The Karanis Housing Project: Integrating archival archaeological material and a contemporary, user-centered geospatial database

1. Introduction

The discipline of archaeology possesses an enormous wealth of unpublished or under-published data. These datasets derive from excavations over the past two centuries at sites throughout the world. Much of this material is legacy data, a term that refers to any records preserved within an obsolete system. Largely moribund in analog format, hand-drawn maps, lists of finds, printed photographs or negatives, and handwritten or typed excavation diaries, are not easily accessible to scholars and the public. (Allison 2008) These data sets, however, provide critical contexts for the artifacts housed in museums and storerooms in the United States and abroad, and have the potential to open vast reserves of untapped material for contemporary investigations of the past in multiple humanistic disciplines. (Kintigh 2014, Bevan 2015) Coupled with contemporary analyses of the legacy material or new fieldwork, these data can permit scholars to reimagine sites excavated in the last two centuries. One such site is Karanis, modern Kom Aushim, located in the located in the Fayum region of Egypt.

The village of Karanis is one of the best-preserved sites from Graeco-Roman antiquity. The University of Michigan excavated Karanis between 1924 and 1935; the finds, including thousands of artifacts and papyri, were divided between the University of Michigan, where they are now housed in the Kelsey Museum of Archaeology, and the Egyptian Museum in Cairo. The excavations, like other field projects from the early- and mid-twentieth century, employed hundreds of workmen who cleared the site on a scale unimaginable by today's standards. Hundreds of buildings were excavated and recorded, providing an unparalleled record of life in antiquity and the workings of a colonial settlement. (Bagnall and Rathbone 2004). These data provide a rich source of evidence that can be used to explore critical questions about social life in a village during the Roman Empire, if only they were in an accessible format. (Wylie 2016)

The excavation records, which include detailed field journals, excavation diaries, maps, plans, photographs and even a moving picture, are currently kept at the Kelsey Museum at the University of Michigan. The Karanis Housing Project aims to make much of the legacy data available to a broad audience: scholars wishing to work on artifacts and papyri that were recovered from the site as well as lay public interested in archaeology. End-users of the Karanis Geographic Information System will to be able to both search for the contents of specific buildings and also identify all instances of particular classes of objects or specific finds, spatially locating them in the digital map of the site. This will permit high-level analyses of the data, including neighborhood and network studies, revision of the stratigraphic record of the site, and assessment of archaeological distribution patterns and depositional sequences. (Wesson and Cottier 2014; Cooper and Green 2015). The formal goals of the project are:

- To develop a digital map of the site that translates the analog, hand drawn maps into a GIS format
- To generate a relational database with clear descriptions of the artifacts and papyri recovered from the site

- To populate the digital map with lists of finds, each associated with its respective archaeological findspot
- To link the finds from the relational database with external data sources, such as the Kelsey Museum artifacts database, Trismegistos.org, and Pleiades
- To make the GIS and artifact data available through an online platform and associated webpage

The Karanis Housing Project is based at Oberlin College and has been supported by multiple grants, including the Oberlin-Kalamazoo-University of Michigan Faculty Collaborative Grant, 2012-2016; Oberlin College Research Assistantship, 2012-2018; and an Ohio-5 Digital Humanities Grant in 2015. Progress on the project is ongoing, and we have achieved a number of benchmarks that demonstrate continued success.

2. The Analog Records of the Karanis Excavations

The methodology used by the excavators was revolutionary for the time, as the excavators carefully record the physical location of finds, using architectural structures as a delimiting factor in cataloging the artifacts as they were recovered in the field. (Gazda and Wilfong 2004; Wilfong 2014) Most earlier excavations had paid less attention to precisely locating their discoveries. In contrast, the Karanis excavations employed a variety of interrelated record-keeping methods to keep track of the excavations and finds.

2.a. The scope of the excavation

When excavations began in 1924, Daira Agnelli, a mining company, had been given a permit to extract the nutrient-rich fertilizer, or sebbakh, that was found at the site. (Figure 1) The first year of excavation was dictated by the mining company, as the excavators essentially followed the sebbakh diggers. Subsequent seasons reversed these roles, as the Michigan team dictated where excavation would occur, and the archaeologists were therefore able to approach the site more scientifically. Even with the loss of the center of the town, the resulting records are astounding.

In the first season, along the eastern face of the central pit created by the mining company, three strata of houses were visible, each separated by a layer of sand. The excavators believed that these strata corresponded to periods of occupation; this visible stratigraphy laid the foundation for their interpretation of the site. From the 1926-1927 season under Enoch Peterson onwards, buildings were identified according to one of five different occupation levels, A-E, with A representing the layer below the surface; E, at the bottom, was believed to correspond to the second and first centuries BCE. (Husselman 1979) Collating evidence from across the mound, the Michigan excavators eventually proposed a sequence of site occupation based on architectural change and variations in elevation; each phase was assigned a rough date, and excavation loci that lacked securely datable materials (i.e. coins or papyri) were assigned a date according to their place within the chronological schema. Early on, the Michigan excavators assumed that the site was abandoned around the fourth century, as there is little evidence for Christianity. Recent work, however, has shown that the site continued into the sixth or even seventh century. (Pollard 1998)

The flaws in the Michigan methodology are readily apparent. Often, changes in construction may have been evident in the architecture, but these architectural changes were not reflected in the stratigraphy. In other words, construction changes, such as blocked walls, were not indicative of different depositional periods. There are numerous instances where a structure was dated to both the C and B levels, but finds are assigned to only one level, either B or C, and no finds were listed for the other level. Structures that were built side by side, in the same insula block, with shared walls, were sometimes designated as belonging to different levels, even though they were likely constructed at the same time. (Stephan and Verhoogt 2005) The excavators also assumed that similar patterns of construction and occupation were at work over the entirety of the village, although there was little evidence to support this. Different parts of the site may have been occupied at different times, and some houses may have lacked one or more of the phases. Despite these problems, the Karanis excavations offer a wealth of useful information related to life in a Graeco-Roman village in Egypt, and the quantity of excavated artifacts and papyri from the site remains unparalleled.

2.b. Types of Records

2.b.1 Maps

The excavation team created precise maps of most of the excavated areas. A map of the site was tied to a number of datum points. This allowed the excavation team to divide the site into grids; work proceeded within different regions of the site. (Figure 2) Key Plans, one for each of the Levels A – E, were produced for the site as a whole; these marked major structures, and provided a general sense of the site. (Figure 3) Within each of the grid squares for each level, the site architect drew detailed plans of each preserved building differentiating areas that were reused from earlier phases, or that the excavators believed had once existed. (Figure 4) A much smaller number of streets and houses are documented only through sketch plans, indicating the shape of the buildings, but little else.

2.b.2 Plans of Houses and Buildings

Detailed top-plans (architectural drawings viewed from above, Figure 5) and elevations (images of preserved remains viewed from one of the cardinal directions, Figure 6) were made for certain buildings and houses.

2.b.2.i Elevation Records

In the course of the excavation, the architect made sketch plans of each of the excavated structures, marked with elevations taken in the field (Figure 7). More detailed drawings were also created, annotated with numbers that refer to elevation records kept in separate journals, keyed to the elevation measurements taken each day from one of the datum points using a theodolite. In the elevation log books, the measurements have been resolved to record meters above sea level. (Figures 8 and 9)

2.b.3 Photographic Archive

The Kelsey Museum retains a detailed photographic archive from the excavations. Photographs were taken of many of the structures during excavation, and the field team also took images of interesting or unusual finds as they were unearthed. (Figure 10) A separate record exists of nearly all of the artifacts that were retained. The Division Albums contain photographs that lay out the artifacts by type; these images were taken before the objects were divided between the University of Michigan, the Cairo Museum, and other stakeholders, such as the National Museum of Athens. (Figure 11)

2.b.4 Lists of Finds

Record Keeping was organized around identifiable buildings; when no walls were readily apparent, the space was designated an "Area." As finds were uncovered in each building or area, they were recorded in the field, sometimes with a notation that indicated the approximate location within the structure or its fill. These records were then transferred to the "Records of Objects," one massive volume for each year of the excavation. (Figure 12)

In the "Records of Objects," each room of each structure or area is recorded on a separate page or on multiple successive pages. Artifacts are designated by upper case letters, proceeding sequentially from A – Z, and then, most often, AA, AB, AC, and so on. In some buildings, there is variation in the numbering practices after Z. Finds of ceramic vessels are numbered sequentially using lower case letters, a – z. Only some of the finds were retained, and many are marked with the abbreviation N.T.H., or "Not Taken Home." There were reasons for this choice. The Karanis excavation was designed to create a record of daily life in the village, and the excavators did not think it was necessary to retain every instance of each kind of artifact.[1]

Each object, as it was recovered from the soil, can be identified by an established code that contains the following information: year of excavation, level, building, room, and find number. For example, a small mud dish discovered in 1929 was coded as 29-B224B*-T.

29	= 1929, Year of excavation
B	B-Level
224	Building number
B	Room
T	Sequential letter given to each artifact

In theory, this system should work well, providing a unique identifier for each artifact. However, a more complex example can demonstrate some the interpretive difficulties in using the system. A papyrus discovered in 1929 was assigned field number 29-C123CQ^4-B. CQ^4 is one of the bins along Passage CP (Figure 5). Because of the large numbers of rooms and spaces in the granary, the excavators extended the room numbering beyond Z, following the sequence AA-AZ, BA-BZ, and CA-CZ, and then appending a superscripted number to indicate successive features, such as bins, within a given room. Other units include a "*" after the room number, often used to designate material found beneath the floor, or in a special space within the room, as is the case with B224 above. Thus the artifact designations provide another layer of complexity.

[1] Indeed, this was following contemporary practice. Flinders Petrie had pioneered the practice of typology, that is, arranging artifacts in a chronological framework. Only a few examples of a particular type were necessary to create the typology.

Following the excavation, annotations were added to the Record of Objects. Where accession information was available, the relevant museum numbers were added for artifacts in the Kelsey Museum Collections or in the Cairo Museum. Papyri and coins, which had been read or studied in Michigan, were annotated, and dates were added where chronological information could be ascertained.

2.b.6 Special Finds

2.b.6.i Pottery

All finds of ceramic vessels were recorded using a specialized notation that the excavators developed in the field. For example, in house B224, room B, excavators discovered 14 ceramic vessels or fragments of ceramic vessels, including types 1 a; 55 XI c 1; and 149 I a. (Figure 12) Two separate numbers are given, an Arabic number (such as 149), and a number that is given as Roman numeral – letter – number (such as XI c 1). The Arabic number represents a subset of each Roman numeral type, that is, type I a encompassed multiple types, including 71, 149, 239, 251, 262, 284, 285, and 536. This typology was specifically created to work with the Karanis evidence as it does not seem to be related to other finds of ceramics in the Fayum region.

2.b.6.ii Papyri

The papyri from Karanis were transferred to the papyrus collection at the University of Michigan Libraries Special Collections, and have been the subject of numerous specialized studies. (Gagos, Gates, and Wilburn 2007; Verhoogt 2017) Until recently, few studies of the papyri have been concerned with the precise provenance of the documents. (Claytor and Verhoogt 2018) The digitization of the papyrus collection, undertaken as part of the APIS project under Traianos Gagos, linked the Michigan papyri with their findspots. Regrettably, the two new aggregators, Papyrological Navigator and Trismegistos.org have not retained provenience in the catalog records.

