick), and fifty-two for object type (e.g. Fragment, Hook, Scapula, Sherd).

With considerable foresight, links between the Finds and specialist tables were created in the database, so that, for example, while viewing a Ground

Stone X-Find in the Finds entry form, one can open its corresponding record in the Ground Stone entry form to find additional information provided by the Ground Stone team[3]. This link to the specialist table, along with a link to the data recorded by the excavation team, allows researchers to follow an artefact through all stages of its study: from the trench to the Finds office and on to specialists' labs. During this journey, the interpretation of the same object may differ, of course. For example, an excavator may believe an artefact to be a bone, but it may be identified as a clay object by the Finds officer. Once in the hand of the specialist, it may be determined a clay figurine.

The centralization of the data tables afforded the ability to create 'views' on the central database that span specialisms, teams, and areas. Views are dynamic tables based on queries in Access terms that are automatically updated and respond to any changes in the content of the tables their query is based on. Like virtual windows that let users look at selected data from one or more tables, they allow users to group materially unrelated artefacts by interpretative category and run customized cross-table queries tailored to specialist requirements. For example, 'team members interested in Building Materials can link from a feature to related units, find the heavy residue data for those units, and tie it together with the relevant diary entries' (Ridge et al., 2005: 264). Naming conventions for the tables in the central database were introduced to help make their contents more apparent for users' designing queries. An additional Access interface was provided to help team members create effective queries and save them for future use.

Finally, centralization of the excavation and specialist data in a database server also allowed direct access to its content from other databases (for example, GIS) or applications[4]. Potential application connectivity originally envisioned was to the project website, GIS, Excel, FileMaker, SPSS, student portfolios, and more (Ridge et al., 2005).

In the years following this major re-structuring, responsibility for the Çatalhöyük system has passed through the hands of several database developers. Nevertheless, it has undergone continuous development, including improvements to the usability of the interfaces, the introduction of further data integrity checks, and cleaning of data. In 2006 a new interface to the central database was made available through the public website, and during 2006 and 2007 a mechanism for exchanging metadata between the multimedia server and the database server was implemented, so images can be directly linked and viewed from their related records (Ridge & Jones, 2006; Jones, 2007). In 2013 the central database was migrated to a new and upgraded server machine, with no modification to users' usual direct interaction with the data.

Numerous references to adjustments and additions to the data recording interfaces can be found in the annual archive reports. These reflect the dynamic nature of the system, which constantly responds to the changing recording requirements motivated by new team leadership, changing recording priorities and new analysis practices. However, none of these represent any radical changes, so the current system still reflects fundamentally the design principles introduced in 2004.

Spatial data, multimedia, and texts

The spatial database mainly contains footprints of buildings, spaces, and units. It also holds a remarkable set of detailed representations of nearly four hundred human skeletons. Tabular attribute data from the central relational database are brought into the GIS database by connecting to a set of specially designed views. Access to spatial data is presently only possible using desktop client software (ESRI ArcGIS). In 2013, digital tablets were introduced for digital archaeological recording directly in the trench (Taylor & Issavi, 2013). The project was expanded to the entire excavation team during the 2014 season (Issavi & Taylor, 2014). Because the GIS database is self-contained, copies can be carried around on tablet computers, allowing digital images and spatial footprints of excavated features to be entered at the trench directly into the GIS database, thus forgoing paper forms.

During the field season of 2010 a group of researchers began constructing 3D models of the excavation using laser scanning, computer vision, photogrammetry, and feature tracking. The impressive results of this work are dramatically enhancing the spatial visualization of the site (Forte et al., 2012; Forte, 2014; also this volume, chapter 4). Due to the novelty of this approach an integration of 3D models into the existing infrastructure and digital workflows has not yet been attempted.

The majority of digital photography and video are managed on a dedicated multimedia server. The server includes a web service (NetPublish) that provides users access to files in the multimedia catalogues through a password-protected web page. NetPublish also serves imagery to the Çatalhöyük website. Images are also linked to the Conservation team's Access interface and site photos from the multimedia server

[3] References to the respective specialist core tables need to be entered manually and have not been recorded systematically.

[4] This requires the databases and applications to comply with the ODBC (Open Database Connectivity) standard and an ODBC driver that translates between the different systems.

are linked to the main Excavation Access interface via a button on each screen. A smaller portion of the multimedia assets are still stored on the file system and are not part of the database backend infrastructure. A subset of all recorded videos is stored on the web server machine and directly linked into the website. In the past, hand-drawn site sketches from unit sheets and the daily sketches that accompany the diary entries (see below) were scanned and uploaded to a folder on the file system. The tablet recording system rolled out in 2014 allows now to create site sketches digitally directly in a GIS-specific file format[5] that can be copied into the GIS database by the end of the day. Daily sketches are now drawn digitally on a photograph taken with the tablet camera and manually uploaded to the file system (Berggren, 2014b).

