

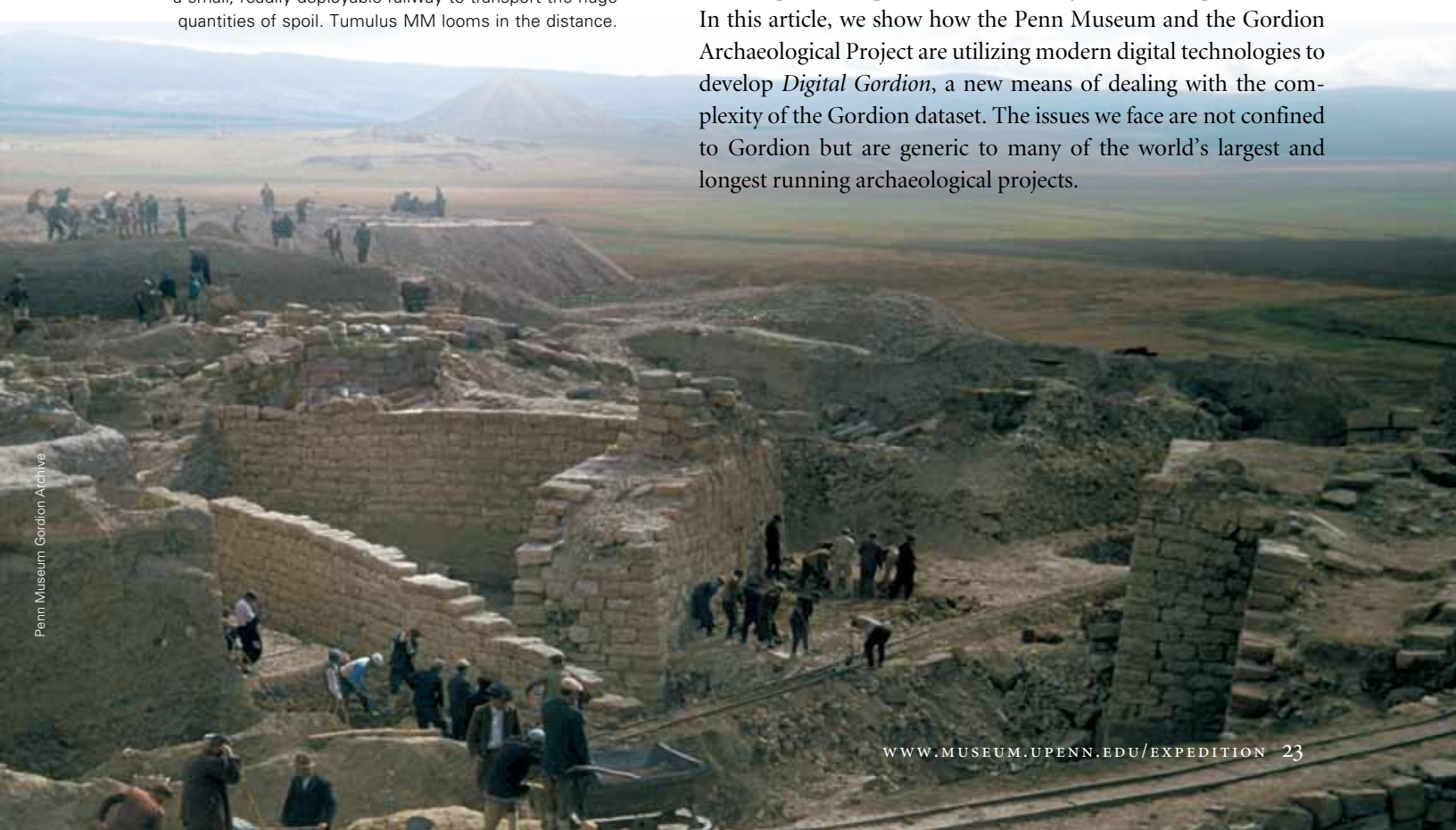
BUILDING DIGITAL GORDION

Coping with the Past in the 21st Century

BY GARETH DARBYSHIRE
AND GABRIEL H. PIZZORNO

Below, excavations in 1955 of the superbly preserved and completely buried Early Phrygian Iron Age citadel gate dated to the later 9th century BCE. Note the large size of the team—local Turkish men, who for decades have done most of the actual excavating at Gordion—and the use of a small, readily deployable railway to transport the huge quantities of spoil. Tumulus MM looms in the distance.

