Last House on the Hill: Digitally Remediating Data and Media for Preservation and Access

Michael Ashley¹, Ruth Tringham², Cinzia Perlingieri²

¹ Cultural Heritage Imaging, San Francisco, California, USA
² University of California, Berkeley, California, USA

Abstract
The idea of embedding, interweaving, entangling and otherwise linking the data and media from archaeological excavations with their interpretation and meaningful presentation in an open access sharable platform has long been an ambition of those of us working in the digital documentation of archaeological research and the public presentation of cultural heritage. Formidable barriers still exist to making it possible for projects to achieve these aims, ranging from intellectual property concerns to providing commitments to the long-term sustainability of the digital content. Working in collaboration with the contributors, archaeological project managers, publishers and information technologists, we devised a content licensing agreement that makes it possible for the primary research media and data, combined with the monograph texts, to be freely and openly accessible in perpetuity.

The aim of our project, Last House on the Hill (LHotH), is to holistically reconstitute the rich multimedia and primary research data with the impressive texts of the monograph, the printed final report of the Berkeley Archaeologists at Çatalhöyük (BACH) project, in which a team from UC Berkeley excavated a group of Neolithic 9000-year old buildings at this famous cultural heritage location in Central Anatolia, Turkey. The Last House on the Hill brings together the published text, complete project database (including all media formats such as photographs, videos, maps, line drawings), related websites, data and media outside the direct domain of the BACH project, and recontextualised presentations of the data as remixes, movies, and other interpretive works by BACH team members and many others. We are achieving this through an event-centered, CIDOC-CRM compatible implementation ontology, expressed with the open source Omeka web-publishing platform, providing open access, transparency and open-endedness to what is normally the closed and final process of monograph publication. This paper describes the strategy, goals, architecture and implementation for the project, emphasizing the novel and innovative approaches that were required to make the project successful.

Categories and Subject Descriptors: Publishing, metadata, semantics, long-term preservation and access.

1. Introduction

In his commencement speech at Arizona State (ASU), U.S. President Obama reminds the 2009 graduating class that life is not measured only in accomplishments, but by perpetuating actions: “That is what building a body of work is all about - it’s about the daily labor, the many individual acts, the choices large and small that add up to a lasting legacy. It’s about not being satisfied with the latest achievement, the latest gold star - because one thing I know about a body of work is that it’s never finished. It’s cumulative; it deepens and expands with each day that you give your best, and give back, and contribute to the life of this nation. You may have set-backs, and you may have failures, but you’re not done - not by a longshot.” [Zim09].

Archaeological fieldwork produces a fragile legacy, requiring remarkable attention and effort in order to assure that our many individual acts will be sustained into deep time. Access to our legacies, analog or digital, are not assured, not by a longshot, so long as the de facto preservation standard for publication are synthetic accounts in the form of monographs. The sharing and the preservation of human traces digitally, coherent access to these traces for future generations [Ash08], depends on a coherent reckoning of all of the evidence streams from our practice, not just the hand-picked pieces chosen for their aesthetics or relevance to the particular assertions we make in the authoritative texts of the final field reports. As challenging as it continues to be, our discipline requires us to not be satisfied with a definitive accounting as told by a select few, rather to present to our present and future audiences a full, unmitigated documentary of the choices, both large and small, that led to our conclusions about the archaeological past.

It became more and more frequent in the late 1990s and 2000s to supplement the written text of a scientific published report or a textbook with a CD-ROM or (later) DVD-ROM that was slipped into an envelope inside the back cover of the volume. Such CD-ROMs accompany most of the preceding volumes of the Çatalhöyük Research Project, specifically those reports of the 1995-1997 excavations of the 9000-year old Neolithic settlement mound in Turkey. They contain supplemental tables, texts, and figures, as well as videos that would either have been
left out of the printed version because of space limitations or because of format limitations (e.g., color photos, videos). The supplemental CD-ROMs were useful because they meant that archaeological publication could transcend publishers' restrictions on details of data and media. But they proved to be a librarian's nightmare, since they were difficult to catalog and very easily went missing.