The most extensive free-text documents that reside in the current Çatalhöyük database are the diaries. They perhaps are most immediately linked to a reflexive practice. Since 1997 researchers at Çatalhöyük have been entering comments about the daily experiences in the main database via an Access entry form. During the 2012 and 2013 seasons we implemented changes in the diary tables and entry form to encourage a deeper involvement in the reflexive methodology. We added the ability to reply to a post, to view recent posts of other contributors, and to tag posts. We also added a prompt to convey expectations and objectives for posts in the diary form (Mickel, 2013). However this effort to mimic elements of a web blog was severely limited in its implementation by the software and existing data models (Berggren, 2013).

A daily sketch, consisting of an annotated photograph, serves as visual component of the diary. Although implemented in 2007 (Jones, 2007), it did not become part of the daily documentation routine for all excavators until 2013. Since digital images cannot be handled as objects by the existing database, only metadata about the sketch (file name, location on the file system, unit number, etc.) are entered into the database and can be searched by unit or feature number (Berggren, 2013).

Over the decades, researchers at Çatalhöyük have produced a wealth of publications in academic journals and books, including a series of monographs from the British Institute of Archaeology, that make use of the data archive in their analyses. In addition, an unknown number of secondary data and tables—stored at unknown locations—has been derived from the original material in the central database. The annual Çatalhöyük Archive Reports can be retrieved as pdf files from the project website, but they are not linked into the database. Thus, a considerable amount of research material exists that, while building on the contents of the central database, is not connected with the original data.

A notable exception is the final report on the excavations at the northern end of the East Mound directed by Ruth Tringham (Tringham & Stevanović, 2012). The publication is mirrored by an online version[6], a 'Digital Multigraph', which links original data, multimedia materials, analysis, and interpretations to the contents held in the printed edition in an open access, sharable platform.

Challenges

The major re-design of 2004 and 2005 described above and the ongoing development over a twenty-year span have mainly focused on the most immediate needs in coping with the amount of data produced at the field site and in the specialist labs during the field seasons. The current system is tuned primarily to putting data in rather than getting it out. There are currently eighteen data entry interfaces and just one query interface.

It is still rather cumbersome for the researcher to make use of the wealth of data held in the databases, which hampers analysis and particularly multidisciplinary work. The single query interface is not easy to use. For example, discovering all the grave goods associated with a particular burial is challenging at best. According to one of the former database developers, 'there are many times where I've admitted I would have to think and probably write some code to achieve the complicated result' (Jones, 2012a: 5).

In order to perform complex searches, researchers need familiarity with the Access software approach to constructing queries and deep knowledge of the available tables and attributes. At the backend, data are still more or less 'silo-ed' in fourteen sets[7] of mostly material-based team specialist tables. Each set of tables has distinctive field attributes and distinctive approaches to schema normalization. Specialist teams typically use specific code values, and looking up the meaning of those values poses an additional hurdle when constructing a query. It would require several queries into different databases to combine a geo-referenced outline of a building with imagery of its floor or walls and information about faunal or human remains in that building. One-to-many table relationships, which typically occur in normalized relational databases, contribute to the complexity of queries. For example, if a space has multiple associated

[5] ESRI shapefiles.

[6] http://lasthouseonthehill.org

[7] These are Archaeobotany, Ceramics, Chipped Stone, Clay, Faunal, Figurines, Ground Stone, Human Remains, Heavy Residue, Lithic, Microfauna, Phytolith, Shell, and Finds.

units, a query joining the unit table with the space table would result in as many repeating rows for the space as it has units. Researchers might prefer in this case to have a single row aggregating units in a computed array, but such queries require an uncommon level of expertise in the underlying query language (SQL). Additionally, as discussed above, several other databases hold important components of the research data that cannot be integrated easily, or they exist even outside any repository. Recently Mazzucato (2013) published a remarkably extensive, systematic overview and analysis of the wide corpus of the database, including spatial data, which demonstrates the potential of this resource if these obstacles can be overcome.

While the technology infrastructure of Çatalhöyük evolved along with the data, ever-changing specialist requirements, and the increasing number of different data types that needed to be accommodated, completely new technologies have emerged, which revolutionize the ways research data can be organized, stored, and analysed.

