Title
OpenDig: Digital Field Archeology, Curation, Publication, and Dissemination

Journal
Near Eastern Archaeology, 77(3)

ISSN
1094-2076

Authors
Vincent, Matthew L
Kuester, Falko
Levy, Thomas E

Publication Date
2014-09-01

DOI
10.5615/neareastarch.77.3.0204

Peer reviewed

OpenDig:
Digital Field Archaeology, Curation, Publication, and Dissemination

Nikki Oakden and Monique Vincent reviewing field notes using OpenDig on an iPad. Photograph courtesy of the Madaba Plains Project.

Matthew L. Vincent, Falko Kuester, and Thomas E. Levy

The image of the archaeologist is changing; there is little we can do about that. The dusty boots, timeworn trowel, and coffee-stained notebooks may themselves become the images that we see in museums in a not too distant future. One only has to open up any archaeology publication today to find stunning images of drones, tablets, GPSs, laser scanners and a myriad of many other devices that were hardly thought of a decade or two ago (For example, see: Levy 2013; Levy, et al. 2010). The rapid evolution of technology has meant that there are new approaches to field archaeology that can enhance our work, help to preserve the past from our rather destructive discipline, and also help to disseminate our research to the public. Among all these tools, perhaps in the shadow of drones and laser scanners, is the archaeologist's notebook. That notebook, containing the primary observations for excavations is fundamental to understanding work conducted in any trench at any site; and like many of the other facets of archaeology, is also going through a process of change from analog to digital.

In this article, we will look at one such tool, OpenDig, developed in the midst of many of these changes occurring all around us. OpenDig should be understood in the context of the circumstances out of which it grew, the problems which it was created to solve, and as a result can be of use to other archaeologists today. OpenDig should not be seen as just a single tool for excavation, but rather should be understood as a growing framework to record, analyze, publish, and disseminate these data. OpenDig is not the only framework available for archaeological data, nor is it intended to be the framework that all archaeologists should be using. Rather, it is a pragmatic look at its use with a specific project, while still focusing on building in access for unrelated projects that might like to incorporate these data into their own project.

OpenDig: A Brief History in Jordan

Archaeologists at Tall al-ʿUmayri employed digital databases since the beginning of their excavations in 1984 (Brower 1989; fig. 1). At the time, these databases required a specialist to operate and maintain them. Over time, the database was migrated to a Microsoft Access database (Borstad 1999), a much more accessible database that did not require the same skill-level as previous databases. OpenDig grew out of this second stage of the database, meeting new needs that arose over time. When the project originally adopted Microsoft Access for its database, it was the best available option at the time; web-based databases were not yet commonplace, and were difficult to create and maintain. Within the last decade, web-based applications have grown in maturity, sophistication and ease of creation. Therefore, where

Microsoft Access could only run natively on a Windows-based computer, shifting to a web-based application meant that users on any platform were able to access the data. Furthermore, with the Access database, every change to the database meant a physical copy had to be mailed to each researcher, a process that often involved months. A web-based application meant that updates could happen real-time, and researchers had immediate access to any changes in the databases.

Another challenge that ʿUmayri faced was the reality of retirement. For years, co-director Larry Herr had the task of digitizing all the dig notebooks each season. While this task was tedious, it meant there was data-consistency, as Herr knew the excavations intimately, he was able to adapt any language used so as to ensure consistent language throughout the database. While in the field, Denise Herr reviewed the notebooks (fig. 2), ensuring that all the necessary data was recorded before the team left the country. The combined work of both of these people meant that the quality of data collected at ʿUmayri was exceptional. OpenDig addresses these problems by beginning with digital entry from the field. Constraints in the data-entry system ensure data consistency and the system allows for data-validation, making sure that both of the Herrs' tasks are continued through a digital framework.