2.b.8 The Formal Publication of the Excavation

On Peterson's return to Ann Arbor at the end of the 1935 season, the work of dealing with the massive amount of accumulated data began in earnest. Peterson spent the next 35 years working on this material, producing an 867-page manuscript on the architecture and topography of the site that was typeset and edited in 1973. This document was never published. Elinor Husselman soon undertook the task of editing it, producing a condensed report on the topography and architecture in 1979, a year after Peterson's death. Other publications from the site followed an approach that is typical of traditional archaeology – the presentation of specialist reports on individual classes of artifacts. Indeed, the period between the 1930s and 1990s witnessed the output of separate volumes on glass, coins, ceramics, lamps, and terracotta figurines as well as multiple volumes of papyri. To this day, the site lacks a final comprehensive publication of the topography and architecture, and much of the work of re-assessing and presenting the data remains to be done.

3. History and Scope of the Project

In 2004, through a fellowship from the Rackham School of Graduate Studies at the University of Michigan and the Collaboratory for Advanced Research and Academic Technologies, I began the process of digitizing the Karanis maps within ArcGIS. After a brief hiatus, I returned to the project at Oberlin College and began to bring undergraduate researchers into the project as individual and capstone projects, and subsequently used the project to introduce students to computational and database software within the classroom.

3.a The GIS Map of the Site

The first step of the project was to create a new GIS map of the site with increased flexibility to permit restructuring of the levels and units. I worked closely with Ryan Reynolds (OC '14, currently, a PhD candidate at UC Berkeley) over multiple years to complete the full mapping of the site, creating separate layers for each insula block found in each of the five occupation levels. Within one architectural enclosure, there may be multiple rooms, and within each room, multiple levels. The excavators also recorded material discovered beneath floor level with a '*'. Our method for representing these units was to draw a polygon within each room or area. This polygon acts as a container that is linked to the database, allowing us to populate the map with objects associated as precisely as possible with their unit location. This permits users to determine which objects were found together.

The process of digitizing the maps was complicated. Some of the individual maps had been scanned, or had been photographed previously. For others, it was necessary for me to travel to the University of Michigan and scan the maps, often on the large format roll scanner at the Map library. The scanned maps first had to be georeferenced, that is, the boundaries of the map were identified in space, because it is not possible to import images (in raster format) into the GIS framework. Then, the individual features of the map (walls, stairs, doorways, etc) were plotted using a mouse to define each feature as a vector-based polygon. (Figure 13) These polygons were then labelled in the GIS with the excavation year, level and room numbers in two formats. The first label preserves the numbering of the original excavation notebooks, while the second adds a new locus number that is machine readable (see below, section 3.b.2).

3.b. Moving the finds into a better database storage software: PostgreSQL

At the University of Michigan, the staff at the Kelsey Museum had entered each of the artifact records from the Records of Objects into a large Excel spreadsheet. This dataset provides a repository for much of the information in the Records of Objects, but it is not easily searchable, and it cannot be linked directly to the GIS data. Through consultation with GIS specialists, I decided to migrate the excel data to PostgreSQL, a multiplatform, open-source database system that is widely used in industry and in the geographic community.

Database tables in PostgreSQL reside on a server that is stored in a dedicated datafile on the host computer. The data is accessed from a client side either through a Graphical User Interface (GUI), or, more commonly, through the command line. In order to move the artifact data into the database, I developed a rough structure and created the database using SQL statements. (Figure 14)

I initially installed PostgreSQL on a server machine that was resident in my office. This server would also be used for ArcGIS for Server, to permit the Karanis GIS to reach the wider world (see below). A power surge prompted a major server crash in 2017, which in turn corrupted the server. I subsequently re-established the server and the PostgreSQL database on a more stable system setup hosted by the Oberlin College Computer Science department.

3.b.1. The organization of the database

The PostgreSQL database was initially envisioned as a simple repository to hold the finds' information from each locus. In the past year, however, I have been developing a more coherent and flexible structure for the database, aligning it with metadata standards outlined in Archaeocore (http://www.ifaresearch.org/archaeocore/), established by the Institute for Fine Arts at NYU, and the Archaeological Resource Cataloging System (https://github.com/matrix-msu/ARCSCore), developed by Ethan Watrall and Jon Frey at Michigan State University.

3.b.2 Renumbering

One of the most challenging problems for integrating the GIS maps and the records of objects has been the notation system applied to each of the artifacts. For scholars consulting the excavation records, the system developed by the UM excavators was workable, if sometimes confusing. Translating the excavation records into a digital context, however, is complicated by the fact that the excavation numbers are not machine readable. Computer systems read from left to right, and sort according to the characters that are encountered sequentially. Alphabetic characters are not easily differentiated. The character * is disastrous for the machine environment, as a search for * returns all entries. Solving these problems has been complicated.

The project developed a consistent renaming strategy. The original field number is retained in the digital records, but a new field has been added to convert the field number to machine (and human) readable text. The discrete parts of the field number are maintained, separated by decimal places. Alphabetic characters beyond 26, such as AA or AG, are resolved into numerical representation. The * is converted into a ^, which is not frequently used in computer languages. An example of the new naming format for 29-B224B*-F is shown here:

	Exc Year	Layer	Building	Room	Under Surface	Artifact designation
Old Numbering	29	B	224	B	*	-T
New numbering for GIS		B_	0224.	002.	^	
New numbering for database		B_	0224.	002.	^_	020

Ceramic finds are differentiated from other artifacts through the use of "c" prior to the number, so that 30-C113AB-b would resolve to C_0113.027_c002.

There are more than 34,000 records in the excel spreadsheet of finds. To work with these very large datasets, we have been transforming the data using OpenRefine, a web-based open-source tool. Data are exported as csv files, and then manipulated within OpenRefine using the GREL language. The field numbers can be split by character type or other signifiers, padded with leading zeros or manipulated in other ways, and then concatenated to achieve the ArtifactNumberNew designation. (Figure 15) Even with the automation provided by OpenRefine, this is still a time-consuming process. The resolved data can then be imported back into the PostgreSQL database.

In preparing the database, we also discovered that there are numerous instances where multiple artifacts (such as coins or papyri) were assigned a single field number. In these cases, it is necessary to replicate the records, creating unique field numbers that append .001, .002, and so on to the record.

3.b.3 Standardization and the use of Keywords

OpenRefine has also been used to standardize the terminology used in the Record of Objects, which includes multiple terms to denote the same classes of objects. We are employing the Getty Museum's Art and Architecture standardized terminology to update the Karanis records and facilitate accurate searches of the material.

3.c Linking the GIS and the database

GIS polygons within each architectural space function as containers into which the artifact records can be stored and accessed. This creates a one-to-many relationship between each room (e.g. C0272A or B224B) and the artifacts that are associated with the fill inside the archaeological unit. When I initially began working with the database system, ArcGIS did not provide native support for PostgreSQL, and it was necessary to add PostgreSQL client libraries to ArcGIS and configure the client and server. Each of the GIS polygons representing the rooms has been labelled with the new, expanded field reference; this number is paralleled in each of the database records in the "locus_number_new" column. The GIS polygons are associated with the database through a "relate," which permits one-to-many relationships; "joins" only allow one-to-one relationships.

4. Outcomes of the Karanis Housing Project

4.a The In-House Geographic Information System

The GIS permits users to view the full extent of the site, and to consider each layer separately. Users can understand which excavation loci overlay one another with reference to the different levels of the site, and how discrete parts of the site interrelate. (Figures 16, 17, and 18)

Once the GIS map had been completed, I selected a small dataset to use in a pilot for integrating the map and the finds. Building B224, a discrete structure that was associated with both artifacts and papyri, was known as the "House of the Nilometer." This building is situated on the western side of the site, and is near to a number of other structures. The building has a number of other features that make it appropriate as a test case, most notably the discovery of material beneath the floors, which necessitated creating multiple polygon containers within the same architectural unit in the GIS. The goal of the pilot was to permit users to query items found within each of the units, and to search for specific objects (e.g. "socks" or "figurines") across the site.

We processed the data from the architectural unit and created a new PostgreSQL database, karanis.public.finds.b224. After linking the GIS database to the database, we created relates between the two parallel fields. Users can select a polygon in the GIS map and view all of the artifacts found in the locus. (Figure 19) Searching across the site for instances of finds and mapping their distribution has proved a more challenging task because of difficulties associated with the one-to-many relationship. Each artifact in the PostgreSQL database is represented by a point placed within the center of the polygon, using a geodatabase that is resident within the GIS. This allows users to search for artifacts and papyri across the site. (Figure 20) At present, because each artifact is represented by a single point, it is not possible to differentiate between single and multiple occurrences of artifacts with the same characteristics. I am exploring potential solutions: either placing the point at a random place within the polygon, which may give a false sense of precise geospatial location, or finding a way to represent multiple instances through different symbology, such as larger points. The newest version of ArcGIS (10.6), which includes native support for PostgreSQL integration, may have added features that will provide a solution.

I have shared this work with other scholars, who have used it for research and in publications.

4.b Map Data Served over the World Wide Web

A goal of the project is to make the data from the GIS and the artifact database accessible to the wider world. I worked extensively with ArcGIS for Server and installed a server on a host machine that was resident on a computer in my office. At around this same time, ArcGIS Online was released, and I migrated our work to that platform. Other digital humanities projects related to the ancient world, such as the Pompeii Bibliography and Mapping Project (https://digitalhumanities.umass.edu/pbmp/) also use this service. We simplified the GIS shapefiles so that they could be displayed more easily. The pilot building, B224, and its surrounding area are available through

http://oberlincollege.maps.arcgis.com/apps/webappviewer/index.html?id=8637a1bf499a4d11be88d5b2b0375609

Users can adjust the layers that are displayed and view the objects found within B224 by clicking on one of the three circles positioned at the center of each room.

As we uploaded the data, one problem came into sharper focus: our maps were developed using the 1920s maps, which do not use a modern map projection. The Karanis plans are shifted by one to two meters, a difficulty with precision that we are attempting to solve, in part through collaboration. This error does not mitigate the success of the project goals, but does prevent the online map from lining up precisely with satellite imagery.

4.c. The Karanis Housing Project Website

In 2015, I worked with a student from the Oberlin Computer Science department to develop and implement a website to host the data and provide a summary of the project. karanishousingproject.org includes multiple pages that provide an overview of the project as a whole; a digital map of B224 that is populated with artifacts hosted through ArcGIS online; a basic catalog of the ceramics from the site; and a page with brief biographies of the project members. The website is intended to showcase the possibilities offered by the project, and is still under development. As time permits, I am adding data to ArcGIS Online and revising the webpage to include information about my current students and the contemporary direction of the project.