2. The Paper Volume and its Digital Mirror

The UCLA Cotsen Institute of Archaeology, who will be publishing our printed report of the Berkeley Archaeologists at Çatalhöyük (BACH) entitled House Lives, have decided to give up on supplemental material (CD-ROMs and DVD-ROMs) in favor of on-line digital versions of their publications. We, as archaeologists who have long been involved in digital documentation and publication of archaeology and cultural heritage were delighted with this news. We felt that these media had a limited lifespan, not only because they could easily go missing, but also because of their regular need for migration due to physical degradation and the inevitable (unless carefully archived) obsolescence of their software. We were also aware of the disadvantages of read-only CD-ROM media, as with the printed word, of offering a definitive closed narrative, whereas we feel that the narratives about archaeology, history, cultural heritage, and the past - including those about the data themselves - should be anything but closed, but should always be open for expansion, critique, and modification.

The online digital mirror of House Lives that is presented in this paper - entitled the Last House on the Hill project - is one which goes much further than to bring together supplemental materials along with digital versions of the published texts. Its ambition, one which we have long wished to satisfy, is to embed, interweave, entangle and otherwise link the data and media from the archaeological excavations with their interpretation and meaningful presentation in an open access, sharable platform [WT00, JT07, TA01, Tr04]. The project brings together the published text, complete project database (including all media formats such as photographs, videos, maps, line drawings), related data and media outside the direct domain of the BACH project, along with recontextualised presentations of the data. We are achieving this through an event-centered, CIDOC-CRM compatible implementation ontology, expressed through the open source Omeka web-publishing platform, providing access, transparency and open-endedness to what is normally the closed and final process of monograph publication.

The architecture and content management practices with which the Last House on the Hill has been constructed will act, we think, as a model and an encouragement for our archaeologist colleagues to share their work with the public for the long-term. Our attitude to sharing our knowledge with the public in which we make the process of our archaeological interpretation transparent in order to engage them more intensively in our work, and our attitude to breaking the strict bondage of the empirical data is, we feel, very close to that of the Çatalhöyük team as a whole, who have made all of their data and media accessible through Creative Commons 3.0 licensing. The digital expression of our work is able to be richer, more colorful, and more engaging than a black-and-white publication, and certainly represents a more intricate and entangled expression of what we do and how we think. The narratives about the BACH excavation have been built out of the rich body of data and media that are available and accessible in this digital on-line version. It is an open-ended data stream that can grow and - as long as it is well curated – can live for many decades.

The paper volume is in essence the sixth major publication by the Çatalhöyük Research Project. It comprises a record of the BACH from 1997 to 2003. The first volume of the Çatalhöyük Research Project related to the initial phase of surface work 1993-95 [Hod97]. The second volume focused on issues of the reflexive methodology [Hod00]. The third, fourth and fifth volumes present the results of excavations of three areas at Çatalhöyük, known as South, North and KOPAL, between 1995 and 1999, and their analysis and interpretation [Hod05a, Hod05b, Hod07]. This monograph presents the excavation results of the BACH project and their analysis and interpretation in one volume. Twenty-three authors contributed to the volume, many of which were participants in the BACH project for all of its seven years.

As with the other Çatalhöyük volumes, the results of the excavations (chapters 1-5) are followed by chapters that present the analysis of excavated data (chapters 6-19), which are followed by chapters of synthesis and interpretation (chapters 20-22). However, in keeping with the reflexive methodology we have attempted throughout the volume to avoid the separation of the presentation of data from their interpretation. There are two color photographs in the volume. The rest are black and white versions of originally color photos. There will probably not be an index. Appendices of lists and detailed data are not included in the printed volume. Nor are complex tables and figures such as the Harris matrices that document the stratigraphic position of each of the over 2000 excavation units. It is these limitations, the result of cost and produced by the text media format, that we have overcome through the LHoH project.

3. The Mind Map

A mind map (also known as ‘cognitive map’) was used in the planning phase of the project to record our shared concepts and ideas, and to better represent the holistic and non-linear reasoning about the existing documentation and the relationships to be created between the different sources of information. The archaeological project itself was considered the central node, the original event that brought together people and things and gave origin to the complex network of information about the project. Five branches were created around the central concept in order to link up all the existing documentation, but also to open possibilities for remediating and remixing, and to allow new strategies for access, visibility and usability of this rich and multiform content.

