The Report Committee for Kathleen Forrest

Certifies that this is the approved version of the following Report:

**Approaches to Digital Preservation: Re-examining the recovery of the *Chora of Metaponto, The Necropoleis***

**COMMITTEE:**

Dr. Patricia Galloway, Supervisor

Dr. Ciaran Trace

# **Approaches to Digital Preservation: Re-examining the recovery of the *Chora of Metaponto, The Necropoleis***

By Kathleen Forrest

A Report
Presented to the Faculty of the Graduate School
of the University of Texas at Austin

in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in Information Studies

The University of Texas at Austin
April, 2019

Acknowledgements:

I would like to thank the Institute of Classical Archeology for giving us the opportunity to help preserve this significant body of research and scholarship, especially Lauren Jackson, who worked closely with us throughout the original archival project and later continued to work with me as I completed this report. I would like to warmly thank my committee members and mentors, Dr. Galloway and Dr. Trace, for their guidance and support. And lastly, thank you to the group members that I had the pleasure to work with for the original project: Josh Conrad, Birch Griesse, and Nicole Lumpkins.

# **Approaches to Digital Preservation: Re-examining the recovery of the *Chora of Metaponto, The Necropoleis***


By

Kathleen Forrest, MSIS

The University of Texas at Austin, 2019

SUPERVISOR: Patricia Galloway



ABSTRACT: In 1998, the Institute of Classical Archeology published the *Chora of Metaponto: The Necropoleis*, a two volume book that represented over twenty years of archeological research and scholarship. This paper describes how graduate students in Dr. Galloway's Digital Archiving class, working together with the ICA, utilized emulation as a tool for recovering data in obsolete legacy formats for long term preservation and increased digital access. This paper analyzes their process, surfacing two different and sometimes competing approaches to digital archiving–preserving working files in their original context as evidentiary information, versus preserving the intellectual content of the material in new formats. Awareness of the motivations behind these two approaches and their different methods will help archivists more clearly identify their goals and allocate resources in other projects of this nature.

# Table of Contents

# INTRODUCTION:

The Institute of Classical Archaeology was founded at the University of Texas at Austin by Dr. Joseph Coleman Carter in 1987.  Since then, Dr. Carter has overseen the ICA's research initiatives surrounding the excavation of the territory of *Metaponto*, an ancient Greek colony on the southern coast of Italy. The excavations of a necropolis at Metaponto, known as the Pantanello Necropoleis, took place between 1982 and 1989; in 1998, the Institute published the *Chora of Metaponto: The Necropoleis* in two volumes, the first publication in a series that would describe their work at Metaponto.[1]

The Pantanello Necropoleis was significant as one of the largest necropoleis in the Greek colonies, which "provided valuable evidence that bridges numerous topics for scholars of Classical archaeology: Greek colonies in southern Italy, Greek and indigenous Italian burial practices, Greek and indigenous interactions, Greek mystery cults, and diet and health of ancient populations, among others." [2] With 890 richly illustrated pages, *The Necropoleis* was at the time of publication, "the fullest study of the physical state and health of a Greek population."[3] It won two awards: the James R. Wiseman Book Award of the Archaeological Institute of America (1999) and the Gold Medal of the Fondazione Raffaele Guariglia in Vieste, Italy (1998).

As the planned retirement of their director, Dr. Carter, approached, the Institute of Classical Archaeology focused on archiving their records. Their goal was to consolidate the work of a globally distributed team and preserve their research and make it available for future researchers. They planned to house their physical records at the Briscoe Center of American History, the University of Texas' official archival holdings. But the ICA staff was responsible for finding a way to preserve their digital materials themselves.

The ICA approached Dr. Patricia Galloway, professor at the University of Texas School of Information, to propose the archiving of their digital materials as a project for the graduate students in Galloway's Spring 2018 Digital Archiving course. Developed and taught by Galloway since 2001, the course provided students with an active learning environment, experience experimenting with digital archiving methods, and a chance to practice adapting their methods to different situations. Structured as a reflective-practice based learning environment, the goal of the course was to properly train archivists in working with digital material so that they were prepared for the ongoing challenge of rapid technological change and evolution throughout their professional careers. [4]

The ICA proposed a project that would offer the students an opportunity to learn both new technological skills and how to "recognize and solve novel problems."[5] The Institute wanted to preserve the digital files that were created as part of the process of publishing the Necropoleis—excavation data that was collected and recorded digitally on site in Metaponto, and the photographs, illustrations, and files that were used to create the final layout of the book before being sent to print. Additionally, the ICA wanted to produce a complete, digital copy of this book to be shared online with scholars of Greek archeology, and to share the data collected at Pantanello online through the Texas Data Repository. Central to this project, which dealt almost exclusively with legacy file formats from more than twenty years ago, was the use of migration and emulation for increased preservation and accessibility.