4.d. The Ceramic Database

When I began working on the project, it was not possible to translate the notations used to record the ceramic finds into vessel types, as the key had been lost. Understanding the ceramic finds is critical, because in contemporary archaeology, ceramic vessels are one of the key elements used for dating an archaeological unit. The publication of the ceramics (Johnson 1981) was undertaken without reference to the in-field typology. In 2012, after extensive research in the Michigan archives, I located the ceramic key and associated line drawings and profile images, which can now be used to cross-reference the Records of Objects. (Figure 21)

The rediscovery of the ceramic "key" meant that it would be possible to discover the meaning of the enigmatic codes that the excavators used in their ceramic typology. Thomas Landvatter, at Reed College, has been working with the key and with the vessels at Michigan to reassess the ceramic assignments made by the original excavators. Samantha Mater (OC '15), a Computer Science and Archaeological Studies major, developed a database in PostgreSQL that associated the ceramic designations with descriptions of each ceramic vessel type. The database has been incorporated into a web interface based around PHP. Users can query the ceramic numbers, either by the Roman numeral code, or by the Arabic number subset. Each query also produces a profile view of each pot type.

5. Digital humanities/GIS/database training for students

5.a. Research Students

Between 2012 and 2014, I worked with Ryan Reynolds (OC '14) to build the basemaps for this study through ArcGIS. By the end of AY 2013-2014, Ryan had digitized all layers from the original excavations into a functional Geographic Information System. From this point, the project moved into the pilot phase with the B224 data, which involved (1) linking the excavation records to the digital map through PostgreSQL, (2) creating a separate database of pottery records, and (3) making this material available on the web through ArcGIS for Server. I worked with two research students, Miranda Rutherford (OC '15), who worked on the PostgreSQL database, and Olivia Fountain (OC '17) who moved the GIS data to ArcGIS online. Samantha Mater (OC '15) worked on the ceramic material as her Senior capstone project in Archaeological Studies. I detailed Samantha's work on the ceramic database above.

During Spring 2016, I worked with Susanna Faas-Bush (OC '18) and Aaron Henry (OC '18) to process more finds data from the excavation using OpenRefine, and integrate this data into the GIS. This past year, I worked with Christian Bolles (OC '18) to refine the GIS maps by standardizing the room designations and adding additional shapefiles to represent successive units within a single archaeological enclosure. This will permit links to be created with the in-house database across a wider proportion of the site. I am currently working with Will Lynch (OC '21) to refine the database structure, process the finds data, and add these data to the database.

The students who have served as research assistants for the project have done well since graduation, using their experiences as a foundation for their first jobs or for admission to graduate school. Ryan Reynolds is completing his PhD in the Classical Archaeology program at UC Berkeley, where he became known as “The GIS guy.” Miranda Rutherford is currently in law school at Berkeley, with strong interests in digital rights protection, having served as a paralegal at a firm focused on immigration. Samantha Mater was hired by Epic Systems in Madison, WI, and worked with database development, largely because of her experience with PostgreSQL. Olivia Fountain, following graduation, worked at the Allen Memorial Art Museum as the curatorial assistant for academic programs. Christian Bolles spent summer 2018 as a paid provenance researcher at the Allen Memorial Art Museum. Susanna Faas-Bush is now a graduate student at UC Berkeley, also in the Classical Archaeology program.

5.b. Classroom use and Student Research Projects

In 2016, I worked with my advanced seminar class, “Roman Egypt: Art, History Culture” to model new ways to use the Karanis data in a classroom setting. For the final project in the class, many of the students used material from the Karanis database to study the contents of individual buildings, or to explore the distribution of finds across the site. The students built online exhibitions in Omeka to present their findings. This past academic year, Susanna Faas Bush, for her Archaeological Studies senior project, analyzed the occurrence of textiles in two houses, looking specifically at production tools and textile remains. She made extensive use of the GIS data to consider the spatial arrangement of finds.

6. Future work

In the short term, I am beginning to work upwards level by level from the E-Level, the earliest layer, which also has the smallest dataset. Christian Bolles re-numbered the rooms in the GIS and Will Lynch is currently working on some of the E Level finds in the database. It is now necessary to link these data sources, and to upload more GIS data to ArcGIS online. Subsequent work will focus on streamlining the process to add more data. Concurrently, I am looking for venues – both through conferences and publication in journals, where I can present the recent work and solicit feedback from others in the field.

I have been exploring a collaboration with Dr. Emily Cole, a postdoctoral fellow at the Center for the Tebtynis Papyri at Berkeley, and Dr. Bethany Simpson, University of California, Los Angeles, both of whom worked on the recent Karanis excavations overseen by Prof. Willeke Wendrich at UCLA. The UCLA excavations in the early 2000s created a systematic map of the site using the datum points that were originally installed by the Michigan team. Working together, it might be possible to integrate the PostgreSQL data into a more accurate map of the site.

A key area of growth in the project is incorporating chronological information. As noted above, one of the main goals of the ceramic component is to correlate the excavation-assigned pottery types with contemporary ceramic typologies. This will permit us to associate each ceramic vessel with a period of use. Dates can also be associated with other artifacts, most notably papyri and coins; these data can be incorporated into the database as well, using pre-existing fields. Once this has been completed, an algorithm can be generated to assign a date to each excavation unit by selecting the latest datable material from the artifacts recorded in each unit. These data can then be employed to study the insulae, or groups of houses, and may yield a method to reassess the chronology of portions of the site.

With this project as with others, is necessary to ensure the continued life of the data by committing to long-term data storage solutions such as OpenContext. I have been exploring funding sources to facilitate this, and included data storage as one element in my NEH applications.

One of the most important tenets in contemporary digital work in the humanities is integration across projects. Researchers who study the ancient world are committed to linked open data, and digital projects increasingly recognize the value of interconnectedness for the goal of data sustainability (Elliot, Heath, and Muccigrosso 2012, Kansa 2014, Averett *et. al.* 2016). Using the new artifact record numbers, I hope to associate each object record with a URI (Universal Resource Identifier), to facilitate integration with linked open data. Many of the artifacts and papyri from the Karanis excavations are accessible online, available either through the Kelsey Museum of Archaeology object database or through papyri.info and trismegistos.org. I have had some success adding URLs to some of the artifact records, but this process may be complicated by the lack of provenience information in trismegistos.org and papyri.info. I am also identifying ways to associate spatial data with the finds records through integration with the Pleiades Project (https://pleiades.stoa.org/), a gazetteer of ancient places. These steps will further integrate the Karanis data sets into large, open data repositories, making the Karanis excavation records available to a broad segment of the population.

7. Conclusion

I embarked on the Karanis Housing Project to showcase one avenue for reimagining legacy datasets in a modern framework and to demonstrate the value of large-scale digitization of archaeological finds. The project sought to demonstrate proof of concept and to identifying problems with large-scale digitization of legacy archaeological data. In pursuit of these ends, I have achieved a number of important benchmarks:

1. Creation of full GIS of site
2. Creation of specific information about B224
3. Online presence through website

As it moves forward, the project looks toward ways to present our findings and to expand the data that is housed in the project to benefit the broader scholarly community by making these important datasets freely available.

8. Bibliography

Allison, P.M. 2008. “Dealing with Legacy Data - an Introduction.” *Internet Archaeology* 24.

Averett, E.W., J.M. Gordon and D.B. Counts. 2016. *Mobilizing the Past for a Digital Future.* Digital Press at the University of North Dakota.

Bagnall, R.S., D. Rathbone and J. Paul Getty Museum. 2004, *Egypt from Alexander to the early Christians : an archaeological and historical guide,* The J. Paul Getty Museum, Los Angeles, Calif.

Bevan, A. 2015. “The Data Deluge.” *Antiquity* 89 (348). 1473-84.

Claytor, W.G. and A.Verhoogt. 2018, *Papyri from Karanis : the Granary C123 (P. Mich. XXI),* University of Michigan Press, Ann Arbor.

Cooper, A., and C. Green. 2015. “Embracing the Complexities of ‘Big Data’ in Archaeology: The Case of the English Landscape and Identities Project.” *Journal of Archaeological Method and Theory*, February, 1–34..

Elliott, T., S. Heath, and J. Muccigrosso. 2012. "Report on the Linked Ancient World Data Institute." *Information Standards Quarterly* 24: 43-45.

Gagos, T., Gates, J. & Wilburn, A. 2005. "Material culture and texts of Graeco-Roman Egypt: creating context, debating meaning", *Bulletin of the American Society of Papyrologists,* 42: 171-188.

Gazda, E.K. & Wilfong, T.G. 2004. *Karanis, an Egyptian town in Roman times: discoveries of the University of Michigan expedition to Egypt (1924-1935),* Kelsey Museum of Archaeology, Ann Arbor.

Husselman, E.M. 1979. *Karanis excavations of the University of Michigan in Egypt 1928-1935: topography and architecture: a summary of the reports of the director, Enoch E. Peterson,* University of Michigan Press, Ann Arbor.

Johnson, B. 1981. *Pottery from Karanis: excavations of the University of Michigan,* University of Michigan Press, Ann Arbor.

Kansa, Eric. 2014. “Open Context and Linked Data.” *ISAW Papers* 7 (10).

Kintigh, K.W., J.H. Altschul, M.C. Beaudry, R.D. Drennan, A.P. Kinzig, T.A. Kohler, W.F. Limp, *et al*. 2014. “Grand Challenges for Archaeology.” *Proceedings of the National Academy of Sciences* 111 (3): 879–80.

Pedley, J.G. 2012. *The life and work of Francis Willey Kelsey: archaeology, antiquity, and the arts,* University of Michigan Press, Ann Arbor.

Pollard, N. 1998. "The chronology and economic condition of late Roman Karanis: an archaeological reassessment", *Journal of the American Research Center in Egypt* 35: 147-162.

Verhoogt, A. M. F. W. 2017. *Discarded, discovered, collected: the University of Michigan papyrus collection,* University of Michigan Press, Ann Arbor.

Wesson, C.B., and J.W. Cottier. 2014. “Big Sites, Big Questions, Big Data, Big Problems: Scales of Investigation and Changing Perceptions of Archaeological Practice in the Southeastern United States.” *Bulletin of the History of Archaeology* 24.

Wilfong, T.G. 2014. *Karanis Revealed: Discovering the Past and Present of a Michigan Excavation in Egypt*, University of Michigan Press, Ann Arbor.

Wylie, A. 2016. “How Archaeological Data Bites Back: Strategies for Putting Old Data to Work in New Ways.” *Science, Technology and Human Values.* 1-23.

Figure 1: Aerial photograph of the mound at Kom Aushim (Karanis) showign the area of damage (center) from sebakh mining.