A Living Archive Vision

The Çatalhöyük Living Archive[8] was conceived in 2014 with two closely related goals: first, to ensure that Çatalhöyük data remain accessible and useful well beyond the duration of the excavation activity and second, to enhance and extend system functionality for current researchers, while honouring the project's history of practising reflexive archaeology. Excavation at the Çatalhöyük site will end after the summer field season of 2016, and analysis and data entry will continue through December 2017. At that time, a considerable volume of digital material will need to be archived as a permanent record of the project's activity and outcomes over a twenty-five year period.

Whereas traditional archaeological archives preserve downloadable copies of databases and associated files along with searchable metadata, a living archive will make the data itself directly accessible for the foreseeable future, in a sophisticated interface that permits both simple annotation and the creation of new analytic interpretive 'layers'. To support interpretative arguments, researchers will be able to reference specific sets of records in the database, such that others may view them along with relevant spatial and/or statistical visualizations. For example, one might find evidence supporting assertions that 'a clay cooking ball is actually a sling shot, or a house actually a shrine, or that not climate change but social tensions caused the abandonment of the settlement' (Hodder, 2014 personal comm.).

[8] http://catalhoyuk.stanford.edu

A Pilot Project

The Çatalhöyük Living Archive project has re-organized and published a significant proportion of the project's tabular records as LOD, and has built a distinctive web application providing new ways of viewing and analysing evidence. The design objectives for the web application were to (1) facilitate the re-interpretation of objects and re-assembling of their contexts at multiple scales; (2) allow presentation of such evidence in supporting new arguments from multiple voices; and (3) incorporate new interpretations as annotations upon the data store itself. In a third segment of the project, researchers developed a new network dataset and visualizations documenting and analysing team membership and knowledge production over the more than two decades of excavation and research. That work is presented in detail elsewhere in this volume (Mickel & Meeks, chapter 3).

Linked (open) data

Our approach to meeting the two project goals of permanent accessibility and improved capability for queries and visualization builds on two relatively new paradigms for extending the World Wide Web—the Semantic Web and Linked Data (LD). These closely related initiatives provide conceptual frames and technologies for connecting not only documents (web pages), but also structured information within documents and within publicly accessible triple stores. They are enabling new ways to find, share, re-use and combine information.

The Semantic Web is a global collaborative initiative, led by the W3C Consortium, that introduced the notion of tagging data within web pages such that the semantics of the data is machine readable and therefore linkable with data in other web pages. The model adopted for that tagging is the Resource Description Framework (RDF), which has a core structure of statements in a 'triple' form of <subject>,<predicate>, <object>. When extended with the RDF Schema, the model allows the embedding of a relatively simple computational ontology within the data itself, structuring and codifying assertions about the concepts, relationships, and constraints pertaining to it. This permits the semantics (i.e. the meaning) of data to 'accompany' it in a computational format, helping people and their machine agents manipulate, interpret, and integrate it. More expressive formalizations, such as the Web Ontology Language (OWL), build upon the RDF model.

Linked Data, a term coined in 2006 by W3C director Tim Berners–Lee, refers to guidelines for web data publication that prescribe the use of standards including RDF and SPARQL[9]. The other so-called ‘rules’ of LD publication are as follows: using unique Uniform Resource Identifiers (URIs) to denote things; using the HTTP web protocol as a means for de-referencing URIs; and including links to related things in your data. LOD is LD made freely available; in principle, LD could be behind a pay-wall; in practice virtually all LD are open.

Çatalhöyük in the LD cloud

We have published a sizable quantity of tabular records now held in the core Çatalhöyük relational database to an RDF triple-store (OpenRDF Sesame) accessible via a SPARQL endpoint. In this way, products of Çatalhöyük research have been added to the LOP cloud. These roughly 2.3 million triples describe some basic attributes of approximately 250,000 finds and their containing units, spaces, features, buildings, and areas. The classes of finds published so far include Human Remains, Faunal Bones and Artefacts, Microfauna, Ground Stone, Figurines, and Chipped Stone. Those tables holding more detailed measures and reflecting refined classifications for these finds—the ‘interpretive layers’ defined in our earlier description of the reflexive modelling strategy—are not yet exposed within this framework. Discussions concerning publication embargoes are under way.

Data have been published referencing RDF, RDF-S, OWL, and SKOS ontologies, with all current Çatalhöyük classes, relations, and vocabularies intact. In a future phase, all Çatalhöyük data will be re-organized according to one or both of two experimental archaeological ontologies in development elsewhere, CRM-EH and CRMarchaeo[10]. Publication of data in the context of a formal ontology will to an even greater extent help people and their machine agents manipulate, interpret, and integrate Çatalhöyük data with other computer applications and data stores.