As part of the Spring 2018 class, students Josh Conrad, Nicole Lumpkins, Birch Griesse, and myself, partnering with the ICA, arranged files from the *Necropoleis* project, added metadata, and produced multiple working emulated environments that enabled future researchers to view the legacy files in their original context of creation. Finally, we migrated these files to digital archival formats. We documented this project in a report titled, "Greek to Me: Using Emulation to Digitally Re-publish the *Chora of Metaponto: The Necropoleis.*"[6]

This report is a reflection on the lessons learned and fundamental concepts we took from that project. Specifically, I believe that the goals of the project—preserving the working files and preserving just the book and the database files—represented two different approaches to digital archiving—one that treated the materials as evidentiary information, and another that focused on the intellectual content.

Understanding the differences in methodology required by each approach can help archivists clearly identify goals and better allocate resources in other, similar projects.

*Figure 1 Example of pottery found at the Pantanello Necropoleis*

# SECTION I: THE PROJECT

*The Necropoleis* project was a massive collaborative endeavor, requiring the hard work and expertise of many individuals over the course of twenty years. In Volume I of the book, Dr. Carter warmly acknowledges seventy-eight team members and their contributions. The long list, which includes their roles and areas of specialization, gives us some insight into the varied and complex labor involved. Carter worked with a team of excavators, statisticians, biologists, chemists, anthropologists, architects, draftsmen, conservators, and archeologists specializing in ancient materials like ceramics, metals, and stone. They utilized complex technologies in the project with the help of expert photographers, x-ray and remote sensing technicians, and database managers. And after the fieldwork and scholarship was done, the book was assembled with the help of editors, graphic artists and designers.[7]

This work generated a large amount of digital records. Before excavation even began, project logistics such as funding and travel arrangements were negotiated in documents written in word processing programs. During excavation, each step was documented thoroughly with photography, data tables, and journals. Those photographs were digitized and turned into illustrations for the book with vector-based computer graphics programs. Manuscripts were written, edited, and shared between team members digitally. And finally, the text, illustrations and data were combined in beautiful layouts in desktop publishing programs before being sent to print.

The digital material created during the *Necropoleis* project was kept by the ICA after the project ended, but, like many institutional files, they had not been formally archived. Since the book was published in 1998, twenty years had passed before the ICA and our student group approached the materials to create a long term preservation plan.

Our first step was to make sense of what we had by taking inventory, arranging the files as close to the original order as possible, and creating metadata to describe the circumstances of their creation and how they were used.

The files were kept on computers (run on servers maintained by University of Texas) and on legacy media storage devices—ten Iomega ZIP drives, twelve 3 ½ inch floppy disks, and twenty optical compact discs (commonly known as CDs). Between these different devices, they had approximately 2GB of material that staff had identified for archiving. The file names and dates, however, showed that many of the different storage devices held duplicates of the same materials. The collaborative nature of the *Necroploleis* project meant that different actors within the ICA had created and retained different versions of the same documents during their use. Duplicates and copies had been made after they were out of use as well. As one of many examples, the ICA created a copy of all the *Necropoleis* materials when their office transitioned to new servers in 2012 and they had to migrate everything from the old server.

In our attempt to make sense of these different versions and find the original files, we started to approach the concept of the document in a digital environment as fundamentally different from paper files. The digital document extends beyond how it is rendered on the screen and includes the data encoded on the storage device as bits.[8] When a file is opened, the computer reads the bits and transforms them into something the user can understand and interact with. The document as it appears on the screen is, in effect, a new copy of the document that the computer creates each time the user accesses it. Opening the document and re-saving it, moving it to another folder, or re-naming it may not change the text, but it does alter the bits and therefore changes the document by creating a new version and (in some cases) destroying the old one. Because of the difficulty in maintaining the document at the bit level, some archivists say that in the digital world, there is no such thing as an "original."[9] Others argue that all versions of the material are "original" as long as they were created by the same author.[10] With this logic, an original document can have multiple instances.

In fact, it became clear that none of the files we were working with were the "originals." They were all derivatives of the originals which had existed on the terminals and servers that the team used in 1997 to create the book, and on the computers in Italy where the archeologists first recorded their data.