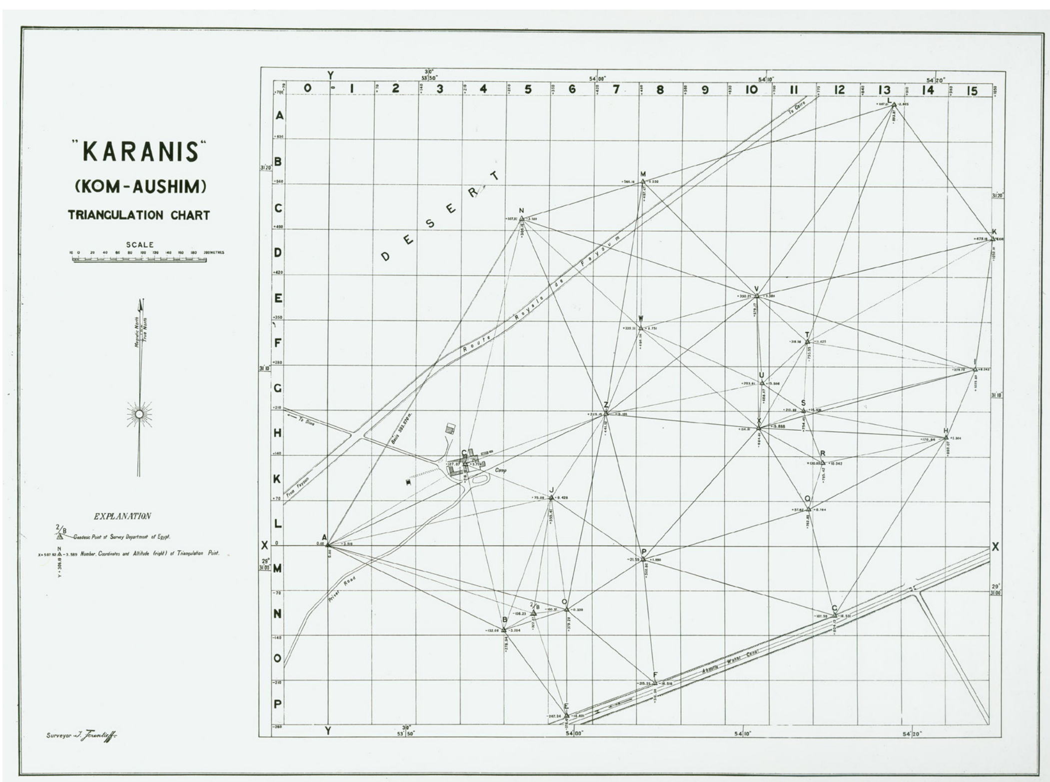


Figure 2: Tringuation map of the site showing the location of datum points and the excavation grid. Excavation squares were designated by the X and Y axes. Datum point Z, for example, was placed in grid square G7.

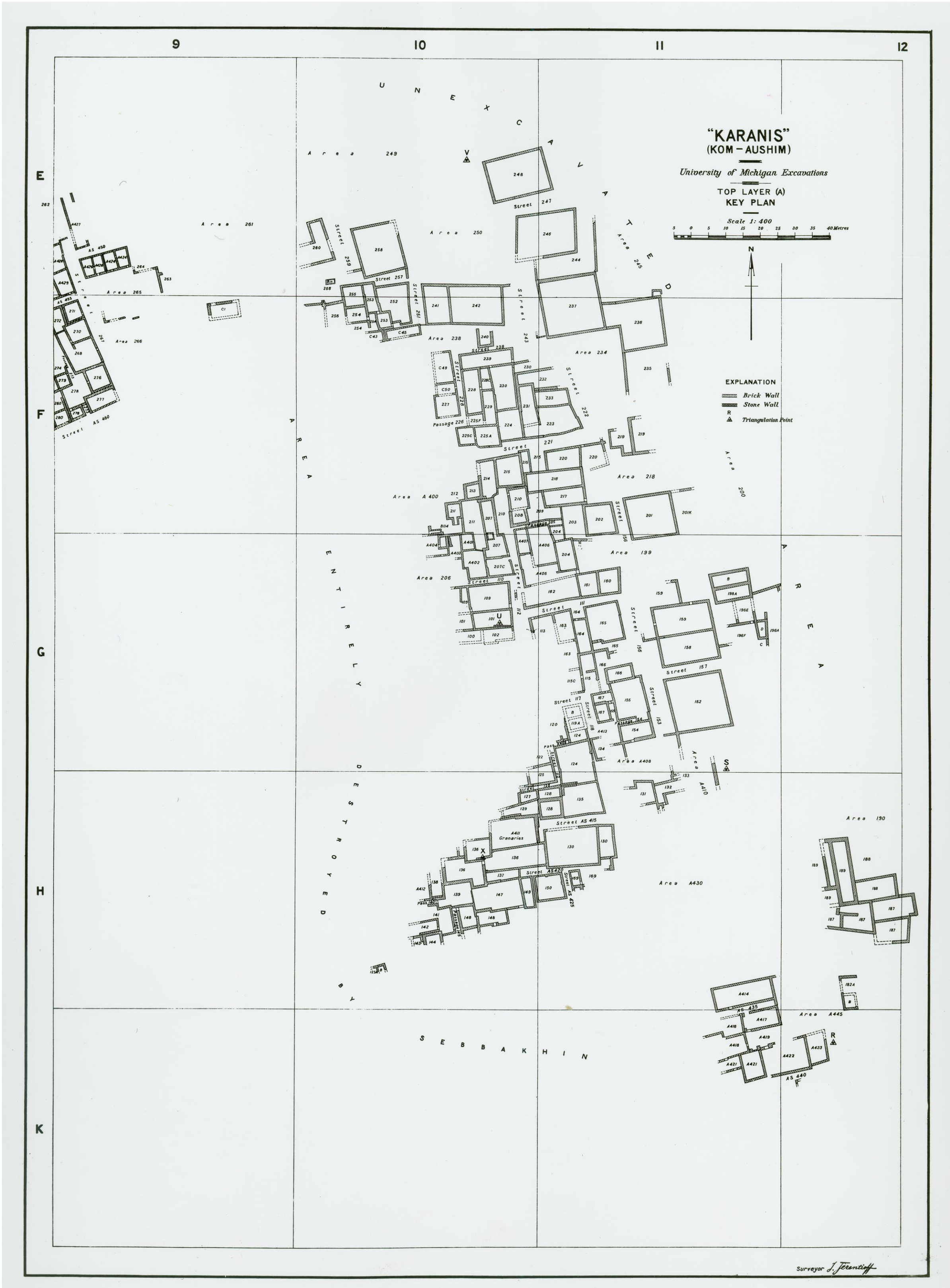


Figure 3: “Key Plan” showing rough outline of all houses and buildings at site preserved in the top (A) layer of occupation.

Figure 4: Section F10 from the top (A) layer, showing top (A) layer construction as hollow, reused walls from second layer (filled with horizontal lines) and third layer (cross-hatched fill), as well as significant features such as ovens, stairs, and bins.

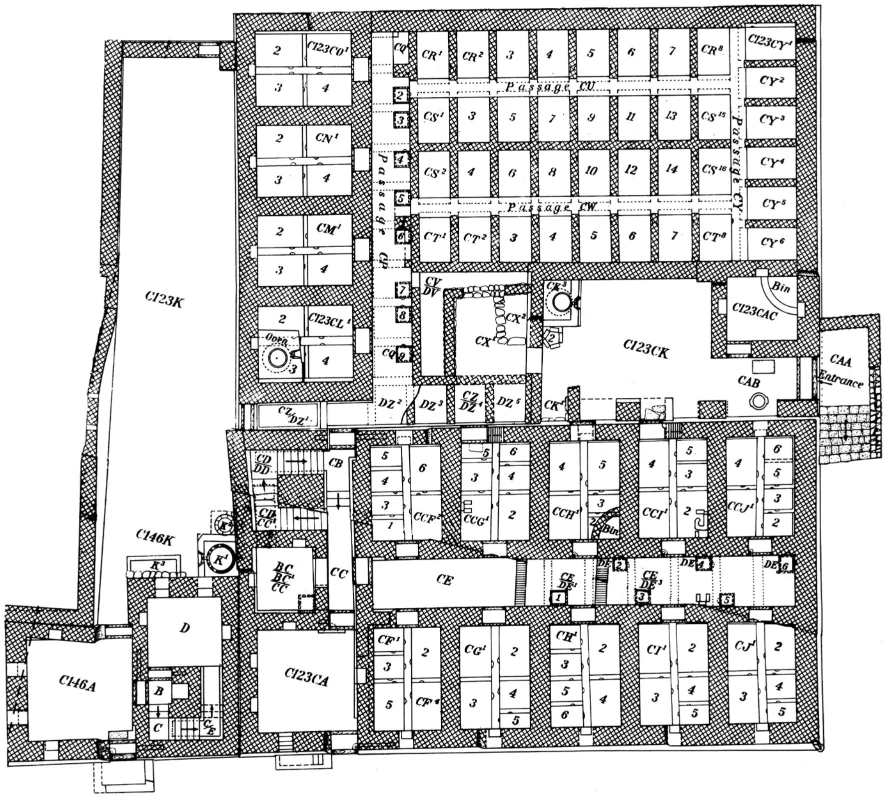


Figure 5: Granary C123, showing placement multiple bins for grain storage. Each feature is clearly labelled in the plan.

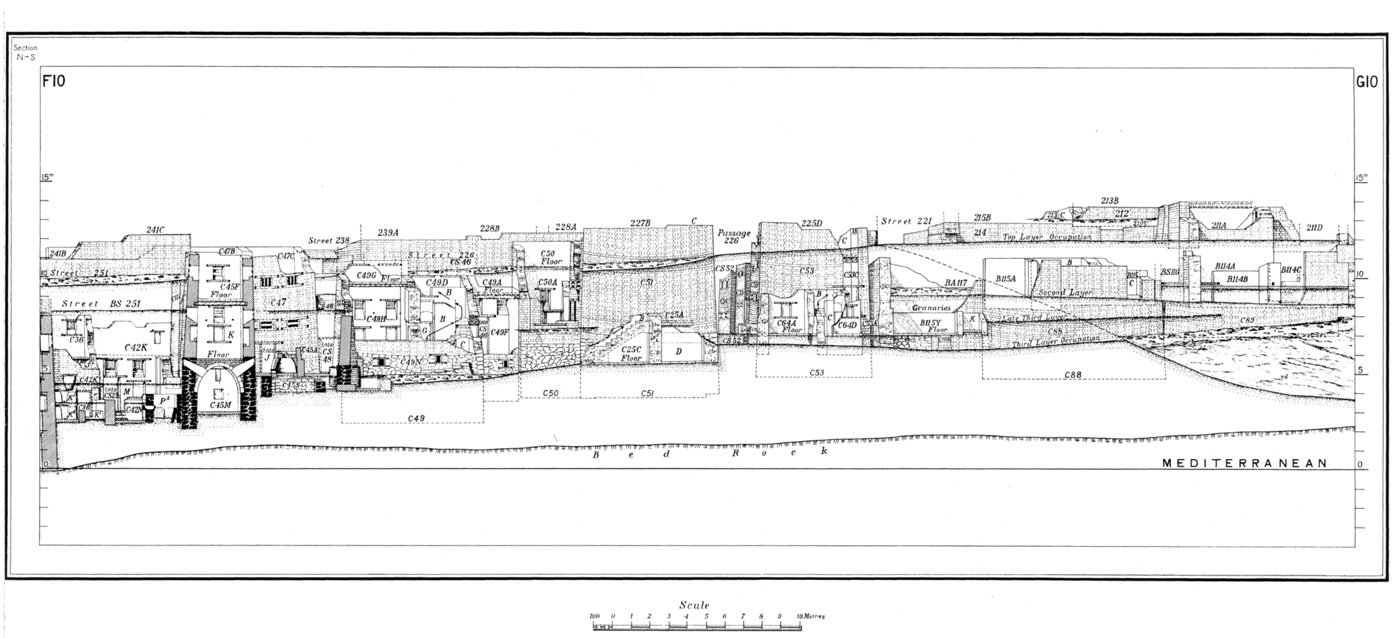


Figure 6: View of sectors F10 and G10, showing elevation and relative height of preserved remains.