© The Eurographics Association 2009.
Figure 1. Mind Map of LHoH.

The existing databases and the online digital mirror of the book represent two of the main branches of the map. Data from the excavation, artifacts, media and the results of lab analysis are represented in numerous sets of different data sets that are managed by a CIDOC-CRM compliant Digital Asset Management system (see below). The branch for "remixes" represents the tools and plugins available on the web-platform that allow users to remix content, develop online exhibits, or build special web pages and showcases, combining digital objects in the archive with new narratives. The "media" branch represents the multiform digital content that several initiatives around the BACH project have brought online and that integrate the content of the book.

4. The Publishing Platform

Our intention for the digital version of the volume is to holistically reconstitute the rich multimedia and primary research data with the impressive texts of the monograph. The requirements for an online, collections-based web-publishing platform are formidable due to the complexity and sheer mass of data and media we wish to reconcile. The paper volume comprises some 22 chapters, 327,000 words, 1161 pages, 602 figures, plus tables, references and appendices. The primary database contains millions of records and hundreds of tables, in various formats and states of completion. The media database consists of over 15,000 images, hundreds of video clips, plus CAD drawings, illustrations, sketches and plans. Furthermore, we do not want to merely build a repository of content, but an extensible framework through which the researchers, visitors and future scholars that make up the Çatalhöyük community can all make substantive contributions. Some of our key requirements include:

• Open Data / Data Portability - Underlying data must be in a transparent and sustainable format.

• Open Source - Platform should be open source and community design driven, with a solid development roadmap and sustainability plan.

• Non-proprietary, free of protection mechanisms and external dependencies - Platform and data should be free of archaic security protocols or require other software/hardware in order to work.

• Open Rights Management - Patents and copyright constraints can severely curtail digital content sustainability. We want anyone to be able to openly publish and remix the content of LHoH. Wherever possible Creative Commons licensing is used to disambiguate how content can be used.

Figure 2. Last House on the Hill site built in Omeka.

For the platform, we would prefer to use a LAMP-stack (Linux OS, Apache server, MySQL database, Perl/Python/PHP scripting) solution that can be installed and supported by virtually any hosting company [Wik09]. The system would need to support robust mechanisms for bulk editing, batch media processing, and open metadata standards. For all of these reasons, we opted for the Mellon Foundation supported, open source and community developed Omeka platform [Ome09].

Omeka provides several user interface metaphors that are perfectly suited to presenting and interacting with rich media content. A collection is a full archive of material, related by event. An exhibit is a selected set of materials from a collection, related by theme, topic or other curatorial decision. Collection administrators and collaborators can contribute, annotate, organize and manage the content using various plug-ins and tools in the Omeka back-end. End users can do almost all of these things in the front end through the MyOmeka plug-in, designed to make the process of creating an exhibit or selecting materials incredibly simple to do.

5. Unburying the Past Semantically: Feature 634

As a test case by which to measure the effectiveness of the Last House on the Hill in user engagement and research potential over the paper format House Lives, we have chosen to focus on a single feature: Feature 634 – the burial of a mature woman under the north-central platform (F.162) of Building 3 in the BACH Area at Çatalhöyük.
This same platform in fact was cut by three additional burials whose pits, together with that of Feature 634, broke almost the entire surface of this low max 25 cm high 2x2 m. plastered platform. Such group burials under occupation floors were known from many other Neolithic buildings at the site, but this was the only one in Building 3.

Feature 634 is described in a number of chapters in House Lives, especially the excavation report (Chapters 2 and 4) and the Human Remains report (Chapter 11a), in 25 images (few of which pertain only to F.634). It also figures indirectly in other chapters because of its contained basket phytoliths (Chapter 10), flaked stone (Chapter 15) and beads (Chapter 17) and the practice and personnel involved in excavating the burials (Chapter 3).

When these texts are integrated online with the full inventory of original data (unit and feature sheets, data from flotation, micromorphology and other soil samples, complete skeletal sheets, 600 high resolution color images, over 40 video sequences, QTVR movies, 11 detailed field drawings, as well as detailed diary entries throughout the two years (2000 and 2001), the increase in the possibility for the public (whether professional or laymen) to understand, evaluate, and interpret the investigation of this feature is enormous. This potential grows logarithmically not only when we realize that these data and media are just one part of similar datasets for all the other spatial elements of the BACH research project, but also that these data and media are integrated and made accessible with the architecture described above, and with standards and protocols that enable its utilization for the long-term.