Without these, we decided to treat the backups as originals, and focused instead on using other distinguishing characteristics to differentiate between the duplicate versions.

**File Types and Structure**

We mapped the organization of files on the different storage devices to compare their contents and identify the most complete and accurate records. For instance, the CDs and the ZIP drives contained the same types of files (.pmg, .tiff, and .esp), while the floppy disks contained .dbf .mss and .doc files. We also utilized existing metadata—some of the filenames contained hints such as v3 or v4, which we took to indicate version number, and "64 (20) metalsTOC" suggested "table of contents," indicating files related to the book. The files also came with date and time stamps that were created by the authoring program. The floppy disks contained files created between 1989 and 1996; the CDs the files were created between 1992 and 1999; the files on the ZIP drives were also created in the '90s. While date and time stamps can be unreliable, these dates seemed to correlate with the ICA's own narrative about the project.

The file types themselves proved challenging to work with. Unlike physical pages in a book or folder, digital materials can only be viewed with the aid of a digital device and specific software. Without the right software the contents of a file remained hidden. And since existing software changes and is replaced over the years as developers periodically distribute new versions, contemporary computer programs may not support the same file formats that they did twenty years ago. Many of the older file types were not supported on our modern machines, which meant they couldn't be opened.

In 2018, our group members were working with a combination of Windows and Apple computers, running contemporary operating systems such as Microsoft Windows 10 and Mac OS 10.12 Sierra. The systems were equipped with the Microsoft Office Productivity Suite, the Adobe Creative Cloud Suite, text editors, and image, audio and media software. With these tools, we were able to open the more ubiquitous file types such as .doc and .txt files, but these suffered from loss of information from encoding differences. Other files could not be opened at all, and were often missing their file extensions so the system couldn't determine what type of file it was.

Since we were unable to look at the majority of the files, or determine what type of information they contained, we relied heavily on the metadata, the file structure, staff testimony and the printed version of the *Necropoleis* for clues. We started to piece together a narrative that described where the files came from and how they were used. Still, it was frustrating how little information we had to work with since we had the files in our hands but couldn't access their contents.

One of the Zip drives, for example, contained six folders and nine files. Our computers didn't recognize the file types of those nine files, nor did they have any file extensions we could look up. Four of the six folders represented data that computers automatically create as by-products of their normal functionality, and were unrelated to the work of the ICA, such as the "desktop" and "trash" folders. Only two folders had content related to the *Necropoleis,* these were titled "60-61 (19) Terracottas fld" and "62-64 (20)metal fld." Inside these were subfolders with ten and forty .tif files, respectively. These .tifs were either photos or computer-generated schematic drawings of iconography found on vases that were found in the tombs as grave goods. We could open the drawings and images in Adobe Photoshop.

Next to these image folders were two files with no file extensions, which we later identified as Aldus PageMaker files. These would not open in any of the programs on our modern computers. The PageMaker files, as we later discovered, contained the final layouts of the book where the images and text were brought together and formatted for publication.

The other Zip drives were largely the same—the majority of the files wouldn't open, many of the ones that did open suffered some information loss that was evident in the appearance of uncommon characters and formatting, and it was difficult to determine how these files related to each other or how they were used. We had to rely heavily on metadata such as file names and structure to guess what the files contained. We could tell, for example, that groups of illustrations and the unknown file types were related to each other based on the folder they were in, and most likely represented chapters or other distinct sections of the book. The file names sometimes correlated with the organization of the book itself, although we couldn't make any assumptions and had to check our guesses against the printed copy of *The Necropoleis.* The file titled "60-61 (19) Terracottas fld" refered to chapter 19, Terracotta Figurines, by Mary Link Malone.

Staff testimony also helped us create descriptive metadata. Research Project Manager Lauren Jackson, our liaison at the ICA, thoroughly answered our many questions and sometimes even contacted individuals who were no longer at the ICA to gather more information for us. Lastly, we relied heavily on the book itself for insight on how these files related to each other. We used this information to put together a description of these files.

**Security**

During this process, we were extremely careful not to alter or overwrite any data we were working with. Because of the way computers read and write information and the way that files are formatted on different machines, an action as simple as copying a file to another folder can change its metadata and alter the basic encoding of the file itself. In the interest of preserving as much metadata as possible, we used established, digital forensics-type techniques during the transfer and handling of the materials to ensure their safety. We used the open source, Ubuntu-based software environment BitCurator to implement write-blocking protection when we opened the Zip drives that would prevent us from accidentally making changes to the material with a stray click of the mouse or keystroke.