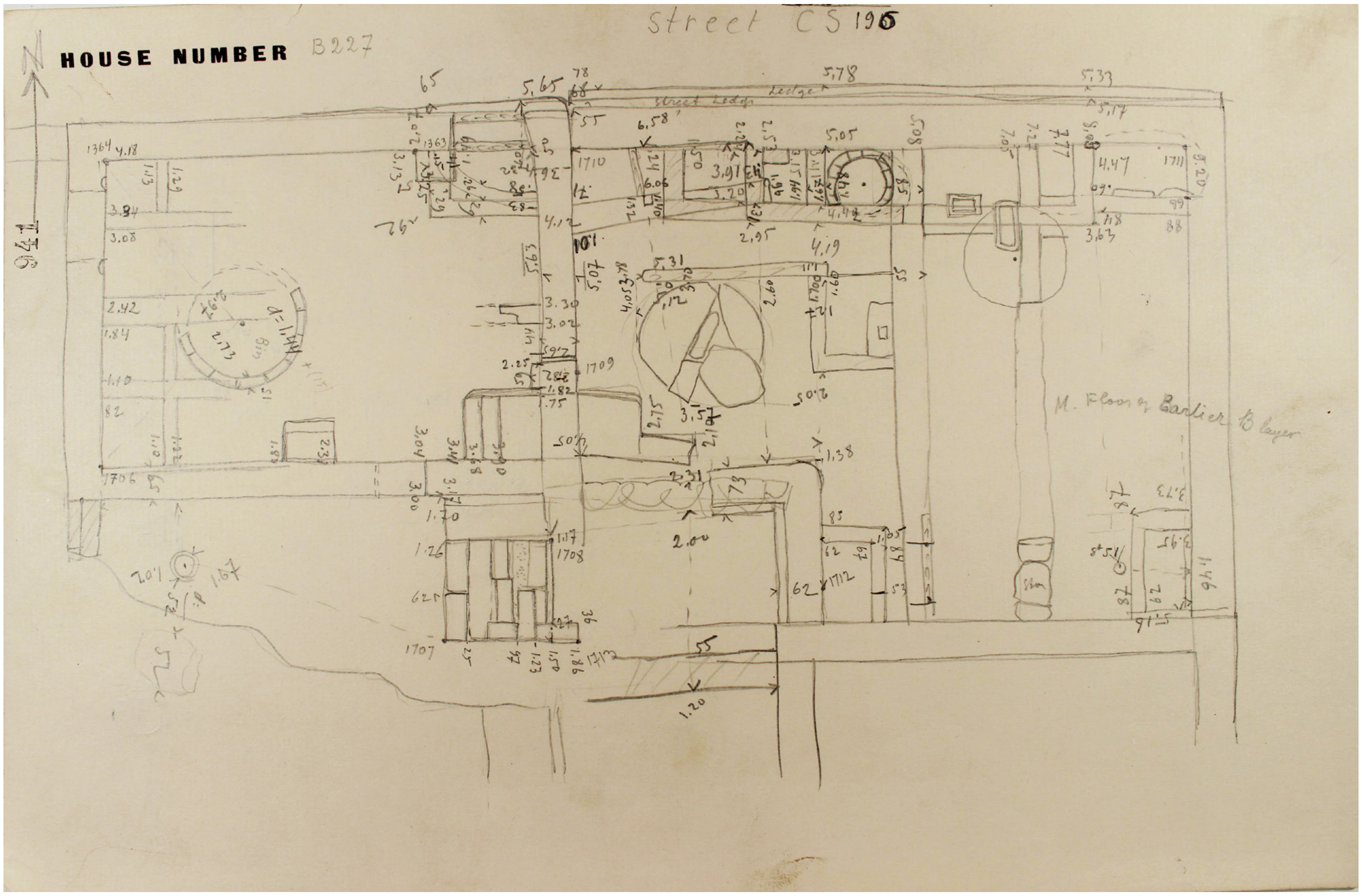


Figure 7: Sketch of house number B227, showing locations and values of elevation points.

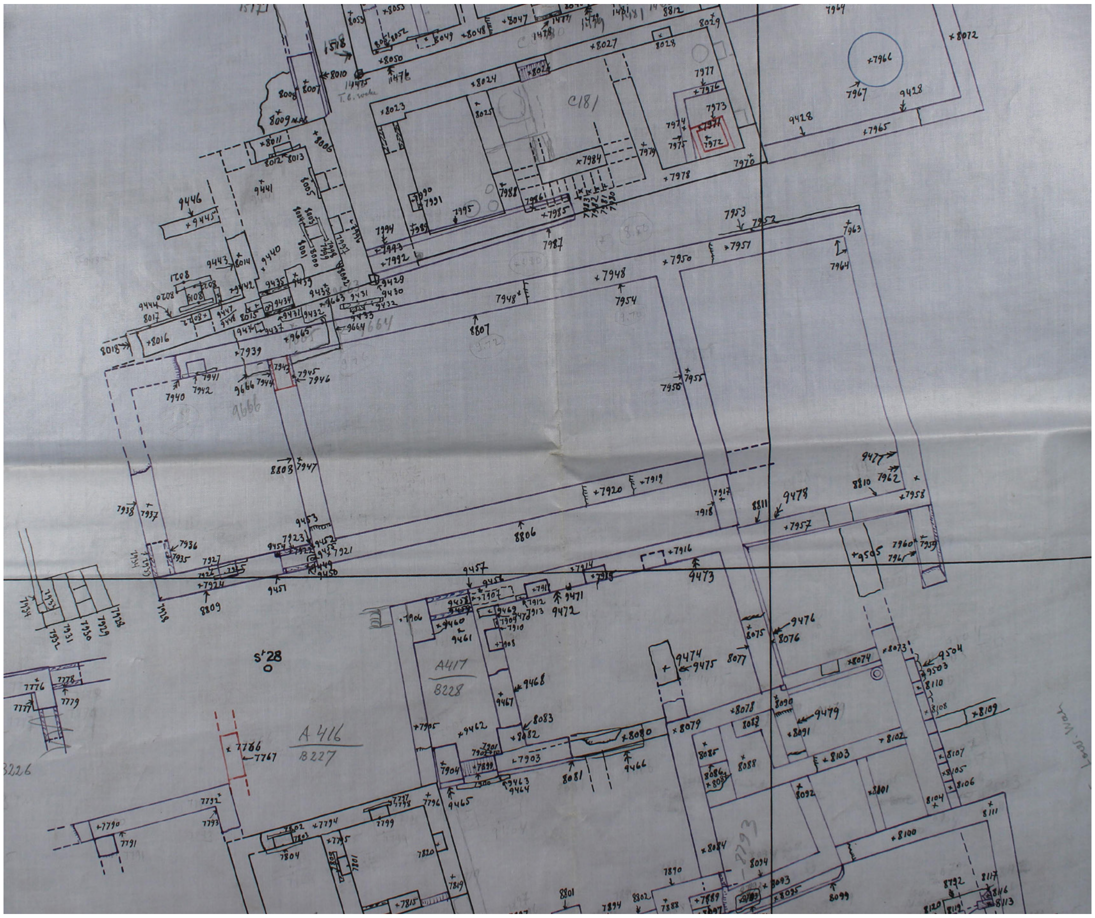
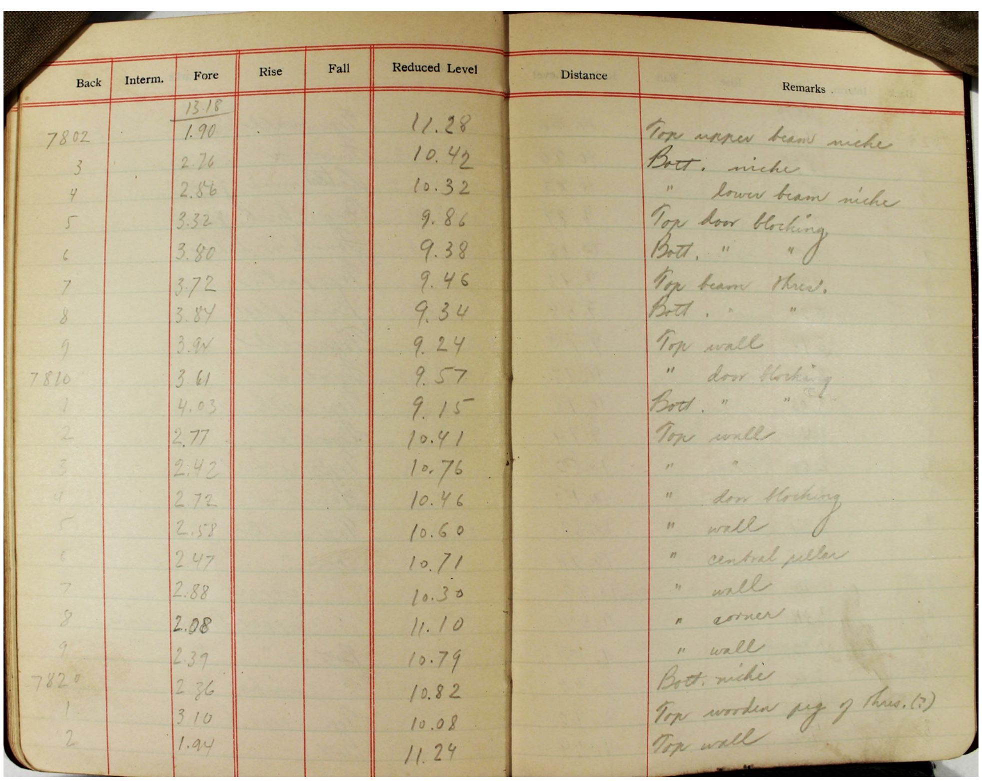


Figure 8: Larger map of B Layer, showing Building B227 at bottom. Elevation points are keyed to the elevation log book, where readings from the theodolite have been resolved for actual m above sea level.