Ever since the Penn Museum began excavations there in 1950, Gordion has remained a key site for the archaeology of 1st millennium BCE Anatolia. The significance of the site derives from the intrinsic historical importance of the place—a longstanding center of power and population—and from the long-running excavations that have revealed the physical dimensions of that history in remarkable detail (see “Gordion in History,” this issue). The resultant archaeological dataset is correspondingly of great size and complexity, and therein lies its value. These same attributes, however, when combined with the dataset’s relative inaccessibility to researchers, have proved to be a serious obstacle to the completion of post-excavation analyses and final publications. In this article, we show how the Penn Museum and the Gordion Archaeological Project are utilizing modern digital technologies to develop *Digital Gordion*, a new means of dealing with the complexity of the Gordion dataset. The issues we face are not confined to Gordion but are generic to many of the world’s largest and longest running archaeological projects.





Satellite view of the Gordion Citadel Mound. The huge scale of the archaeological excavations is apparent, in particular the Main Excavation Area, above center, where the remains of the Early Phrygian Destruction Level are still visible. Other sizable excavation trenches, as well as the great tips of excavated spoil, can also be seen at the edges of the mound. The excavations at the bottom right revealed the extensive Lower Town fortifications.

A PROBLEM OF DATA

Despite a substantial record of achievement that spans six decades of work and includes many publications, much remains to be done with the Gordion excavations dataset. Key artifact assemblages still require analysis and publication, and the complex details of the site's stratigraphic sequence, and its many architectural and depositional contexts, have yet to be fully defined, synthesized, and reported. Without such a comprehensive contextual framework readily accessible to all researchers, none of the Gordion analyses are properly geared to the others, and all are proving more difficult and time consuming to complete than they might otherwise be.

The problem is that all of these analytical projects constitute major endeavors that inevitably take many years to finish, usually much longer than was originally conceived. In addition, the size of the research team and available resources have not matched the magnitude of the task. Indeed, the Gordion project has already outlived several of its members, including the first two directors, Rodney Young and Keith DeVries. As these

people have passed away with some of their key work unfinished, extremely valuable and hard-won matrices of knowledge have been lost. To put the situation in perspective, and with the benefit of hindsight, we suggest that one season of digging will generally require a minimum of three years for post-excavation analysis and publication. While this equation would vary depending on factors such as the scale of individual excavations, the size and input of the post-excavation research team, and the project's priorities, a 1:3-years ratio seems a reasonable guideline. By this yardstick, Gordion, with over 30 seasons of digging, would need at least 90 years to prepare the significant excavated data for full publication. Such uncomfortably realistic time-scheduling seems to be rarely articulated and would probably be very unwelcome news to potential funders. Nowadays, excavations tend not to be as extensive or as long-running as those in the past, a sensible attempt to keep analysis and publication more under control (but one that may offer far less expansive vistas). With hindsight, some might now question whether the Gordion excavations should have run for as long as they did. But given that they have, and

much of great value has resulted, what can be done to improve the situation? The first step is to identify the root causes of the problem. There are three of these.

The first is the sheer size and complexity of the Gordion dataset. The original excavation records alone comprise hundreds of thousands of pieces of information: photographic negatives, slides, prints, maps, plans, catalogue cards, lists, documentary reports, and artifact drawings. Furthermore, ongoing post-excavation analyses have naturally produced much new data, and these large-scale additions to the dataset will continue for years to come. As the product of 60 years of development in archaeological method and theory, the data also vary considerably in character. Quantifying and qualifying the various kinds of data, and integrating and organizing this information in such a way that projects can be accurately and rapidly defined and prioritized, has always been problematic. And once a project has been defined, a considerable amount of time and other resources is required for interpretive study of the material. Completing the stratigraphic and contextual work would certainly demand long-term and regular access to the records, a sizable team of people, extensive collaboration between the researchers, and an effective means of continually reorganizing data.

This leads us to the second difficulty, one that compounds the first: access to the data. The original excavation records can only be studied in their entirety in Philadelphia, where they are housed in the Gordion Archive, a well-organized paper repository at the Penn Museum. However, the active Gordion research team is scattered across four continents. As a consequence, only a selection of the original excavation records has been fully studied, through rare, costly, and usually inadequately short research visits to Philadelphia. In particular, the enormous photographic record has never been easy to access. Due to high costs, many of the black and white negatives were never developed into prints. And the color slides could only be viewed at the Archive, on a light table or using a projector.