Inside OpenDig

OpenDig is a framework consisting of three different parts: a mobile application, an in-field lab "server," and a full-fledged web application (fig. 3). The mobile application is one of the principle components, as it handles all the data entry in the field and is one of the most used parts of the framework. Due to limited Internet connectivity, for many projects it is not feasible to maintain an active connection to a server hosted outside of their lab. A local server makes it possible for multiple devices to synchronize data among themselves, while at the same time providing a centralized repository for supervisors and excavators to review daily excavation data. This is a crucial component, providing a way for the extroverts to have access to a central server without necessarily having access to the Internet. Finally, the original, and fundamental, component is the full web-framework application that is permanently hosted on a particular server, allowing researchers to access the data.

The framework has been in active development since 2007, primarily by the author although contributions sponsored by UC San Diego have helped to increase the speed of development. OpenDig today uses an open-source database, CouchDB, as the central component in all of its systems. CouchDB is a document storage system, versus the traditional SQL, or tabular storage that one might see in something as simple as a Microsoft Excel spreadsheet, Microsoft Access database, or a FileMaker database. As a document store, it accurately represents the sheets previously used in the field to record the data, while still making it possible to create virtual tables, such as pottery samples, lithic samples, stratigraphy tables, or any other regular data found across the sheets. Inherent in the database is the ability to synchronize multiple devices, thus ensuring regular backups of the field data on the server. This replication also means there are complete duplicates of the database on each device, adding to the redundancy of the data security.

OpenDig: Web

OpenDig grew as a web application in 2008 (fig. 4), replacing the Microsoft Access database previously used by the researchers at ʿUmayri. As a web application, it created a central place for the researchers to access the data, allowing for corrections to be made in one place, in real-time, without the reliance on a single individual to handle the process. The major flaw was the need to continue to digitize the sheets recorded in the field, adding a significant burden on the various field supervisors at the end of the season. It was around this time that developers began actively developing for smart devices (phones and tablets) at which point it was obvious: we needed to go digital from the field, thus eliminating the need to transcribe documents.

OpenDig: Mobile

Field recording requires ruggedized equipment, particularly when working in harsh conditions such as the deserts of Jordan. In the past, computers in the field were either too expensive or too impractical to be of any value. Any digitization had to take

Figure 1. James Brower entering data in the first ʿUmayri database, 1984. Courtesy of the Madaba Plains Project.

place in the labs where a clean lab with computers, prints, scanners, and other equipment could be safely setup. However, the smart phone revolution, and perhaps more importantly, the technology to produce ruggedized cases for those devices, means that power mobile computers can be brought out to the field for relatively low costs. These devices are capable of storing many gigabytes of data, as well allowing the user to take photographs from the device, localize using a GPS, and take advantage of the powerful processing power now found on these devices today (fig. 5). Assuming the presence of a data connection, a user can use their mobile device as a terminal from which they can produce three-dimensional models of excavation units or connect directly to remote data sources. For OpenDig, these devices are central to the data-cycle, as the data is born digitally on these devices.

Figure 2. Data validation before OpenDig: the weekly notebook review. Photograph courtesy Madaba Plains Project.

OpenDig: Lab

We live in a digital age, that is to say, most of the time. Often times in archaeology, we are working in underdeveloped, remote areas where high-speed, readily available Internet is simply not yet there. While we often think about our data ecosystems with data connectivity in mind, that cannot be assumed in many parts of the world. For this reason, OpenDig uses a temporary server while in the field. This handles the synchronization of data across devices, as well as providing a central place to review and edit the data while in the field (fig. 6). The nature of the database used to run OpenDig allows for the direct embedding of the web application into the database itself, making it extremely easy to install in the field as any computer can act as a server. This also adds redundancy, as multiple servers can be setup and synchronized thus creating secure copies of the data. In the cases where a data connection is present, it is possible to periodically send the data back to the principle server, thus ensuring an off-site backup at the same time.