Archaeological documentation consists of a series of "contextual objects" that are datasets, media, reports, maps and all the information objects created to represent the archaeological knowledge. This critical mass of information gains its full significance if represented in all its aspects and relationships between people, actions, circumstances, time and space that altogether contribute to the creation of the information.
with the heterogeneity of the representation, the diversity of the material and media, and with the fact that "virtually nothing can be understood or interpreted without its relation to a context of thousands of other directly or indirectly related information assets" [SAC’06].

The CIDOC-CRM ontology is one of the best tools available to date to define the underlying semantics of database schemata and document structures used in archaeology [DK00], [Cid09]. The classes and properties that CIDOC-CRM provides to explain the logic behind a database structure, to explain a database content, as well as to describe the relationships between different datasets related to the same object or event, have proved to be very effective. And yet, the use of CIDOC-CRM in archaeological practice is still very limited, due in no small part to the required complexity of the CRM mapping in order to represent archaeological events.

![Diagram](image_url)

**Figure 6.** Implementation framework.

Although the abstraction of archaeological information in small and simple properties and relationships is totally possible, it is still far from being accepted by archaeologists as a standard practice for archaeological documentation, and standards are still very far from being adopted in archaeological research practice. Proof of this is the scarce amount of CRM-encoded original archaeological data available on the Internet or in any other open form.

### 6.1 Modeling the Ontology and the Challenges of Ambiguity

Archaeologists tend to consider databases as limited structures for their "textual thinking" or their "interpretive thinking". Fields like Description, Notes, Comments, Interpretation, Discussion abound in archaeological databases. Structured archaeological information is commonly non-standardized, even within the same project. Often different categories are used to describe the same things, and shortcuts are used that are sometime very difficult to interpret. Not secondarily, archaeological databases make undifferentiated use of categories like object, object abstraction, object appellation and identifier, and object representation, which are critical in the CIDOC-CRM reasoning. In the Çatalhöyük Research Project database, expressions like "Feature", "Feature ID", "Feature Name", "Feature Code" all express the identification name or number of a feature; fields like "Name" "Initials" "Entered by" or simply "By" all give us information about the actors involved in the project, and depending on which data set contains these fields, the actors’ roles change. A similar situation is common also for fields related to dates and chronology of events and to location through spatial coordinates. To conceptualize an ontological model for LHotH we rely on substantial and clear methodological guidelines elaborated at the beginning of the excavation that has been used as the base for establishing the relationships between the different data sources.

The aim of the LHotH project is to harmonize and make available the final excavation report as a comprehensive digital resource enriched by related data archives and by the massive media databases produced over the years. To link media, contexts, objects, events, and people we needed to link all sources starting from the archaeological context and excavation method used. Our need was to develop an explicit path for the relationship between the find and the context, so that rather than storing data as attributes of either the object or the context, it should be possible to store data linked directly with the events to which they relate. Therefore we needed a data structure that could not simply map, in the sense of replicate, the existing data items, but an enriched and complete new dataset that could represent a more conceptual and enlarged structure. This work has been conducted largely referring to the experience done at the English Heritage’s Center for Archaeology (CfA) and the principles established over the time by the team that developed it [CGF*04], [MBT08], even though we believe that some further reasoning on how the archaeological deposits and the stratigraphic relationships should be represented in the CIDOC-CRM schema is still needed.