The third aspect of the problem is that effective collaboration between team members has been difficult to



realize given their widely scattered distribution, infrequent opportunities for convening, the limitations of traditional communications systems, and in some cases the inadmissible but inevitable specter of

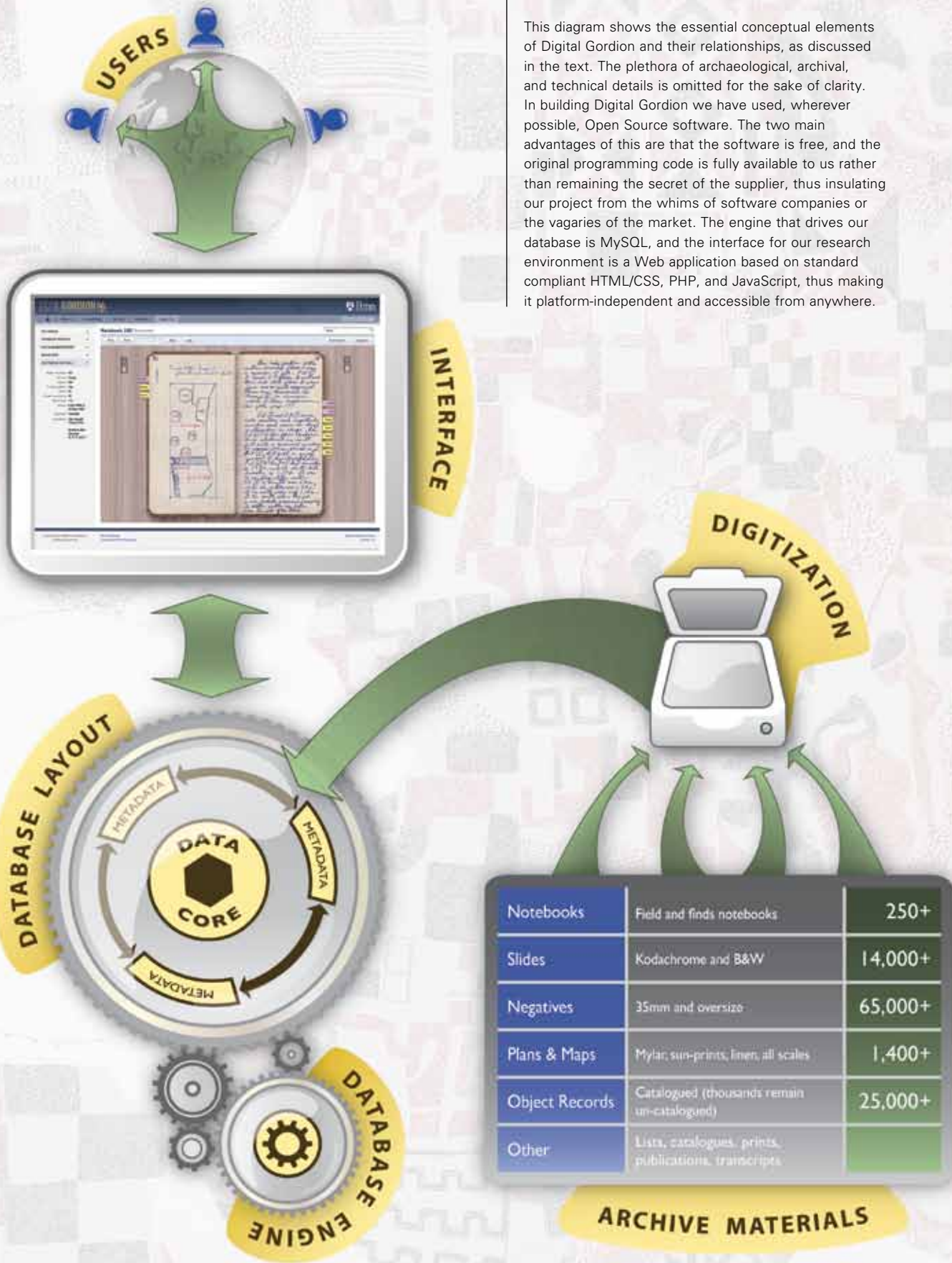
personal rivalry. This has inhibited timely pooling of knowledge and joint approaches to tackling problems.