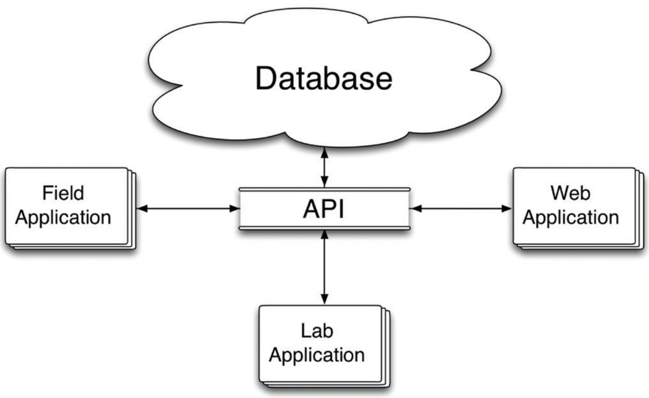

Figure 3. The OpenDig framework. Courtesy of M. Vincent.

Curation, Publication, and Dissemination

As a framework, OpenDig provides a platform for the curation, publication, and dissemination of archaeological data. By treating these data as digital from the beginning, it is possible to significantly increase the speed of data processing. For example, as data are recorded in the field they are automatically synchronized once the team returns to the lab. Within minutes of everyone synchronizing their devices, supervisors can review their day's excavation without the need to retrieve notebooks from various individuals. Likewise, they have access to all the previous excavation records in the palm of their hands, without the need to carry photocopies of the notebooks from previous seasons. With the database in place, researchers are then able to quickly produce final reports by having access to all past and present data in one single place. Furthermore, cross-site studies are easily carried out, as researchers are able to query the database for specific interests, returning results from excavation data from all over the site.

The natural next step in a web framework such as OpenDig is to use the same platform for publishing these data. One of the major challenges here is the longevity of data, as can be seen by the multitude of projects that have disappeared after the funding dries up. The only solution is to tap into permanent infrastructures, ideally libraries that have a strong focus on digital curation. One such example is the RCI collaborative project as part of the California Digital Libraries initiative at the University of California, San Diego (http://rci.ucsd.edu). As a pilot project, the UC San Diego Levantine and Cyber-Archaeology Laboratory are using the data from the excavations at Khirbat en-Nahas in Jordan's copper ore-rich Faynan district that will be converted into a read-only, long term storage and accessible through the library interface itself. Due to the ongoing development of such infrastructures, OpenDig does not currently support a way to export data for such projects. However, assuming the adoption of digital formats for long term storage, OpenDig is ready to adopt any data format the international community sees as necessary for the permanent storage and publication of archaeological data.

Figure 4. OpenDig, web-based archaeological database. Courtesy of M. Vincent.

OpenDig has adopted an open-access policy, but for the software as well as the data. As the name implies, access should be open, allowing other researchers to tap into the database and incorporate the primary data into their own research. Some of the challenges associated with this will be discussed below; what should be highlighted here is that access to these data should be both human and machine-readable. As OpenDig is a series of JSON (JavaScript Object Notation) documents, they are inherently machine-readable. The interfaces built on top of the database, both the web and lab applications for instance, are readers that categorize and manage the data in order to make it human readable. The data can be accessed through APIs (Application Programming Interfaces) by machines, allowing other researchers to create their own applications taking advantage of the OpenDig back end, or to simply write queries that pull sections of data from OpenDig to incorporate into their own analyses.

OpenDig: Challenges

The academic system in place today encourages the delay of primary data publication by rewarding analysis and penalizing the researcher if someone else should publish an analysis of their primary data before they do. For doctoral students, publishing primary data could often be damaging to their dissertations, even endangering their candidacy altogether. If an open-access system of data is to be adopted, the publication of primary data needs to be rewarded and embraced at the same level as the analyses of these data. With such a strong emphasis on the analysis, we are often left with only the analysis, and select tables of primary data. Rewarding the publication of primary data needs to be discussed by the scientific community, enabling an open exchange that can only benefit the community as a whole.