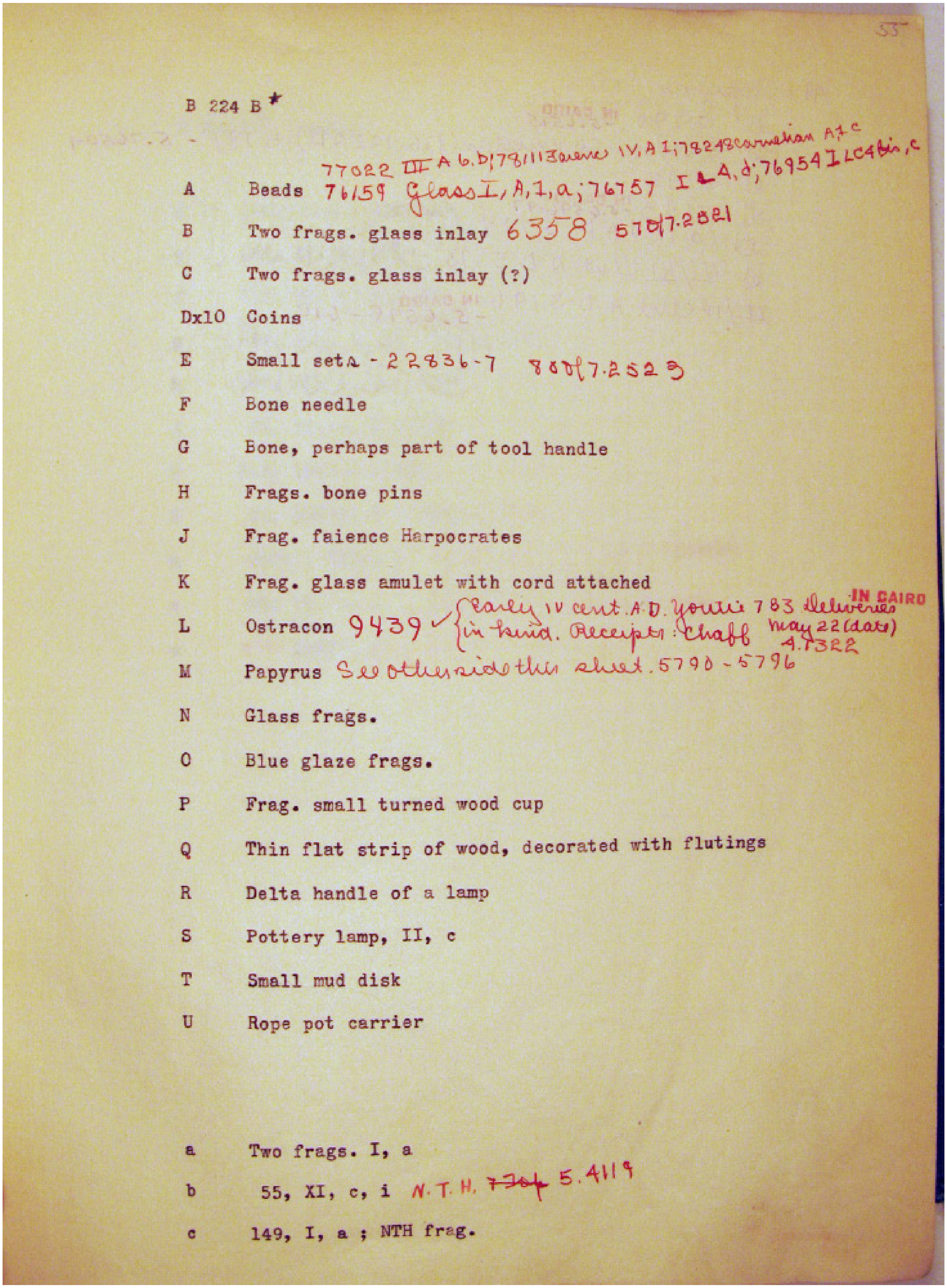


Back	Interm.	Fore	Rise	Fall	Reduced Level	Distance	Remarks
		13.18					
7802		1.90			11.28		Top upper beam niche
3		2.76			10.42		Bott. niche
4		2.86			10.32		" lower beam niche
5		3.32			9.86		Top door blocking
6		3.80			9.38		Bott. " "
7		3.72			9.46		Top beam thres.
8		3.84			9.34		Bott. "
9		3.94			9.24		Top wall
7810		3.61			9.57		" door blocking
1		4.03			9.15		Bott. " "
2		2.77			10.41		Top wall
3		2.42			10.76		" "
4		2.72			10.46		" door blocking
5		2.58			10.60		" wall
6		2.47			10.71		" central pillar
7		2.88			10.30		" wall
8		2.08			11.10		" corner
9		2.39			10.79		" wall
7820		2.36			10.82		Bott. niche
1		3.10			10.08		Top wooden peg of thres. (?)
2		1.94			11.24		Top wall

Figure 9: The elevation log book, recording all of the sightings from the theodolite, with measurements resolved to meters above sea level and descriptions of where the points were taken.

Figure 10: Photograph showing objects discovered in situ. Karanis Photo Archive 5-2831.

Figure 11: Division album photo, showing organic remains, period to division of artifacts between Michigan and the Cairo Museum. Finds have been grouped by type, rather than findspot.



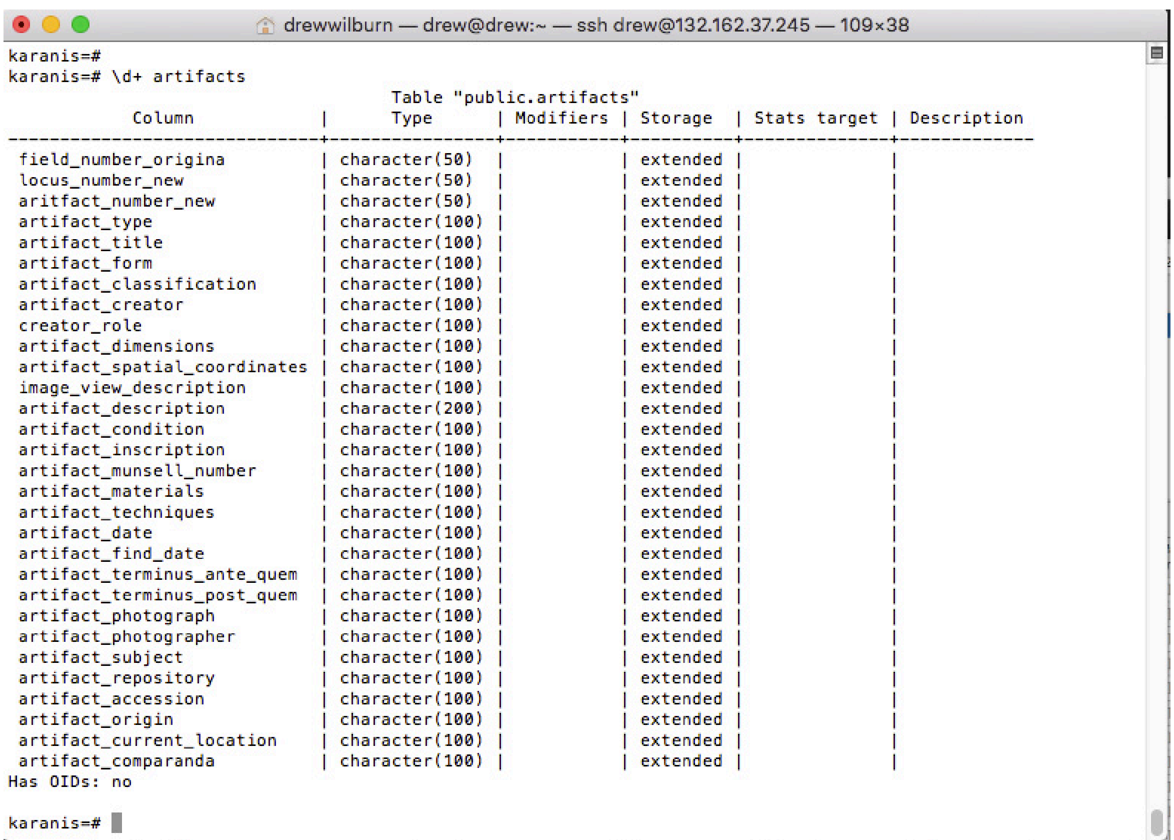
55

B 224 B *

A	Beads	76159 Glass I, A, 1, a; 76757 I L A, d; 76954 I LC4 bis, c; 77022 III A 6, D; 78111 3 areas IV, A 1; 78248 carnelian A 7 c
B	Two frags. glass inlay	6358 5707.2821
C	Two frags. glass inlay (?)	
Dx10	Coins	
E	Small sets - 22836-7	8807.2523
F	Bone needle	
G	Bone, perhaps part of tool handle	
H	Frags. bone pins	
J	Frag. faience Harpocrates	
K	Frag. glass amulet with cord attached	
L	Ostracon	9439 ✓ Early IV cent. A.D. Youths 783 deliveries in kind. Receipts: Chaff May 22 (date) 4.7322 IN CAIRO
M	Papyrus	See other side this sheet. 5790 - 5796
N	Glass frags.	
O	Blue glaze frags.	
P	Frag. small turned wood cup	
Q	Thin flat strip of wood, decorated with flutings	
R	Delta handle of a lamp	
S	Pottery lamp, II, c	
T	Small mud disk	
U	Rope pot carrier	
a	Two frags. I, a	
b	55, XI, c, i	N.T.H. ~~7704~~ 5.4119
c	149, I, a ; NTH frag.	

Figure 12: A typical page from Record of Objects 1930, showing the finds from beneath the floor of Building B224, Room B. Note the annotation in red for the items accessioned and studied at Michigan. The abbreviation N.T.H, visible for item “b” means “Not Taken Home.”

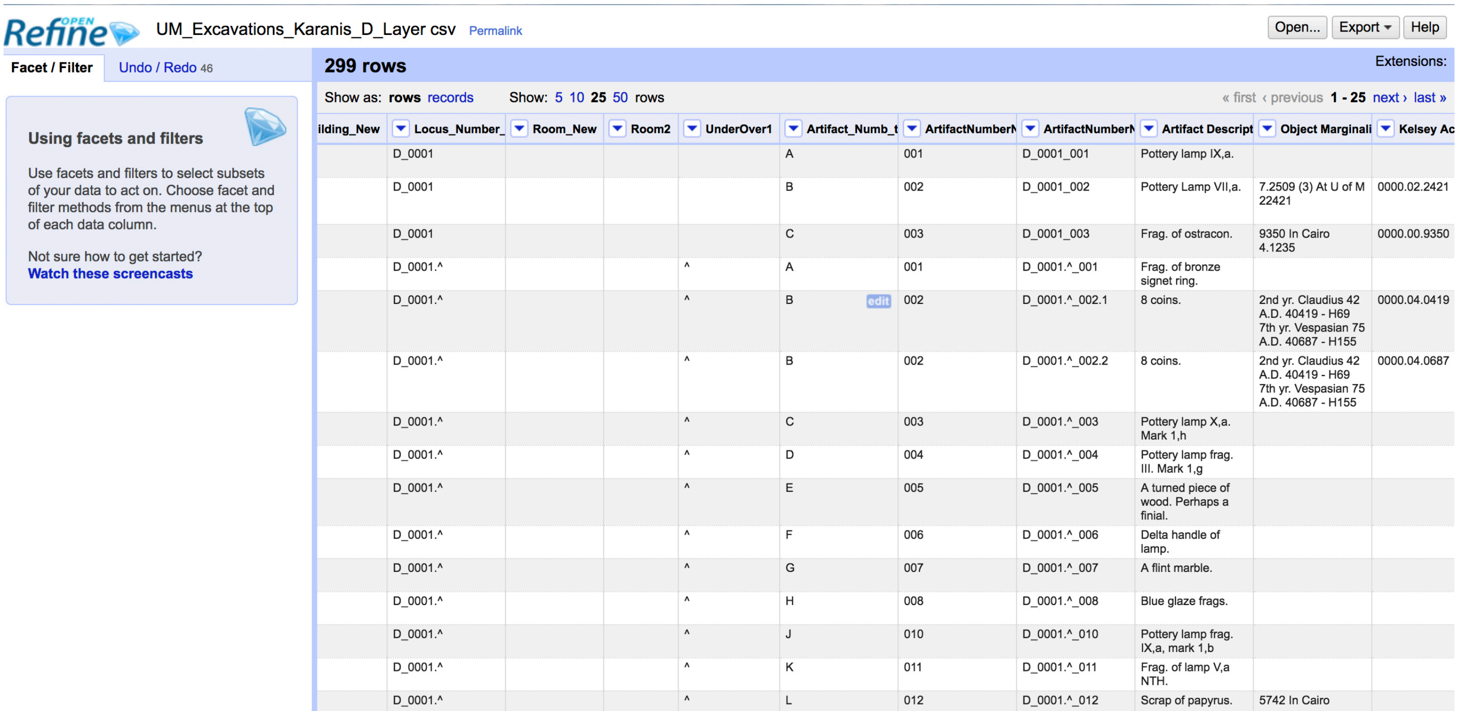
Figure 13: Side by Side comparison of B224 and surrounding insulae, in original plan (left) and new GIS reconstruction (right)



```
drewwilburn — drew@drew:~ — ssh drew@132.162.37.245 — 109×38
karanis=#
karanis=# \d+ artifacts
                                    Table "public.artifacts"
            Column             |      Type      | Modifiers | Storage  | Stats target | Description
-------------------------------+----------------+-----------+----------+--------------+-------------
 field_number_origina          | character(50)  |           | extended |              |
 locus_number_new              | character(50)  |           | extended |              |
 aritfact_number_new           | character(50)  |           | extended |              |
 artifact_type                 | character(100) |           | extended |              |
 artifact_title                | character(100) |           | extended |              |
 artifact_form                 | character(100) |           | extended |              |
 artifact_classification       | character(100) |           | extended |              |
 artifact_creator              | character(100) |           | extended |              |
 creator_role                  | character(100) |           | extended |              |
 artifact_dimensions           | character(100) |           | extended |              |
 artifact_spatial_coordinates  | character(100) |           | extended |              |
 image_view_description        | character(100) |           | extended |              |
 artifact_description          | character(200) |           | extended |              |
 artifact_condition            | character(100) |           | extended |              |
 artifact_inscription          | character(100) |           | extended |              |
 artifact_munsell_number       | character(100) |           | extended |              |
 artifact_materials            | character(100) |           | extended |              |
 artifact_techniques           | character(100) |           | extended |              |
 artifact_date                 | character(100) |           | extended |              |
 artifact_find_date            | character(100) |           | extended |              |
 artifact_terminus_ante_quem   | character(100) |           | extended |              |
 artifact_terminus_post_quem   | character(100) |           | extended |              |
 artifact_photograph           | character(100) |           | extended |              |
 artifact_photographer         | character(100) |           | extended |              |
 artifact_subject              | character(100) |           | extended |              |
 artifact_repository           | character(100) |           | extended |              |
 artifact_accession            | character(100) |           | extended |              |
 artifact_origin               | character(100) |           | extended |              |
 artifact_current_location     | character(100) |           | extended |              |
 artifact_comparanda           | character(100) |           | extended |              |
Has OIDs: no

karanis=#
```