Some Çatalhöyük data have also been made available via a pilot RESTful API in GeoJSON format[11], providing another means for integration and annotation. For example, the URI ‘http://catalhoyuk.stanford.edu/api/units/bldg/89’ entered manually into a browser or sent programmatically within software, returns basic data and spatial footprints for the twenty-nine units contained within Building 89.

Prototype Web Application

Development of the Living Archive application began with an analysis of the existing database, followed by the creation of a new and experimental partial copy. After soliciting ideas for functionality from team members, we designed and built a web interface for exploring many of them (Figure 3). These efforts and interim results are described briefly below.

Database study and re-organization

Over three hundred tables from the post-2013 field season SQL Server database were imported into a new PostgreSQL/PostGIS database for experimental re-configuration. PostGIS is a set of add-on libraries for the open-source PostgreSQL that provide advanced OGC[12]–compatible spatial functions. This enabled the conversion and import of a copy of the project GIS database as well, and a close coupling of spatial data with all attribute data for both finds and their spatial containers within the site: units, features, spaces, buildings, and areas. The generic local metric grid coordinate system of the existing GIS ([[0, 0], [1200, 1200]]) was converted to geographic coordinates (latitude, longitude) for mapping in ‘real-world’ context.

A detailed study was made of the elaborate Çatalhöyük encoding system incorporated in dozens of LOV tables, with the goal of understanding the database schema as the working ontology of the domain—the entity classes, sub-classes (or types) and relationships represented—and the degree to which formal relationships are made explicit in the schema. Although a number of controlled vocabularies are in effect, a substantial amount of classification information is held in free-text fields. It is also the case that each team has its own set of vocabularies, which, if harmonized or mapped to a central standard, would facilitate much richer cross-team querying than is currently possible. Remedying this was outside of the pilot scope, but the data modelling challenges are now well understood and will be addressed in the next phase.

Some experimental transformations *were* in scope for this phase, namely the de-normalization, or ‘graphification’ of a subset of data, to better support

[9] A recursive acronym standing for SPARQL Protocol and RDF Query Language. A SPARQL endpoint is a URI for a web service that interprets SPARQL queries against a triple-store.

[10] CRM-EH (Conceptual Reference Model-English Heritage), developed by the Hypermedia Research Unit at University of South Wales (http://hypermedia.research.southwales.ac.uk/resources/crm/), is an archaeological extension of the CIDOC–CRM ontology (*ISO* 21127:2006; http://www.cidoc-crm.org/). CRMarchaeo is a related research effort at the FORTH Institute of Computer Science.

[11] In this case, a web service (Representational State Transfer; Application Programming Interface) that answers queries formatted as URIs via the HTTP web protocol, returning records with spatial footprints and some attributes in GeoJSON format.

[12] Open Geospatial Consortium, the principal standards framework for geospatial data.