The knowledge domain to be modeled was identified in the BACH excavation project and two groups of events were identified as the cores of the model: the events in archaeological time and the following formation of the archaeological record, and the recent events linked to the BACH excavation project. Starting from the excavation guidelines established by the main project, our first step was to extrapolate from all the different sets of data the information and structures that were then identified as Classes and Subclasses, then we created the main relationships to connect the data and sets of data (Properties). The schema was developed as simple horizontal triplets of class-relationship-class to create the chain of connections. We were very careful during the mapping process not to use shortcuts by simply extracting fields/classes from the existing database structures, instead we created more complete paths to represent the meaning, sometimes implicit, of the data structure, not to lose any of the complexity that we wanted to represent. In the first phase of the project the model was represented in a simple spreadsheet to be easily shared within the team of archaeologists, ultimately fully visualized in the Protégé ontology editor.
The context of the excavation has been constructed to give emphasis to the archaeological depositional context rather than to the event or action of excavating it, and it has been represented by the recursive pattern E53 Place - P 89 Falls within - E53 Place. This pattern has been used not only to represent the site sub-divisions, i.e. the spatial conceptual categories linked to the research methodology (area, space, trench, level), but also to represent the site and its depositional components of volumes and surfaces (stratum, fill, pit, floor). Although the depositional components are de facto physical things, our concern was to represent them as they are intended in the stratigraphic system and, in our opinion, they are much better represented by the CIDOC-CRM concept of "place". A principal goal of the Harris' stratigraphic system is to avoid temporal assumptions during the excavation phase and "freeze" in a mathematical scheme the stratigraphic units and their relative place in the stratigraphy together with their spatial relationships [Har89], [MBT08]. In archaeology, strata are physical things that need to be described in their physical characteristics of material, color, texture, but are also places that may host other features like pits, holes, and of course findings. This is still a challenging problem using the CIDOC-CRM, as E53 Place and E18 Physical Thing don't have shared properties to represent this "duality". In fact, in the ontological model published by HE's CfA, there is an added property 'EH_P3: occupied' that solves the problem of linking the hosting context (E53 Place) and its materiality (E18 Physical Thing) opening the possibility of a physical description [CGF*04].

The experience with LHotH confirmed that a semantic mapping very well meets the need to integrate all data produced by the various disciplines that contribute to an archaeological project. The CIDOC-CRM was an obvious choice be it an ISO standard for event based modelling of Cultural Heritage information. The high level conceptual categories of the CIDOC-CRM proved, in fact, to be very relevant to the complexity of this archaeological project and its related activities, although more remains to be done to make it easily applicable during the archaeological process. More work is needed on the ultimate use and form of an ontological stricture once it has been built.

7. Implementation and Architecture

The immediate aim of the LHotH architecture is to implement a lightweight but effective semantic mapping in order to make it possible to define portable data streams that can be easily moved, transformed and repurposed without losing their underlying meaning. This is the goal of the data portability movement, which has arisen as the walled gardens of social networking platforms such as Facebook and Twitter have demonstrated new challenges in this domain [Biz09]. Our implementation ontology elevates all media and data elements to the same level of relevance, and makes it technically painless to curate the entire universe of rich media and data despite its enterprise level complexity and sheer mass. Data portability assures that security and privacy concerns are met while providing remarkable flexibility to how data streams can be managed and structured.

The LHotH architecture adheres to international standards and practices (CRM compliant), but also inherits many convenient benefits the open frameworks we have implemented. For example, digital objects imported into Omeka can be cited easily cited using Zotero, and their metadata is OAI-PMH compliant, and can be readily harvested [Ome09]. The actual media objects can reside in distributed repositories because the harvesting works both ways (on import and through publishing). Furthermore, end users can create their own exhibitions and collections, using the community-developed Exhibit Builder plug-in. All of these truly Web 2.0 services facilitate sharing and perpetuate preservation through dissemination. At the same time, the LHotH data universe remains intact and is enriched through its use and reuse.

8. The Concept of "Database Narratives"

Steve Anderson recognizes two directions in which historiography has embraced digital technology. On the one hand is the idea of amassing the "total" historical record of events through accessible networked interoperable databases, creating history that is as definitive as possible. On the other hand "digital technologies have enabled strategies of randomization and recombination in historical construction resulting in a profusion of increasingly volatile counter-narratives...and histories with multiple or uncertain endings" [And]. Both these practices are legitimate pathways to creating narratives about the past. The idea of recombinant history, however, has been a guiding principle of the Last House on the Hill project, reflecting the fragmentary open-ended nature of memory and history drawn from a database with a structured architecture.