Whenever we transferred files from their legacy storage devices to our computers, we used software that would replicate the files at the bit stream level, which ensured that the material will be "copied in such a way that all available data are preserved intact."[11] The bit stream image is an exact replica of the bits a computer uses to encode information onto a memory device. Typically, using the computer's graphical interface to "copy" and "paste" a file will re-write the bits in a new order based on the file structure and operating system of the file's new location. We prevented this by using Guymager, the bit stream imaging software available with BitCurator.[12]

**Narrative**

At the end of our investigation, we had identified two categories of materials. The first contained all the documentation that was created during the Metaponto excavation, created between the years 1980 and 1990, and consisted mostly of data recorded about the tombs and their contents, but also included things such as documented methodology, correspondence, and program files. The most complete set of these records were found on the floppy disks. The .dbf files and the text files opened on our computers in contemporary spreadsheet programs such as Open Office Calc 4.1.5, and in Open Office Writer 4.1.5. But, they showed signs of loss (described in the next section). Other files, such as the .mss files, could not be opened.

In the second category were all the files directly related to the publication—illustrations, images text, and the PageMaker files, as they had been assembled for the graphic designer. [See Figure 2] Through the metadata, file structures, and staff testimony, we also determined that the files on the Zip drives were the most accurate and complete versions of the materials related to the book, and that the other versions were derivatives of the files on the Zips.

*Figure 2 Example of images used in the* Necropoleis *publication, saved on Iomega ZIP drives*

**Next Steps: Opening the Files**

At this point, we understood what the files contained but were still unable to access that information. As archival documents, they didn't reveal much about the process behind the ICA's significant work of scholarship. Our task, then, was to find a way to open the files.

We formalized our goals. The first was to find a way for future researchers to open the files that were associated with the book, to provide evidence of the process of publication. Secondly, we wanted to open the PageMaker files ourselves and export those layout pages into .pdf format as a full-resolution, complete digital version of *The Necropoleis*. This digital version of the book would be shared with researchers online, supporting the mission of the ICA to preserve and disseminate their research for future scholars. This had been impossible up until now, not only because the files were inaccessible, but because digital reproduction and dissemination would have violated the terms of the copyright holder, UT Press.

But when the publisher's rights to *The Necropoleis* expired in 2018, they were transferred back to Carter, who wanted to utilize his rights as copyright holder to share this book digitally and for free.[13]

Our goals for the excavation files were similar—provide a way for researchers to view the files and prepare the data to be shared online through the Texas Data Repository. Doing so meant the data would need to be exported into new file formats. Some of the data that was collected during excavation had previously been migrated and re-interpreted in data tables published in the book, but this would be an opportunity to view and share the "raw" data and anything that was left out of the book.

The motivations behind these goals was to preserve as much information from the ICA's digital materials as we could for the future. Our project was typical of many other digital archiving projects today, and posed the same challenges others face—are materials created four decades ago lost to us forever? How do we preserve and provide access to those materials, and how do we implement what we learn into our habits for handling digital files today?

**Emulation in Theory**

We approached these questions through existing research and literature on digital media as archival objects and investigates how they function within the computing infrastructure (the software, operating system, and hardware). These scholars ask archivists to reject the common idea of digital information as immaterial, and re-consider it within its physical properties and mechanical processes. By focusing on the materiality of digital information we can learn how to work around its constraints and effectively handle these files.[14]