A DIGITAL SOLUTION

It seems clear that a new approach to managing the Gordion dataset is needed to do full justice to the importance and complexity of Gordion's history. This new approach must capitalize on the work done by the project, and help it more rapidly repay the long-term debt of reportage owed to the world. Furthermore, it must effectively deliver the huge quantity of data to the researchers, improve their analytical toolkit, and facilitate better collaboration among team members. The means for this new approach have only begun to materialize in recent years with the development of readily available digital technologies. These technologies offer a clear way to overcome obstacles that were practically insurmountable before. Digitized data can be made readily accessible throughout the world via the Internet. Information can be organized and processed in many powerful ways, taking the burden of data processing from the researcher and transferring it to the computer. A digital solution offers an effective format for handling masses of photographic data and facilitates collaboration between people working on the same data from different locations, while at the same time providing a backup of the paper and photographic material. Digitized data offers the potential to shave years off the post-excavation program and liberate the research team from many logistical burdens. Researchers will thus be able to concentrate on the interpretation and presentation of the material.

Our solution to the Gordion dilemma, then, is to develop an online research environment, accessible from any location where the Internet is available, powered by a sophisticated database that stores and correlates digital versions of the Archive's contents, and offers the tools to manage and process those data. We call this environment *Digital Gordion*.

Many conceptual and practical challenges face the creation of Digital Gordion. The range of issues involved comprises an intimate combination of archaeological, archival, and computer-technical matters (rather than the latter alone, as is often assumed). To effectively tailor the database and the digitization program for producing rapid results, careful evaluation must be made of the archaeology of the site and the specific character of the archival data, as well as the knowledge, abilities, and research aims of the users. Technical considerations



This diagram shows the essential conceptual elements of Digital Gordian and their relationships, as discussed in the text. The plethora of archaeological, archival, and technical details is omitted for the sake of clarity. In building Digital Gordian we have used, wherever possible, Open Source software. The two main advantages of this are that the software is free, and the original programming code is fully available to us rather than remaining the secret of the supplier, thus insulating our project from the whims of software companies or the vagaries of the market. The engine that drives our database is MySQL, and the interface for our research environment is a Web application based on standard compliant HTML/CSS, PHP, and JavaScript, thus making it platform-independent and accessible from anywhere.

include selecting the right technology and hardware at a reasonable cost, creating appropriate programs for building the database and interface when these are not readily available off-the-shelf, and devising effective workflows for the digitization process.

Having decided to take on the challenge, our small team at the Penn Museum, under Deputy Director Brian Rose, has spent the past year taking the first steps towards building Digital Gordion. Our strategy had three main components. First, we needed to conduct a thorough evaluation of the Archive's resources and use patterns. Then, armed with that new understanding, we had to design and begin development of the online research environment while at the same time putting in place a program to digitize the archival data.

THE GORDION ARCHIVE

The essential components of the Gordion Archive dataset are field notebooks (the daily written record of each excavation in progress, including sketch plans, sections, and other drawings), photographs (negatives, prints, and slides), large format plans and maps drawn to scale, drawings of artifacts, many lists (of photographs, artifacts, and so forth, with details of their provenance, etc.), and post-excavation written reports. The archived material often constitutes the only extant evidence of what was found by the excavations, since, apart from the artifacts, most of the archaeological material was destroyed by the Gordion project as part of the normal process of digging down to the next stratigraphic level.

All the data interrelate and overlap to varying degrees, according to such interconnections as excavation season, general location on the site, specific excavation trench, specific structural or depositional context, artifact category or type, notebook reference, historical period, and so on. We distinguished the various components of the dataset and identified their generic similarities and differences, and the nature of their interconnections, from both an archaeological and an archival perspective. We evaluated how these data are used in research, and we assigned them a priority value based on the current and impending needs of the project. We then determined how the different types of data could best be digitized. Our task was made much easier than it might have been because of the superb organization of the archive, established by the late Ellen Kohler.

Our thorough evaluation of the archive gave us the necessary information to design a research environment



The Gordion Archive and its creator, Ellen Kohler, in 1989.

that matched the archaeological and archival requirements of the Gordion researchers. This understanding of the Gordion materials and research workflow enabled us to create a digitization program that will be economical and easy to manage. Priorities were established to effectively meet the most urgent data needs of the users, so that the advantages of digitization could be realized as soon as possible.