Looking Forward

OpenDig is not the end-all solution for archaeological data. There are many platforms today that allow researchers to collect, analyze, and publish data from the field (Gidding et al. 2013). Researchers should adopt the platforms that best fit their projects, although it is imperative that we strive for open data access, allowing researchers to access and incorporate data from other projects, thus creating holistic studies that go beyond the subjective interpretations often found in the analysis published in archaeological journals (see Kansa 2007; Schloen 2001). Open access to primary data allows researchers to draw their own conclusions, conclusions that can only be enhanced by incorporating primary research data from other areas.

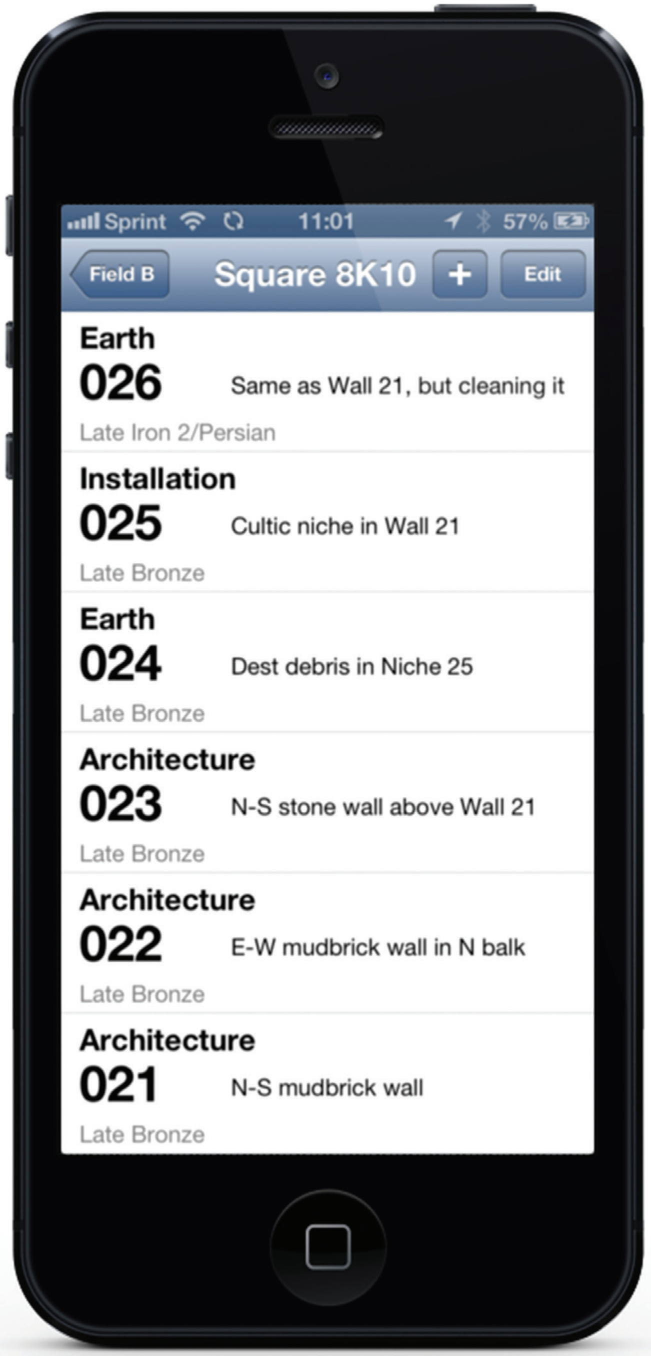

Figure 5. OpenDig, mobile data-entry. Courtesy of M. Vincent.