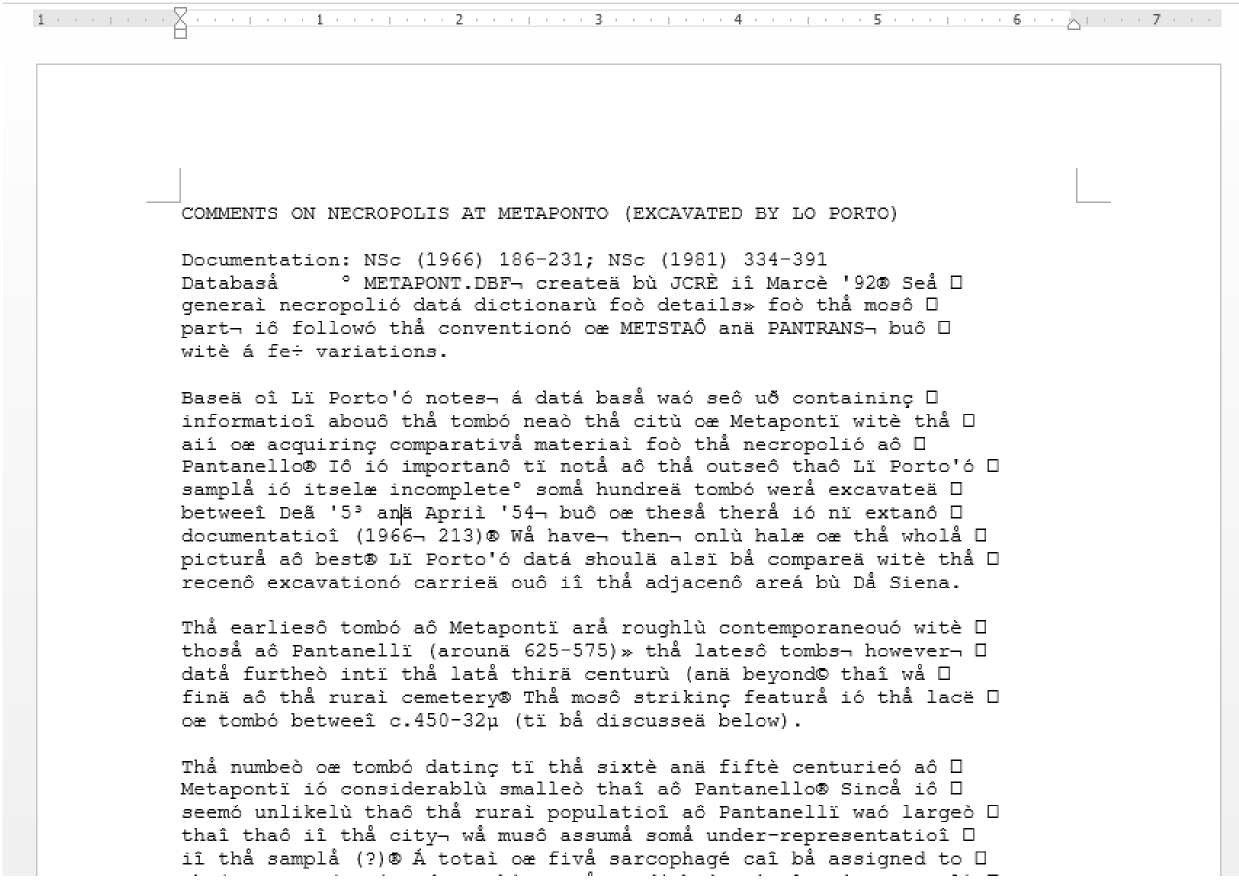
Computers read and write information to memory devices as bits using a series of on/off signals—represented as 0's and 1's, and known as binary code.[15] To turn those bits into something we can understand, those 1's and 0's are run through a set of specific instructions for translation. Those instructions come in packages known as computer programs. Each program has unique instructions for translating bits. A word processing program knows exactly how to take the bits and render them as text, while a media player will read bits and output music or video. Our modern computer programs lacked the right instructions to translate the bits, which were encoded and written to their memory devices by different programs a long time ago. The solution, however, wasn't as simple as installing the old programs on our modern machines. A program is created to be run on specific operating system; a program written for Windows would not run on a Macintosh operating system, and Aldus PageMaker 5 from 1997 would not run on any of the systems installed on our modern computers. To open these files, we had to use the right program in combination with the right operating system.

An operating system, however, is specific to certain hardware. It manages communication between the computer's physical parts—such as hard-drive, processor, and screen—and the software. For example, when we use the keyboard to input data, the operating system takes those keystrokes and delivers them as binary data to the programs that translate them to letters on the screen. The operating system also coordinates the computer's memory devices and processes, deciding where exactly on the hard drive new bits should go, and what tasks are completed in what order. The operating system, which is essentially a large, powerful piece of software, gives other software access to the hardware.

All the parts of the computing infrastructure contribute to the way data is read and interpreted. Without the right instructions, the computer cannot translate the data or will translate it incorrectly.

Many of us have experienced this: try to open an old file in a newer version of the same software, and you will be promoted to "update the file to the most recent file format." The greater the difference in version number, the more trouble the computer will have translating the data. By opening a file from the excavation titled METAPONT.MSS in Microsoft Word 2013, we can see that the computer has mistranslated the data and displayed symbols at the end of each line of text and replaced basic English letters with special characters such as ä, ė, and ŏ. [See Figure 3] As a result, a firsthand account of the excavation from an expert archeologist appears to be written in a whimsically cryptic, unknown language. With a little effort and flexibility, the document could be read and understood by a typical Anglophile.