DESIGNING THE ONLINE RESEARCH ENVIRONMENT

The aims and principles of Digital Gordion are largely determined by an evaluation of the targeted users—the Gordion research team. Understanding their requirements is a crucial matter for it is they who are doing the research and they who primarily need the data. While it helps that we ourselves are also users, the opinions of others are crucial to fine-tuning our approach.

What is required is a powerful research tool that can incorporate all types of Gordion data and interconnect them in the same complex web of relationships that already exist in the mind of the archaeologist, yet is easy—indeed pleasing—to use by someone who is not a computer specialist. One of our primary goals is to replicate, as far as possible, the familiar ways that Gordion researchers use to work on their material. The key concern in designing our online research environment is to give users the ability to do everything they already do with the paper data, but much more quickly and thoroughly, while also giving them the tools to do things they could never feasibly do before. Thus they should be able to search for and sift through all available data (including pertinent data which they may have been unaware existed), pull up any required text and images, and



A screenshot of the Digital Gordion research environment as a user browses and annotates a field notebook. The *Inspector* tool shows all materials connected to that specific page. Various data pertinent to an ivory comb have been called up, including its catalogue card, a drawing and a color slide of the object, as well as a plan which details the depositional context of the artifact. Personal bookmarks have also been added to the notebook.

annotate material, thereby transforming themselves into a virtual Gordion archivist as well as an online researcher. They should now be able to carry out tasks that would formerly have placed multiple demands on the Gordion Archive and which would have taken days, weeks, or months to perform. All of this should be achieved at each researcher's convenience and location, and with great rapidity and confidence.

The first step was the design of the database that would drive our research environment. The process of constructing the data model for this database is the most research-intensive aspect of Digital Gordion, since it requires a deep understanding of the complex theoretical principles that guide both database design and archaeology. Our aim was to create a model that can accommodate the varied data from all the different projects associated with Gordion, past and future. In particular, the datasets from the Young and Voigt excavations differ considerably in their format and details. In developing our data model, we thus had to walk a fine line, not forcing the archaeological data into preconceived frameworks, but at the same time attempting to keep the model itself as close as possible to an ideal, generic archaeological workflow. By mapping the different data in the Gordion Archive to their corresponding entities in the database, each with their own attributes and particular relationships, we defined the nature of the archival data in a way that the computer can understand and process. This is essential since we aim to do more than just store the data. This capacity of the computer to understand and process our data is what makes our research environment a smart tool.

The design of the database only gets us halfway to the accomplishment of our goal. The second half, just as important, is the interface. This is the *face* of the database as it appears on the user's screen. It is the only medium through which the user and the database interact with each other. For most people, the interface *is* the research environment. Because of this, the design of the interface plays a central role in our development process, as Digital Gordion must be user-friendly. We did not want an interface that appeared daunting, intimidating, or confusing to the user, or that suddenly rendered the familiar Gordion data alien in appearance. Rather, we strived to create an experience that is simple to learn and use, and that matches as closely as possible the normal research workflows, while at the same time augmenting these by integrating new capabilities offered by digital technologies in order to create a sophisticated tool. This confers a major practical and psychological advantage that is often lost when trying to utilize commercially available databases and interfaces that have been developed for purposes other than archaeology, or archaeological databases developed *ad hoc* for specific projects without a central vision to guide their design.

Digital Gordion is designed to enable individual researchers to rapidly assemble and effectively process all the data required for their own research. A key aspect is the embedding of a powerful search capability at every level of the interface allowing the user to refine their dataset on-the-fly, with the results being updated live as they type. This search capability is also built into every aspect

A screenshot of the Field Notebooks module of Digital Gordion, showing the results of a search for “bone” and the user’s personal bookmarks.



of the database itself, so that the application searches not only keywords but also every word in the database. That means that a simple search for a word will flag every item where the word is mentioned including everything from publications and notebooks to image captions and mere scraps of paper. A translation matrix ensures that the database also understands different designations for the same object. For example, a search for an Iron Age building such as “Megaron 2” will return also every mention of “Meg. 2” or “M2.” Furthermore, the search algorithm will attempt to understand context in order to return coherent datasets as results, rather than simply a list of words.

Besides allowing for the quick compilation and refinement of datasets, the universal search is also a powerful tool for rapidly and accurately evaluating the scale and character of any Gordion research theme. By immediately indicating which data are pertinent, it allows a more realistic and immediate quantification and prioritization of projects than was ever possible before.