We (or our users) can draw out of our database limitless such fragments - narratives that are drawn out of the data, some being sparked by observing patterns in the spatial or chronological occurrence of artifacts in the excavation record, others by juxtaposing conversations on a video with narratives from a text, or by images with diary entries. Some of these narratives are descriptive or informative, following Anderson's (in press) first historiographic option - what Bolter and Grusin refer to as 'respectful remediation' [BG99], in which texts such as those in House Lives are enriched by use of digital photography or video in a pleasing design, but with little other creativity. Such remediation of the BACH Area (including Feature 634), can be seen by the examples of Remixing Çatalhöyük, a presentation to the European Archaeological Association in 2002 on using the Harris Matrix in platform Feature 162, and the BACH Area website. In the Last House on the Hill, the architecture of the data and media also encourage a more 'radical remediation' [BG99] which encourages a user to critique and improve on other media (and narratives) in the process of mediation when an explorer of our data re-uses and re-contextualizes our excavation products to create in their own narrative an alternative interpretation or 'reading' of the data, challenging the apparent authority of the 'expert', even using such tools as irony and satire [JT07], [TAM]. We have already created some 'remixes' of this kind that are based on Feature 634; RAVE: Requiem for a 'boneyard'; Dido's Lament: Remixing Dido's Lament;
the F.617 movie; time-lapse excavation of F.617 by the Science Museum of Minnesota as part of a multi-year National Science Foundation grant.

We have shown in our test case – the burial Feature 634 – that one relatively small feature can generate an enormous amount of data and documentation in a variety of formats. There are limitless strands that can be followed to explore these data about of people, places, things, and events that are entangled in Feature 534, and to use them creatively to produce new narratives or readings: the Baskets, the bodies, the stratigraphy representing life-history, the practices that surround treatment of the dead, the people who specialize in their excavation, the procedures of their excavation, the story of Dido with her broken ribs and dislocated hip, and so on. In Chapter 22 of House Lives: the Senses of Place at Çatalhöyük – We could do little more than suggest such narratives as these that might go beyond the strict empirical boundaries of the excavated data as tantalizing but so on. In Chapter 22 of House Lives: the Senses of Place at Çatalhöyük – We could do little more than suggest such narratives as these that might go beyond the strict empirical boundaries of the excavated data as tantalizing but disembodied fragments. In the Last House on the Hill we will use the full power of digital technology to take these fragments - vignettes - of created knowledge and combine and recombine them into an open-ended but always accumulating history of the BACH project, that is itself a (often forgotten) fragment of the larger Çatalhöyük Research project, that leads onto the history of Anatolia.

9. Collaborations and Conclusions

On May 28, 2009, a small development team in Sydney, Australia, the same development team that gave the world Google Maps, announced Google Wave, a singular communication platform that will transform the way humans interact with each other digitally [Ras09]. In the next few years, recombinant history will be a way of life in the cloud, as the lines drawn between e-mail, text messaging, blogging and media management are washed away and we simply communicate with each other, instantly, globally. Waves will thrive on high value information, database narratives, whole cloth event streams than can be dragged and dropped into conversations as easily as photos are now handled in iPhoto. In June, UC Berkeley researchers announced the development of a new storage nanotechnology that could preserve information for billions of years, potentially eliminating the digital preservation crisis we are still facing [Hea09], [Ash08]. In the wake of such innovations, it is high time to be casting away the limitations that are self-imposed by traditional publishing. We really can save and share it all. There will be challenges and setbacks ahead, but we can see no other way than forward.

As we mentioned, the Last House on the Hill project is made possible due to the democratic decision made by the Çatalhöyük Research Project to release all data and media under Creative Commons open licensing, freeing the content to be shared widely with little intellectual property restrictions. The Cotsen Institute of Archaeology has agreed to partner with the LHoTH team to take the deep dive into this new paradigm of holistic digital publishing, through collaborations with the California Digital Library and the UC Berkeley Media Vault Program. We are using this occasion to launch ArchaeoVault.org, a digital preservation and access resource for archaeological projects to be perpetually stored and shared. The LHoTH project URL is www.archaeovault.org/lhoth.

References


[Hod00] Hodder, I.: Towards reflexive method in archaeology: the example at Çatalhöyük by members of the Çatalhöyük teams. McDonald Institute for Archaeological Research, 2000, Cambridge.