Figure 14: PostgreSQL artifact table, showing column list for each artifact entry. The column designations are based on the Archaeocore Cataloging metadata system standards for artifacts, with the addition of fields for new artifact numbers and locus numbers.



ilding_New	Locus_Number_	Room_New	Room2	UnderOver1	Artifact_Numb_t	ArtifactNumberN	ArtifactNumberN	Artifact Descript	Object Marginali	Kelsey Ac
	D_0001				A	001	D_0001_001	Pottery lamp IX,a.		
	D_0001				B	002	D_0001_002	Pottery Lamp VII,a.	7.2509 (3) At U of M 22421	0000.02.2421
	D_0001				C	003	D_0001_003	Frag. of ostracon.	9350 In Cairo 4.1235	0000.00.9350
	D_0001.^			^	A	001	D_0001.^_001	Frag. of bronze signet ring.		
	D_0001.^			^	B	002	D_0001.^_002.1	8 coins.	2nd yr. Claudius 42 A.D. 40419 - H69 7th yr. Vespasian 75 A.D. 40687 - H155	0000.04.0419
	D_0001.^			^	B	002	D_0001.^_002.2	8 coins.	2nd yr. Claudius 42 A.D. 40419 - H69 7th yr. Vespasian 75 A.D. 40687 - H155	0000.04.0687
	D_0001.^			^	C	003	D_0001.^_003	Pottery lamp X,a. Mark 1,h		
	D_0001.^			^	D	004	D_0001.^_004	Pottery lamp frag. III. Mark 1,g		
	D_0001.^			^	E	005	D_0001.^_005	A turned piece of wood. Perhaps a finial.		
	D_0001.^			^	F	006	D_0001.^_006	Delta handle of lamp.		
	D_0001.^			^	G	007	D_0001.^_007	A flint marble.		
	D_0001.^			^	H	008	D_0001.^_008	Blue glaze frags.		
	D_0001.^			^	J	010	D_0001.^_010	Pottery lamp frag. IX,a, mark 1,b		
	D_0001.^			^	K	011	D_0001.^_011	Frag. of lamp V,a NTH.		
	D_0001.^			^	L	012	D_0001.^_012	Scrap of papyrus.	5742 In Cairo	

Figure 15: Screenshot showing record standardization using OpenRefine.

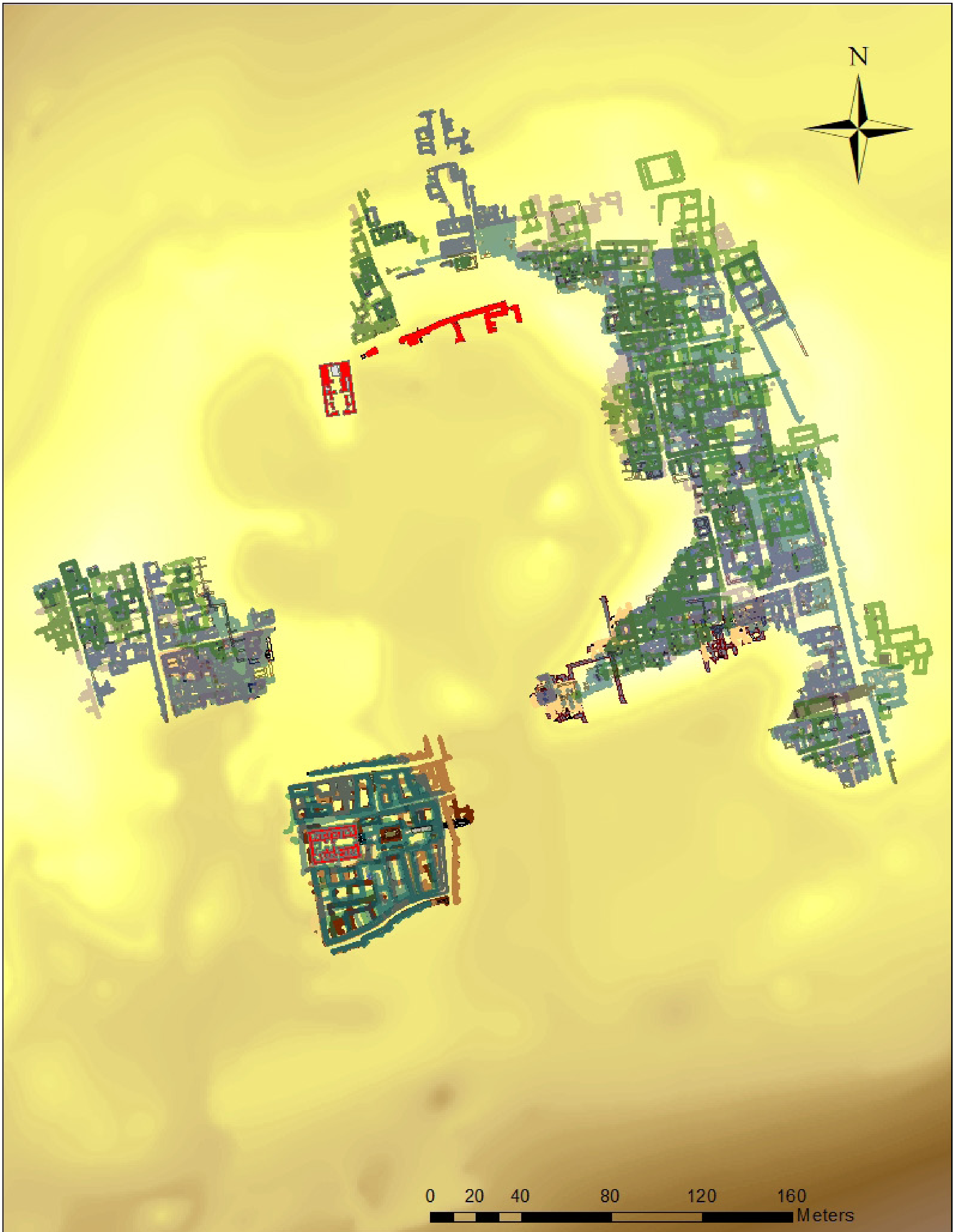
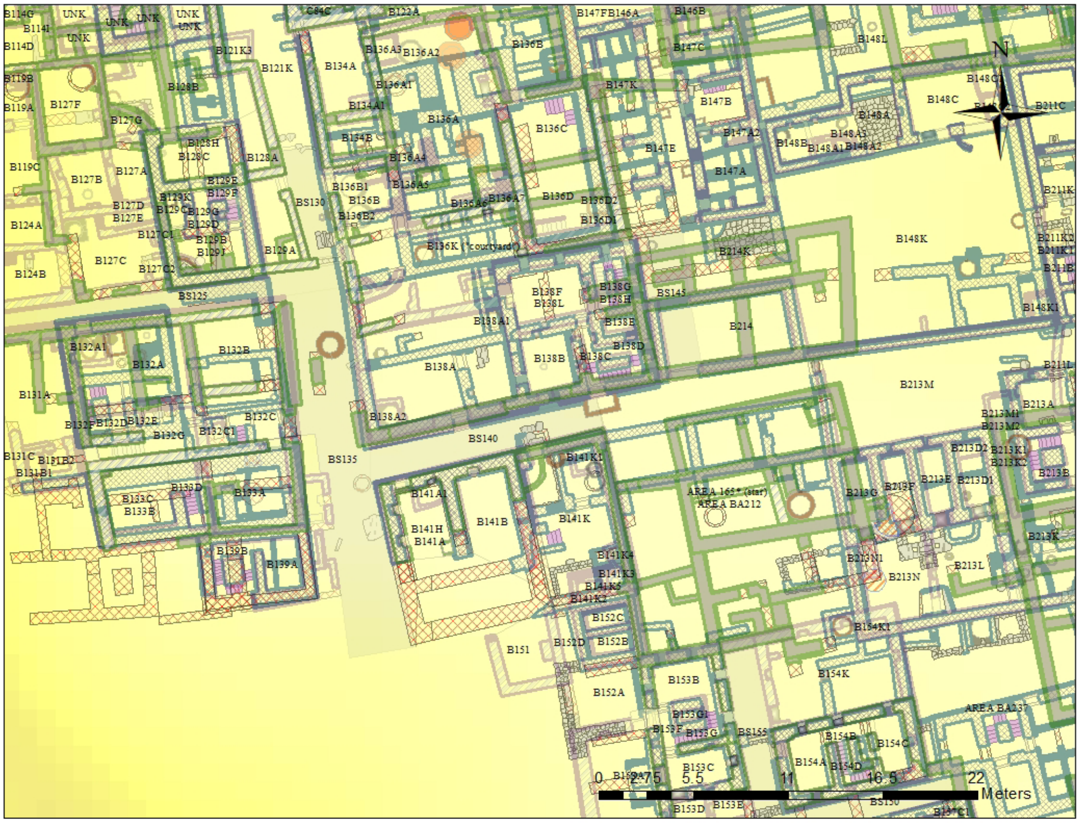


Figure 16: GIS Reconstruction of site of Karanis, showing all excavation layers.