Users can assemble and browse through the core data just as they would with the paper originals. But they can also process and make changes to the data by adding their own interpretations, notes, and keywords, and making corrections, all of which are stored as metadata for each individual user. The original information is never altered, ensuring that one can always backtrack to the original state of the data as it was collected. New information created by users during the process of analysis, or any other transformations or alterations to the original information, is stored in layers of metadata. It is these

metadata that arm us with the “research muscle” for Digital Gordion, allowing researchers to transform the original data into new analytical and interpretative configurations.

Users could, for example, pull up a field notebook, flip the pages, add bookmarks, and at the same time rapidly make a search of the entire text for any word. Images could be sorted according to the researcher’s own labeling scheme, annotated with words and symbols, and viewed in detail with a virtual loupe (the digital equivalent of a magnifying glass). Users can save their growing body of personal work to a central repository where it is safely stored and ready for them to pick up again in their next online session. At the same time, the original dataset of core data is always available. Users can copy and download data as required such as maps for a presentation or publication-quality images for printing.

While the data generated by an individual user is in principle private, users can choose to share their work with the rest of the Gordion team. Digital Gordion is designed to be a collaboration tool to facilitate and encourage cooperation between researchers. Users have great flexibility in deciding what data to share with others. Researchers can also work on the same material at the same time and thus pool their interpretive efforts. By collaborating in Digital Gordion, individual efforts will be more effectively welded together, so as to benefit the collective whole. Cooperation is essential to completing the analysis of Gordion’s complex stratigraphy and archaeological contexts in a timely manner.



A screenshot of the Images module of Digital Gordion, showing a user employing the virtual loupe tool to zoom in on a detail of the selected image

THE DIGITIZATION PROGRAM

Before any of the features of Digital Gordion can be realized, the contents of the Gordion Archive must be converted from paper and film to digital formats through a digitization process that includes the scanning and formatting of the data. Digitization is a time consuming and complex procedure, filled with unexpected challenges and pitfalls. It is a major operation that has required the development of a new organization and innovative ways of working with the data. Our approach has been shaped by our personal experience as archaeologists working with archival material combined with a healthy dose of learning from mistakes. The methodologies we have developed have proven effective in allowing us to efficiently digitize the material.

Given the limited resources currently available, selection of data to be imported into the database is a major strategic issue. Following our evaluation of the Gordion Archive and the needs of the researchers, our basic doctrine has been to focus on those data that will be of broadest and most immediate use to the most people, and which can be digitized quickly at relatively low cost. Thus we have initially concentrated on the excavation notebooks, which form the backbone of any research at Gordion; lists of artifacts and photographs, which are key for elucidating the many and complex relationships within the dataset; and the key photographs detailing the excavations and stratigraphy. In addition, we are responding to various urgent requests for other specific data from Gordion researchers.

In terms of labor, we are currently largely reliant on the contributions of volunteer and work-study person-

nel. These include local high school students, college students, and retired volunteers. Members of the Gordion research team are also contributing, whenever it is possible for them to come to the Museum, by digitizing data pertaining to their own research interests using our scanning facilities.

Regarding the actual process of digitization, we have paid particular attention to identifying the best operational procedures, or workflows, for transforming the paper and photographic data into their digital counterparts. We have broken down even the most complicated digitization tasks into sets of easily learned, manageable, and time-efficient procedures that limit the amount of interpretive thinking required on the part of the worker, thus reducing the chances of

erroneous or inconsistent data entry. This is essential because we rely on a team with varying levels of experience and the workflows allow us to quickly incorporate new people into the project while guaranteeing consistent quality of results.



This is an exciting time to be involved with Gordion, with a great deal of analytical work variously nearing completion and just beginning. Digital Gordion is a collaborative archaeological research project in its own right and, as it develops, it will greatly help the Gordion team to fulfill the mission begun in earnest in 1950.

The problems that we have outlined are not unique to Gordion. Our hope is that the methodologies we develop for Digital Gordion will be applicable to many other archaeological projects. 🏠

GARETH DARBYSHIRE *is the Gordion Archivist at the Penn Museum.*

GABRIEL H. PIZZORNO *is a Ph.D. candidate in Art and Archaeology of the Mediterranean World at the University of Pennsylvania and Research Assistant to the Williams Director at the Penn Museum*

For Further Reading

To learn more about Gordion, visit the official project website at <http://sites.museum.upenn.edu/gordion>