But other documents can suffer much greater loss—the content may not be as easily deciphered, or it may be irrevocably lost.



COMMENTS ON NECROPOLIS AT METAPONTO (EXCAVATED BY LO PORTO)

Documentation: NSc (1966) 186-231; NSc (1981) 334-391
Databaså ° METAPONT.DBF¬ createä bù JCRÈ iî Marcè '92® Seå □
generaì necropolió datá dictionarù foò details» foò thå mosô □
part¬ iô followó thå conventionó oæ METSTAÔ anä PANTRANS¬ buô □
witè á fe÷ variations.

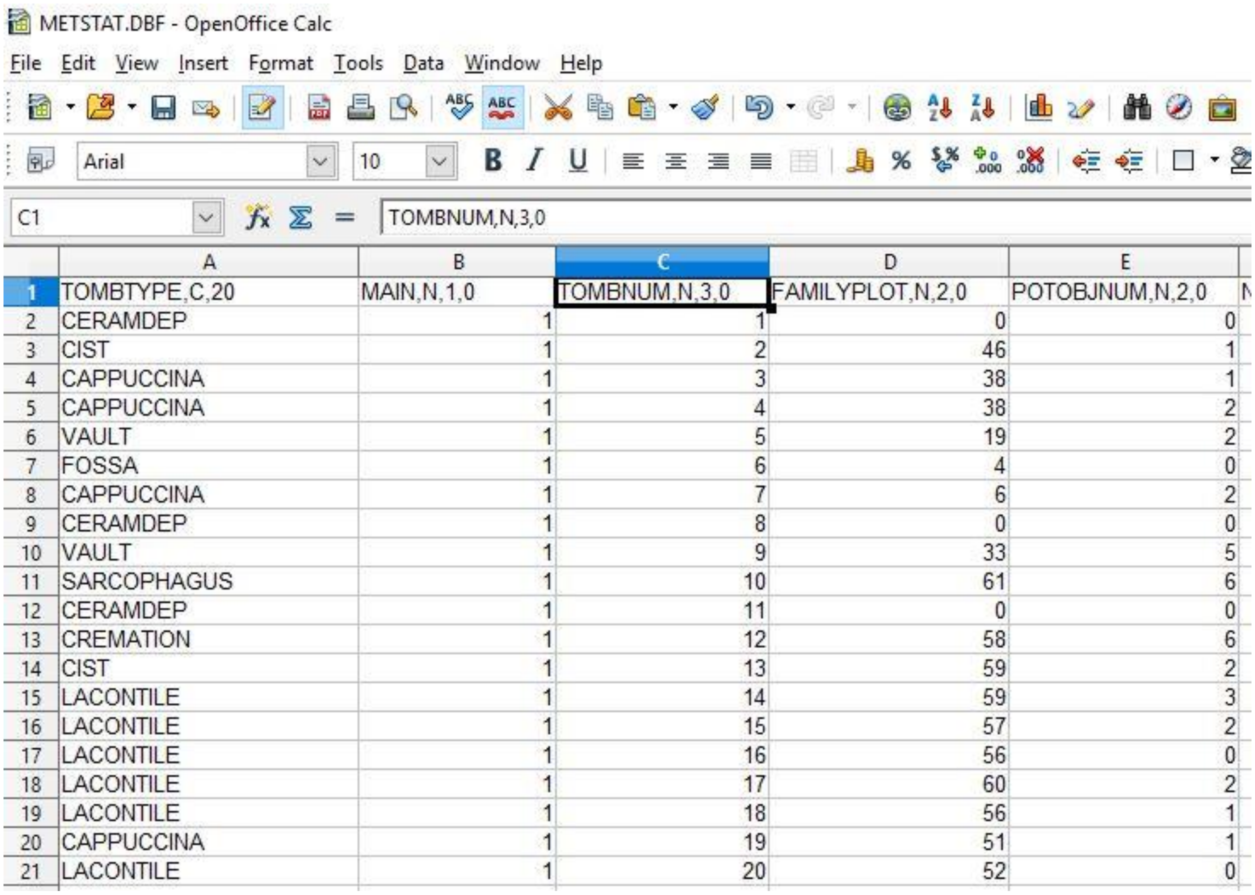
Baseä oî Lï Porto'ó notes¬ á datá baså waó seô uð containinç □
informatioî abouô thå tombó neaò thå citù oæ Metapontï witè thå □
aiì oæ acquirinç comparativå materiaì foò thå necropolió aô □
Pantanello® Iô ió importanô tï notå aô thå outseô thaô Lï Porto'ó □
samplå ió itselæ incomplete° somå hundreä tombó werå excavateä □
betweeî Deã '5³ anä Aprii '54¬ buô oæ theså therå ió nï extanô □
documentatioî (1966¬ 213)® Wå have¬ then¬ onlù halæ oæ thå wholå □
picturå aô best® Lï Porto'ó datá shoulä alsï bå compareä witè thå □
recenô excavationó carrieä ouô iî thå adjacenô areá bù Då Siena.