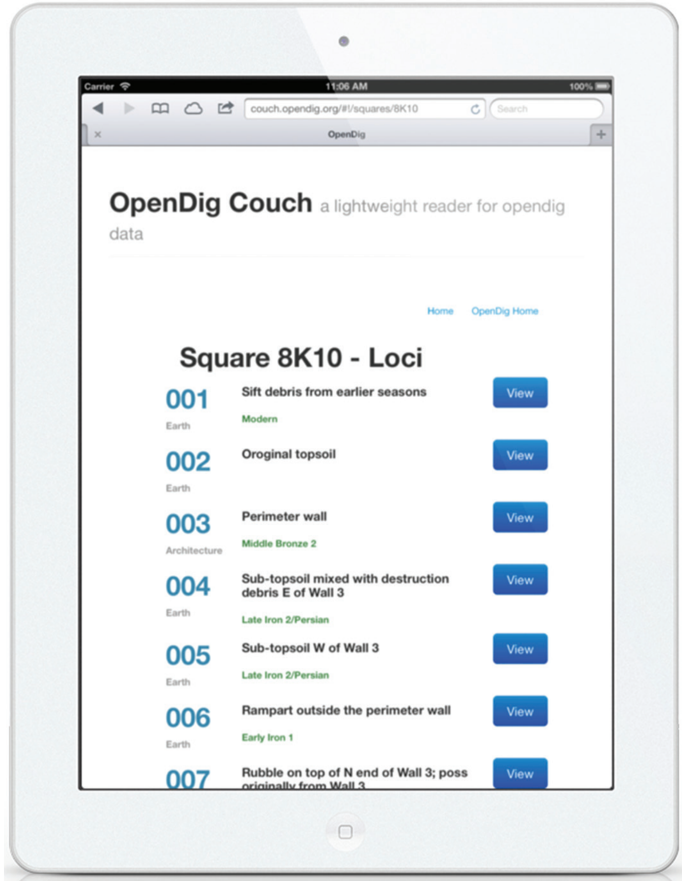

Figure 6. OpenDig, portable server for managing data in the field. Courtesy of M. Vincent.

Acknowledgments

The research leading to these results is partly funded by the EU Community's FP7 PEOPLE under the ITN-DCH Project (Grant Agreement 608013) and partly by the National Science Foundation, under IGERT Award #DGE-0966375, "Training, Research and Education in Engineering for Cultural Heritage Diagnostics" awarded to FK (PI) and TEL (Co-PI). Additional support was provided by the Qualcomm Institute at UC San Diego, the Friends of CISA3 and the World Cultural Heritage Society. Opinions, findings, and conclusions from this study are those of the authors and do not necessarily reflect the opinions of the research sponsors.

References

Borstad, K. A. 1999. *Data Entry and Information Retrieval System Manual: The Madaba Plains Project*. Berrien Springs, MI: Institute of Archaeology, Andrews University.

Brower, J. K. 1989. Archaeological Excavation Data Management System. Pp. 387–401 in *Madaba Plains Project: The 1984 Season at Tell el-Umeiri and Vicinity and Subsequent Studies* Vo. 1, eds. L.T. Geraty, L. Herr, O. LaBianca, and R. Younker. Berrien Springs, MI: Andrews University Press.

Gidding, A. et al. 2013. ArchaeoSTOR: A Data Curation System for Research on the Archaeological Frontier. *Future Generation Computer Systems* 29: 2117–27.

Kansa, E. C. 2007. Publishing primary data on the World Wide Web: Opencontext.org and an Open Future for the Past. *Technical Briefs in Historical Archaeology* 2: 1-11.

Levy, T. E. 2013. Cyber-Archaeology and World Cultural Heritage: Insights from the Holy Land. *Bulletin of the American Academy of Arts & Sciences* LXVI: 26–33.

Levy, T. E. et al. 2010. On-Site Digital Archaeology 3.0 and Cyber-Archaeology: Into the Future of the Past – New Developments, Delivery and the Creation of a Data Avalanche. Pp. 135–53 in *Introduction to Cyber-Archaeology*, ed. M. Forte. Oxford: Archaeopress.

Schloen, J. D. 2001. Archaeological Data Models and Web Publication Using XML. *Computers and the Humanities* 35(2): 123–52.