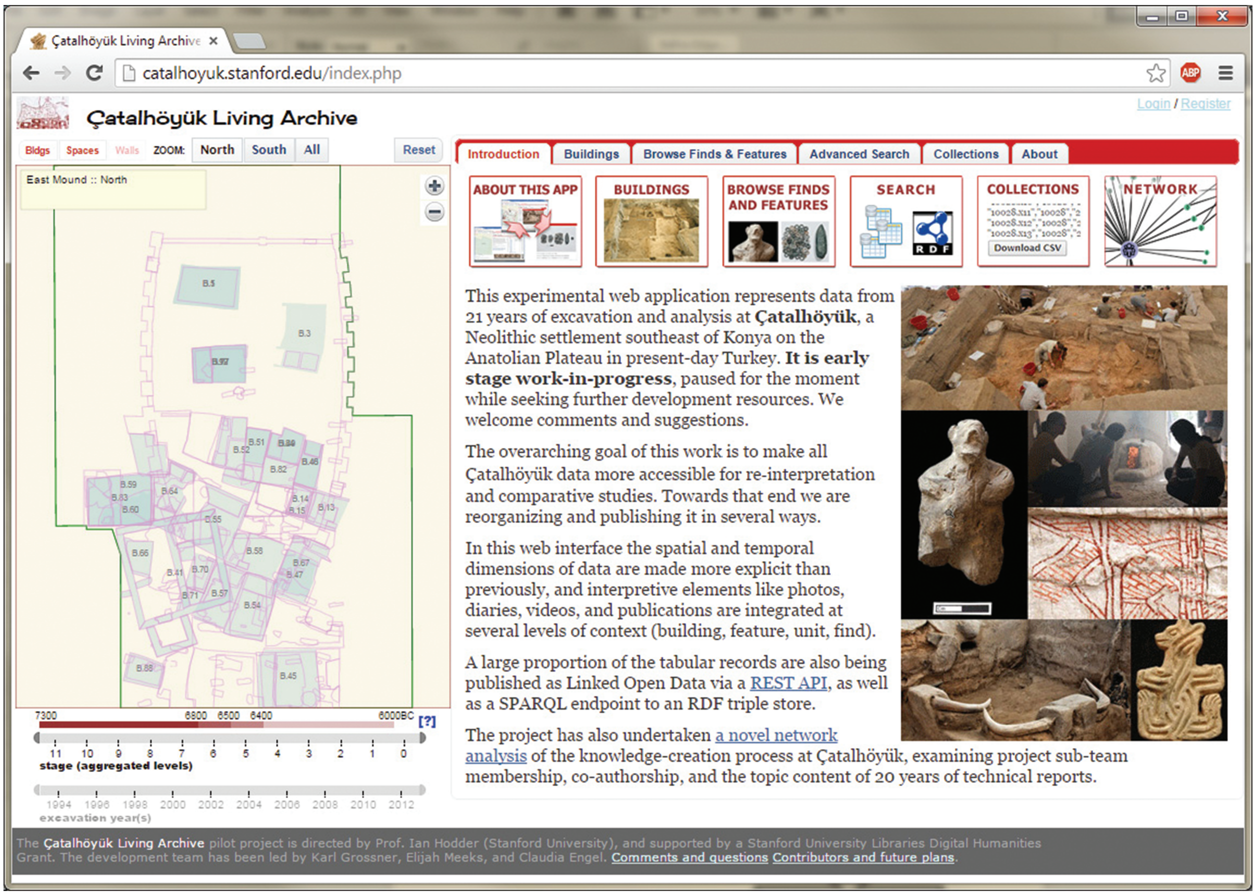

Figure 3. *The Çatalhöyük Living Archive web application (http://catalhoyuk.stanford.edu).*

queries in the new web interface, and to facilitate publication of data in RDF triple form. A new set of tables were generated for areas, buildings, spaces, features, units, and X-finds, aggregating information from LOV tables, or merging related fields (such as spatial footprints) previously joined in queries—effectively a new set of views on the database that simplify the scripted generation of RDF. A related step taken was creating a unified Finds table which joins basic attributes from records found only in separate team-specific tables. One self-imposed challenge at this stage was populating some of the data grids in the pilot web application from an RDF triple-store, in order to test the feasibility of a complete conversion from the relational model in the future archive, and to simulate the future linking of data from other sources.

Multimodal discovery

The new web interface facilitates discovery of patterns and processes by making the spatial and temporal dimensions of data more explicit and manipulable. Familiar browse and search functionality is extended with a map showing objects in their spatial context, and three interactive timelines that can be used to filter the map and data grids for deposition level, excavation year, and building occupation phase. Queries for features, object types, and materials can be made on spatial, temporal, and textual dimensions, as well as the classifications found in team-specific controlled vocabularies. Depending on the object type, query results might include tables, spatial footprints and point locations, diary entries, spatially tagged articles from the archive reports, photographs, video, and 3D models.

The principal sections of the web application group functionality for (1) building-centred browsing; (2) site-wide browsing; and (3) site-wide search. Browsing is accomplished with sortable data grids for units, X-finds, and features. The selection of one or more rows displays a spatial footprint (for units and features) or a marker (for X-finds) on the map. Considerable detail is available for each object, according to its type. For example, results can include data from field recording sheets, photos, videos, diary entries, related technical reports, heavy residue sampling data, summary visualizations of finds per spatial entity, and so on (Figure 4). In many cases, results can be filtered temporally by means of the three interactive timelines.