Figures 17 and 18: Top: Detail view of part of site, showing all layers. Bottom: Detail view of part of site, showing B Layer only. Different construction phases are marked by different symbology.

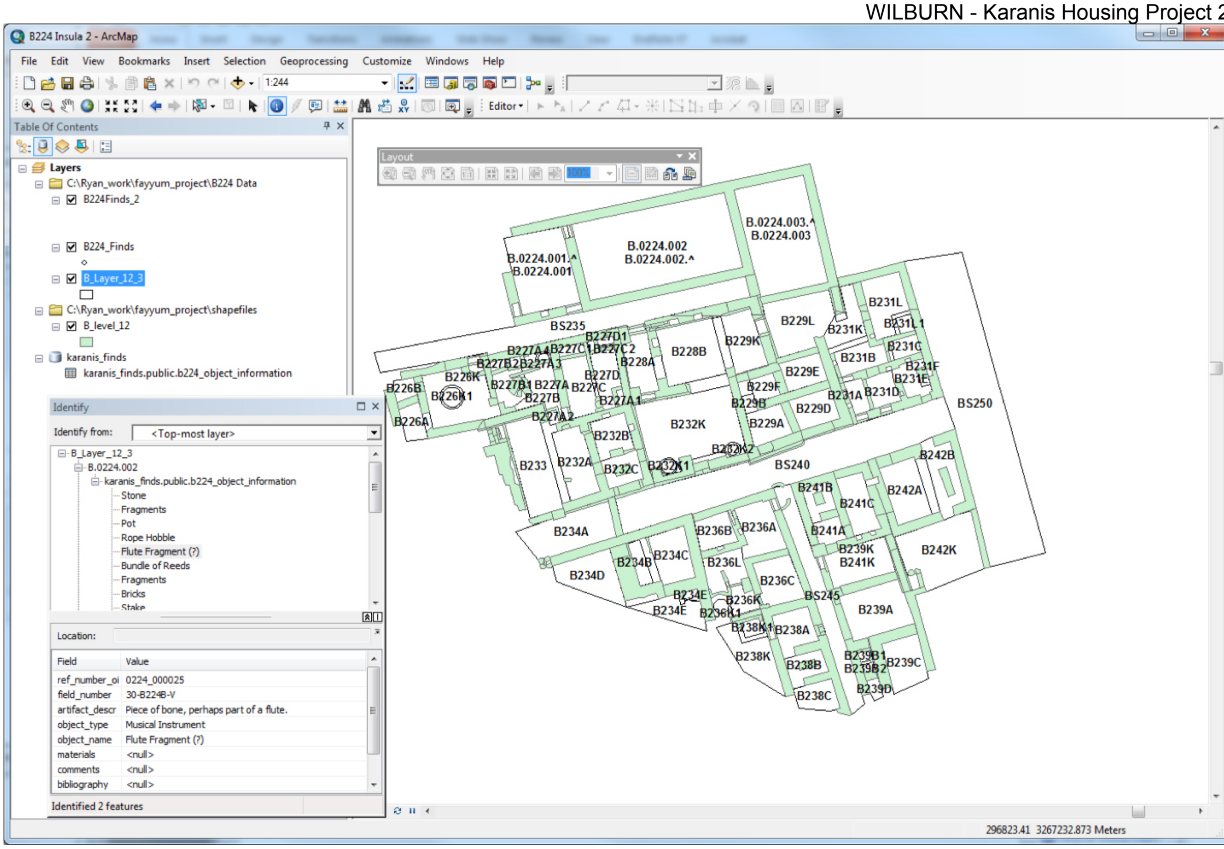


Figure 19: Screenshot showing a query in ArcGIS to retrieve all the finds recorded for unit B224B.

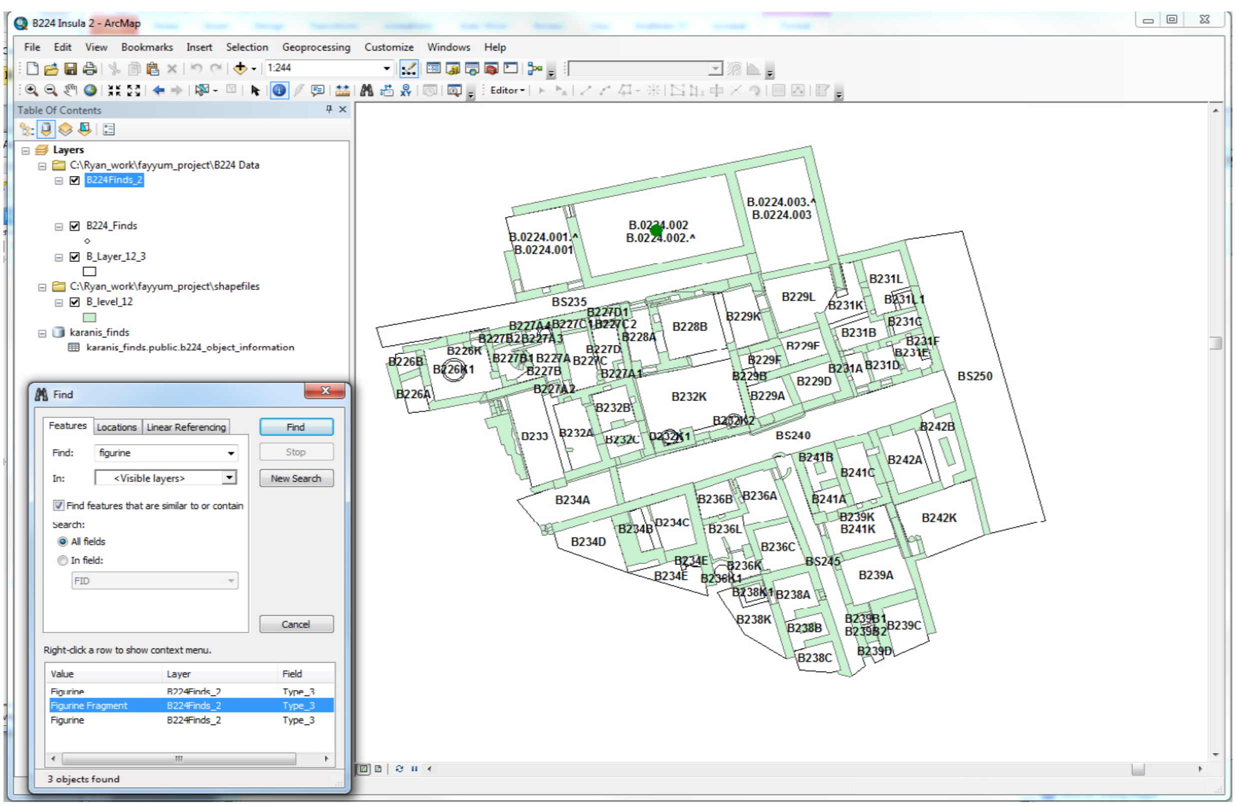


Figure 20: Screenshot of query in ArcGIS for the keyword “figurine.” Three results were retrieved, all in room B224B.

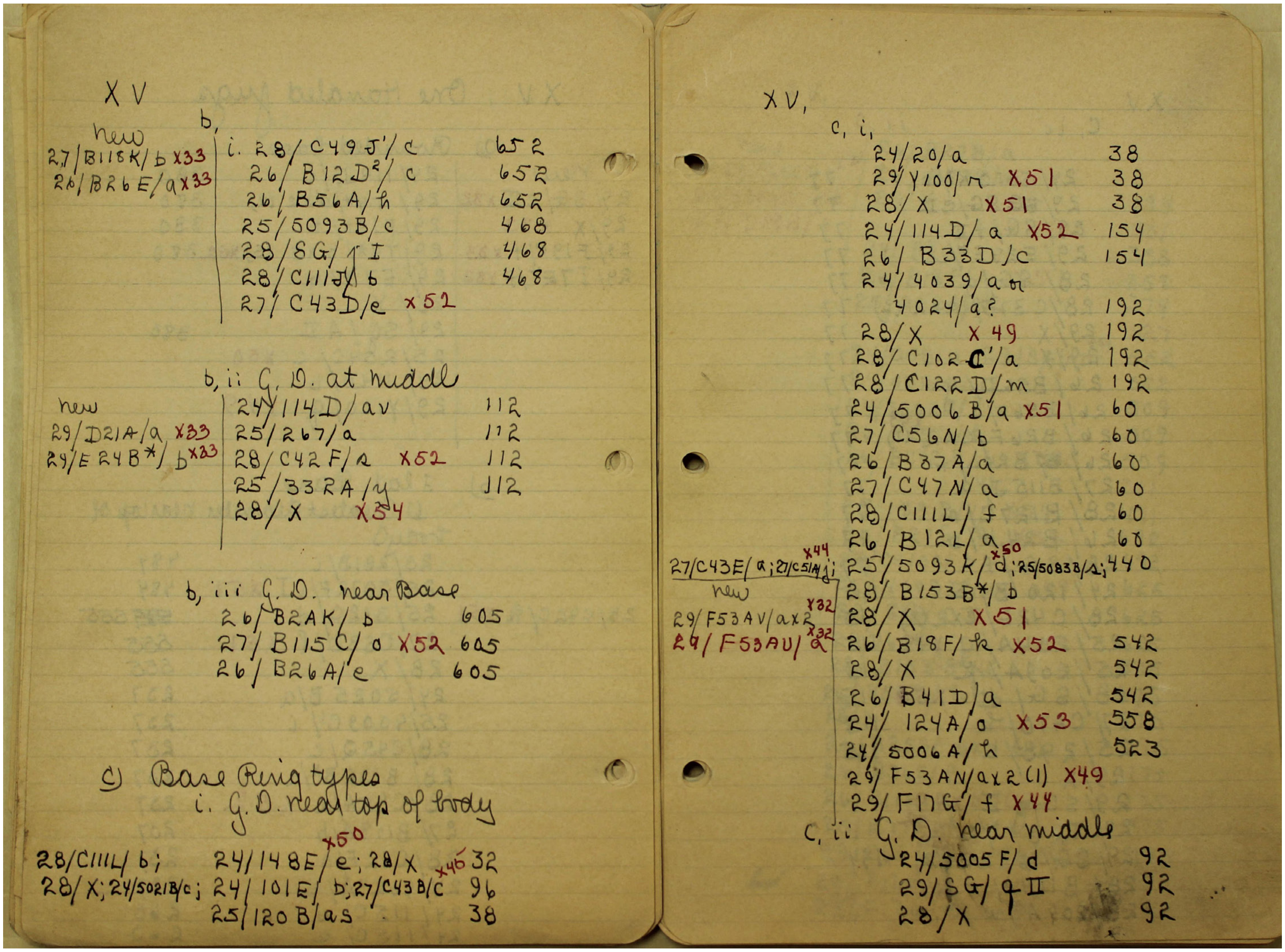


Figure 21: Page from the pottery “key” showing types, e.g. Type XV c is a vessel with a ring base; this type can be subdivided into types 32, 96, 38, etc.