Thå earliesô tombó aô Metapontï arå roughlù contemporaneouó witè □
thoså aô Pantanellï (arounä 625-575)» thå latesô tombs¬ however¬ □
datå furtheò intï thå latå thirä centurù (anä beyond© thaî wå □
finä aô thå ruraì cemetery® Thå mosô strikinç featurå ió thå lacë □
oæ tombó betweeî c.450-32µ (tï bå discusseä below).

Thå numbeò oæ tombó datinç tï thå sixtè anä fiftè centurieó aô □
Metapontï ió considerablù smalleò thaî aô Pantanello® Sincå iô □
seemó unlikelù thaô thå ruraì populatioî aô Pantanellï waó largeò □
thaî thaô iî thå city¬ wå musô assumå somå under-representatioî □
iî thå samplå (?)® Á totaì oæ fivå sarcophagé caî bå assigned to □

*Figure 3 "Comments on Necropoleis at Metaponto" in Microsoft Word 13 on Windows 10*

Sometimes, the errors are not as obvious. Some things might appear to make sense, but small errors can change the meaning entirely. Numbers in a spreadsheet are often formatted differently based on the type of information they represent: dates, decimals, currency, negative or positive integers. The formatting gets encoded with the number itself and might get misinterpreted in a different program. The result could be a misplaced decimal, too many significant digits, or metadata inserted as content. The Metaponto excavation data for instance, was recorded and formatted in a program called dBase, and saved as .dbf files. When we opened those files with Open Office Calc 4.5.1, numbers and letters had been added to the column: “TOMBNUM, N, 3, 0.” [See Figure 4] We couldn’t have known that those were not present in the original version until later, when we were finally able to open the file in the original environment, where the column was simply listed “TOMBNUM.”




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | TOMBTYPE,C,20 | MAIN,N,1,0 | TOMBNUM,N,3,0 | FAMILYPLOT,N,2,0 | POTOBJNUM,N,2,0 |
| 2 | CERAMDEP | 1 | 1 | 0 | 0 |
| 3 | CIST | 1 | 2 | 46 | 1 |
| 4 | CAPPUCCINA | 1 | 3 | 38 | 1 |
| 5 | CAPPUCCINA | 1 | 4 | 38 | 2 |
| 6 | VAULT | 1 | 5 | 19 | 2 |
| 7 | FOSSA | 1 | 6 | 4 | 0 |
| 8 | CAPPUCCINA | 1 | 7 | 6 | 2 |
| 9 | CERAMDEP | 1 | 8 | 0 | 0 |
| 10 | VAULT | 1 | 9 | 33 | 5 |
| 11 | SARCOPHAGUS | 1 | 10 | 61 | 6 |
| 12 | CERAMDEP | 1 | 11 | 0 | 0 |
| 13 | CREMATION | 1 | 12 | 58 | 6 |
| 14 | CIST | 1 | 13 | 59 | 2 |
| 15 | LACONTILE | 1 | 14 | 59 | 3 |
| 16 | LACONTILE | 1 | 15 | 57 | 2 |
| 17 | LACONTILE | 1 | 16 | 56 | 0 |
| 18 | LACONTILE | 1 | 17 | 60 | 2 |
| 19 | LACONTILE | 1 | 18 | 56 | 1 |
| 20 | CAPPUCCINA | 1 | 19 | 51 | 1 |
| 21 | LACONTILE | 1 | 20 | 52 | 0 |

*Figure 4 “METSTAT Database” opened in OpenOffice Calc on Windows 10*

The computing system—hardware, software and operating system—forms the necessary infrastructure for recovering data saved in digital media. According to existing research and literature, the best way to recover that data with no loss of information was to reproduce the original computer system, or the context of creation where the files were first created. This approach was best explained by Jeff Rothenberg in his essay “Ensuring the Longevity of Electronic Documents.”[16]

His description of the challenges archivists face when working with digital material is spot-on, including the early steps where we had to depend on metadata to understand the contents of the Zip drives and other various storage devices when we couldn't open them. Rothenberg clearly describes how computers process binary data and the many ways those data can be mistranslated. He argues against migration,[17] and demonstrates that the only way to preserve the data perfectly is to also preserve the software, operating system, and hardware that they were originally created with.