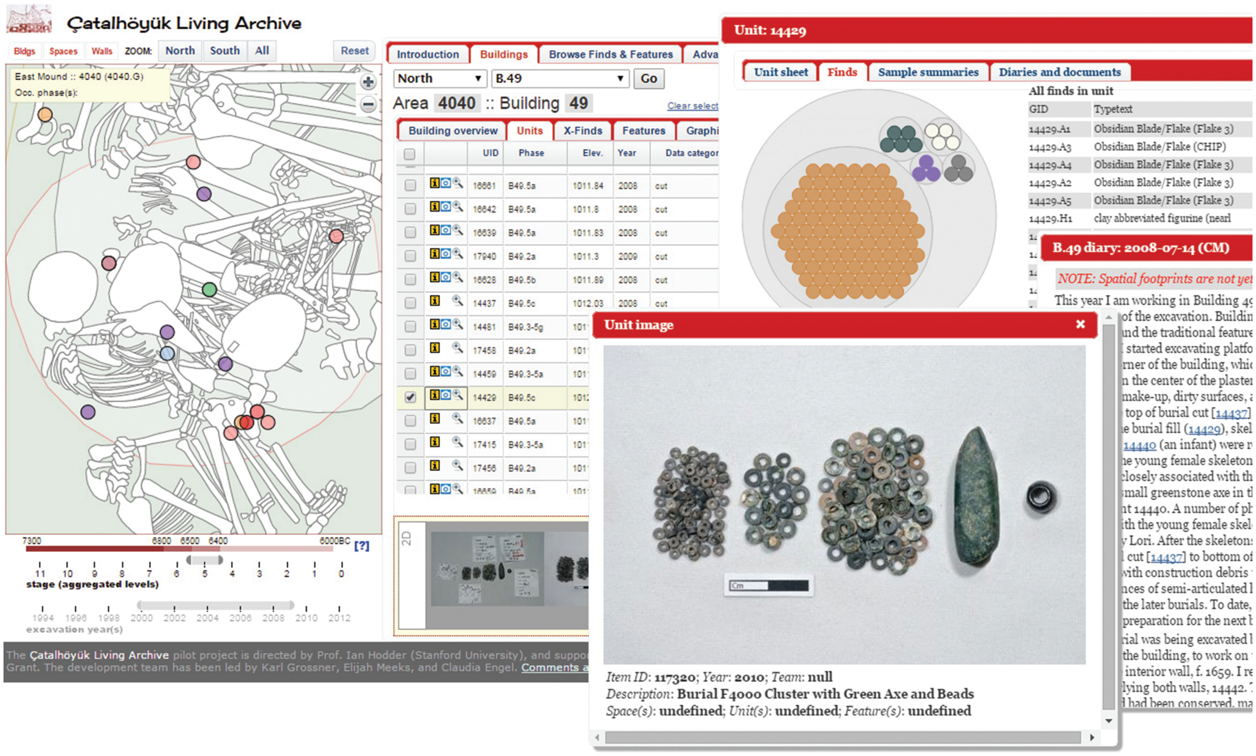

Figure 4. *Multimodal search and browse in a spatial-temporal browser, to reconstitute a burial.*

Two kinds of site-wide search have been prototyped: searches against individual tables within the central database, and searches against the RDF triple-store (Figure 5).

Cross-disciplinary analysis

With the conversion of data into a graph model, we are making it easier for researchers to broaden queries across all tables, thus facilitating cross-disciplinary analysis. One outstanding challenge is completing the development of a classification system that harmonizes vocabularies. Terms used for the same or similar materials or object types by specialists with different backgrounds can vary; queries should be able to locate records that are semantically related. In the RDF triple-store we have represented several existing vocabularies in the W3C standard SKOS[13] format frequently used in LD implementations. SKOS permits encoding of vocabularies as Concept Collections,

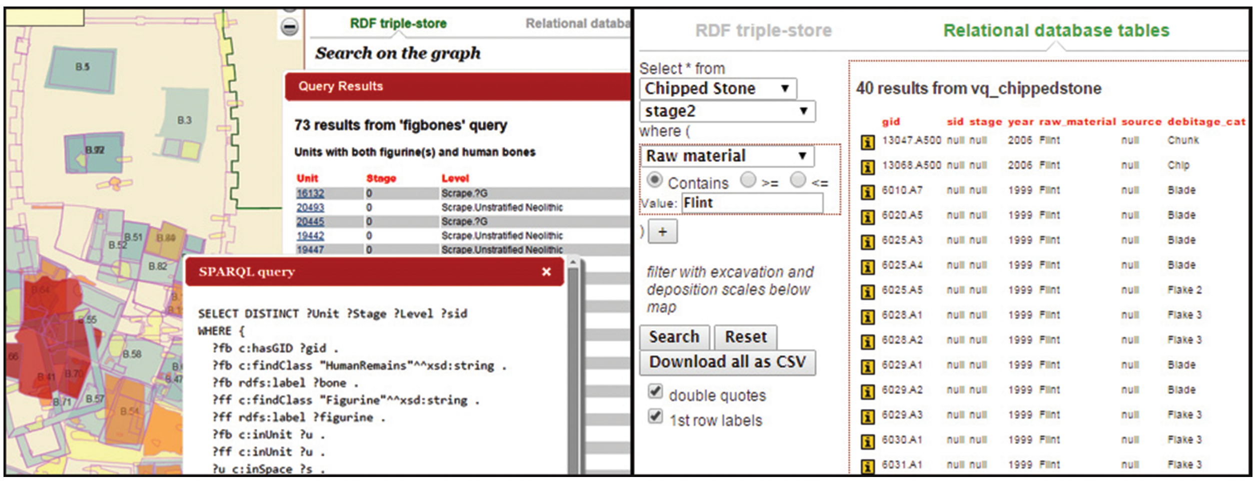

Figure 5. *Searching both the RDF store (left) and traditional relational database (right).*

[13] Simple Knowledge Organization System; http://www.w3.org/2004/02/skos/

with thesaurus-like relations such as *broader*, *narrower*, *related*, and *exactMatch*. In the next phase of work, classification terms with identical or similar meanings will be mapped to canonical vocabularies for materials and object types.

To test our experimental graph-based data model, we developed a set of informal 'competency questions' the system should be able to answer relatively easier than with a relational database.[14] Some of these tests are successfully met in the pilot application (indicated by *); others will require the full transformation to a computational ontology undertaken in the next phase of work:

- Which buildings have baby burials? (*)
- Where are beads, regardless of material (stone, bone, shell)? (*)
- Which spaces/units have both anthropomorphic figurines and human bones? (*)
- Where are the wall paintings? (*)
- What occupation phases have no oven or hearth? (*)
- Are there any wide-mouthed pots in North Level G?
- Which midden spaces intersect spatially (x, y) with earlier or later living spaces?
- What stone or clay objects have been found near burials, during which deposition phases, and what are their interpretive categories?

Integration of written works

We have spatially referenced several hundred technical articles from twenty-one years of Çatalhöyük Archive Reports (1993–2013), and made many of these available as relevant description, analysis, and interpretation concerning buildings, spaces, features, and units. The integration of written interpretive works with visual representations and tabular data represents a significant step towards development of interactive scholarly works supported by the entire data store. A topic model was created from the corpus, and used in the knowledge network analysis mentioned earlier, and detailed elsewhere in this volume (Mickel & Meeks, chapter 3). At the next stage of the Living Archive, we will be able to represent temporally indexed topic 'signatures' for areal features of the excavation, including spaces and buildings.

Widening the spatial scope

Our integration of the central relational database with the project GIS will also allow broadening the scope of analyses beyond the immediate spatial extent of the Çatalhöyük excavation. By converting existing spatial data from a local coordinate system to a global one, data can be displayed on ordinary web maps, and contextualized further with spatial data developed by Çatalhöyük researchers and by other projects, provided it uses a global coordinate system. Çatalhöyük residents were engaged in trade with surrounding regions, involving for example obsidian (Carter, 2011), chert (Nazaroff et al., 2013; Nazaroff et al., 2014) and mollusk shells (Bar-Yosef Mayer et al., 2010), and the understanding of source networks is part of Çatalhöyük research. We anticipate that our approach to data publication will enable more and better comparative studies including sites elsewhere in Anatolia and further afield.

Annotating evidence

The pilot web application demonstrates rudimentary functionality for the recording and sharing of annotations, with an individualized 'My Collections' feature. Registered users are able to save, annotate, and export collections (query result-sets of interest), which can be kept private or flagged as public. In the future, annotations of individual records will be possible as well. Together with collections, these will form a discrete new layer within the data store, and provide a means for proposing new or alternate interpretive classifications. In this way, evidence assembled during the research process supporting a particular interpretive argument can easily be made available to others for evaluation and comment.

Conclusion

The Çatalhöyük project team is large and dispersed. It assembles only once per year over a period of several weeks, when excavation and recording are a priority and little time is available for discussions about team-wide data management and analysis strategies. Over the years there have been periodic difficulties obtaining the resources necessary for new development, upgrades, and maintenance on digital systems. Despite these challenges, Çatalhöyük team leaders and data system professionals have been consistently innovative and remarkably successful in integrating digital methods into research practice.

Technology was considered instrumental in furthering the post-processual methodology. As Hodder predicted: '…it is to be expected that in trying to operationalize these concerns [reflexivity, contextuality, interactivity, multivocality] in archaeology, the technologies themselves come to play a central role' (Hodder, 1997: 7). For the entire duration of the

[14] This term is due to Uschold & Gruninger (1996) who describe methods for evaluation of computational ontologies.

project, the most innovative tools available have been employed alongside the research, from the early launch of a public website to the extensive use of multimedia, and from the implementation of purposefully designed databases to mobile technologies, 3D scanning, and remote sensing.

Çatalhöyük database systems have made the recording and interpreting processes themselves a part of the excavation record, particularly within the past ten years. Layers of interpretation are explicit within the record, although not yet fully visible or actionable. The Çatalhöyük Living Archive will ultimately expose this tiered system fully, for both the researchers within a given specialism, and—a much harder task—for multiple audiences outside it.

Acknowledgements

We thank the following colleagues for their contributions to this work: Ian Hodder, Arkadiusz Marciniak, Elijah Meeks, Allison Mickel, Justine Issavi, Lynn Meskell, Carrie Nakamura, Christopher Knüsel, Maurizio Forte, Camilla Mazzucato, Adam Nazaroff, Jason Quinlan, Margaret Tomaszczuk, and Kristy Henrich.

References

Bar-Yosef Mayer, D. E., Gümüş, B. A. & İslamoğlu, Y. 2010. Fossil Hunting in the Neolithic: Shells from the Taurus Mountains at Çatalhöyük, Turkey. *Geoarchaeology*, 25(3):375–92.

Berggren, Å. 2013. Reflexive Methodology [online]. In: *Çatalhöyük 2013 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2013.pdf

Berggren, Å. 2014a. Development of Reflexive Approaches in Archaeology. In: C. Smith, ed. *Encyclopedia of Global Archaeology*. Berlin: Springer Verlag, pp. 6249–57.

Berggren, Å. 2014b. Reflexive Methodology and Recording [online]. In: *Çatalhöyük 2014 Archive Report* [accessed 8 January 2015]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2014.pdf

Carter, T. 2011. A True Gift of Mother Earth: The Use and Significance of Obsidian at Çatalhöyük. *Anatolian Studies*, 61:1–19.

Cassidy, J. 2006. Finds System 2006 [online]. In: *Çatalhöyük 2006 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2006.pdf

Forte, M. 2014. Virtual Reality, Cyberaerchaeology, Teleimmersive Archaeology. In: F. Remondino & S. Campana, eds. *3D Recording and Modelling in Archaeology and Cultural Heritage: Theory and Best Practices*. Oxford: Archaeopress, pp. 113–27.

Forte, M., Dell'unto, N., Issavi, J., Onsurez, L. & Lercari, N. 2012. 3D Archaeology at Çatalhöyük. *Journal of Heritage in the Digital Era*, 1(3):351–77.

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

Hodder, I. 1999. *The Archaeological Process: An Introduction*. Oxford: Blackwell.

Hodder, I. ed. 2000. *Towards Reflexive Method in Archaeology: The Example at Çatalhöyük*. Cambridge: McDonald Institute for Archaeological Research/British Institute of Archaeology at Ankara.

Issavi, J. & Taylor, J. 2014. Tablet Recording Overview [online]. In: *Çatalhöyük 2014 Archive Report* [accessed 8 January 2015]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2014.pdf

Jones, S. 2007. Çatalhöyük Database Development Archive Report 2007 [online]. In: *Çatalhöyük 2007 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2007.pdf

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

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

Mazzucato, C. 2013. Sampling and Mapping Çatalhöyük. In: I. Hodder, ed. *Humans and Landscapes of Çatalhöyük: Reports from the 2000–2008 Seasons*. Los Angeles: Cotsen Institute of Archaeology Press, pp. 31–64.

Mickel, A. 2013. Diaries, 2013 [online]. In: *Çatalhöyük 2013 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2013.pdf

Nazaroff, A.J., Baysal, A. & Çiftçi, Y. 2013. The Importance of Chert in Central Anatolia: Lessons from the Neolithic Assemblage at Çatalhöyük, Turkey. *Geoarchaeology*, 28(4):340–62.

Nazaroff, A.J., Baysal, A., Çiftçi, Y. & Prufer, K. 2014. Resilience and Redundance: Resource Networks and the Neolithic Chert Economy at Çatalhöyük, Turkey. *European Journal of Archaeology* [online] (25 November 2014) [accessed 7 May 2015]. Available at: http://dx.doi.org/10.1179/1461957114Y.0000000081

Ridge, M. & Jones, S. 2006. IT Team [online]. In: *Çatalhöyük 2006 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2006.pdf

Ridge, M. & May, R. 2004. Database & IT Developments [online]. In: *Çatalhöyük 2004 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/archive_reports/2004/ar04_32.html

Ridge, M., May, R. & Mackie, D. 2005. Çatalhöyük IT Team Archive Report 2005 [online]. In: *Çatalhöyük 2005 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2005.pdf

Taylor, J. & Issavi, J. 2013. On-site tablet recording [online]. In: *Çatalhöyük 2013 Archive Report* [accessed 30 December 2014]. Available at: http://www.catalhoyuk.com/downloads/Archive_Report_2013.pdf

Tringham, R. & Stevanović, M. eds. 2012. Last House on the Hill: BACH Area Reports from Last House on the Hill: BACH Area Reports from Çatalhöyük, Turkey. Los Angeles: Cotsen Institute of Archaeology Press.

Uschold, M. & Gruninger, M. 1996. Ontologies: Principles, Methods and Applications. *The Knowledge Engineering Review*. 11(2):93–136.