Yet he also acknowledges the futility of preserving hardware for these purposes. An institution attempting to do so would have to keep a working specimen of all the standard models ever produced, and all the necessary accessories—mouse, cables, disk drives, etc. Hardware is also at risk of deterioration and malfunction. As the models go out of date, their parts are no longer commercially produced and are difficult to replace. Instead of preserving the machines themselves, archivists can mimic their functions and behavior in a virtual machine—a program designed for building emulations of different systems. The virtual machine partitions and runs a different operating system and software on an existing computing system. It also adapts the input/output from the host machine to the guest operating system. Thus, the virtual machine could enable the emulation of the original operating system and software used by Carter's team to research and publish *The Necropolies*. Based on the descriptions of emulators by Rothenberg, we decided to emulate two environments to recover the ICA's files—one to replicate the systems used to record data from the excavation, and another that replicates the system the ICA later used to design the book.

Carter and his team thoroughly documented their process, even including a description of the tools used to design and layout the book which they published in Appendix 9A.3. The software used for the layouts was Adobe PageMaker; the photography was digitized, then edited in Adobe Photoshop; the graphs and digital drawings were created in Adobe Illustrator and Denebra Canvas; the manuscript was mostly typed and edited in Microsoft Word.[18]

The description is so thorough that it included details about the entire process from scanning the images (at 600 dpi) to the method of transmission of the digital documents to the printer (optical disks). This information was incredibly helpful for building our emulation; it almost completely described all the essential elements of the computing system, except for the specific version of the software used and the operating system, which we could approximate based on the dates and the computer models (Apple Macintosh Centris and Power Macintosh 7200).

In addition, Chapter 9, appendices 9A.1 and 9A.2 were devoted to explaining the databases used to record excavation data and the statistical analyses they performed with those data sets. The data was recorded in two databases created in the program dBase by Ashton-Tate, and exported to be manipulated in Microsoft Excel and a mapping program created by Golden Software called Surfer. This chapter also contained a lengthy explanation of the record types, the fields, and a complete data dictionary. But they did not list the version of dBase they used, nor the operating system and computer model. These we guessed at in order to build our emulation.

**Emulation in Practice**

Building the emulations became the focus of our project, as it was the best way for us to both provide a way to view the files and migrate the contents to new formats to be shared. It also took significant time and resources. Our group members were comfortable using a variety of contemporary technological applications and were invested in learning more about legacy systems. But researching and experimenting with different operating systems and software in virtual machines turned out to be a process with quite a learning curve. Many of the system functions on the modern computers we were used to were automated. The earlier systems required greater involvement from the user, but also offered greater control.

We quickly realized building an emulation was much more complex than simply finding the right program for each file type and installing it on the right operating system. Systems require a variety of support software, such as device drivers[19] and system updates that fix errors in the original version of the operating system. These additional programs are specific to the program or operating system they are intended to support, and can vary by version number and the geographic location where they were released. Our task was to find the right version of operating system and find the matching versions of device drivers and programs used to edit the data, which largely became a process of trial and error.

All the software, drivers, and operating systems were proprietary and no longer produced or sold by the developers. They were saved and made available online for non-commercial use by volunteer contributors that loosely form an informal online community. Thanks largely to their efforts and expertise we were able to source most of the components of our emulated environments ad-hoc, and installed them with the help of instructions provided by the contributors. However, documentation was sometimes sparse and often assumed a higher degree of familiarity with computing. And we often installed multiple versions of the same software before finding one that was compatible. Sometimes the right software wasn't available, so we had to rely on creative work-arounds and alternative solutions.

Most of the software we downloaded from WinWorld, a self-described, "online museum dedicated to the preservation and sharing of vintage, abandoned, and pre-release software." WinWorld is an online platform for the free sharing of legacy software and operating systems that are no longer available for sale from the original vendor. To install and run these programs, we used online serial number generators to 'authenticate' them. We recognized that this might be cause for concern from a legal standpoint, since the software was proprietary. But we determined that using these software in an emulation to recover lost information was justifiably fair use under copyright doctrine, since our actions were driven by research and learning purposes. The recently released *Code of Best Practices in Fair Use for Software Preservation,* recognizes that using these software in the interest of preserving digital media where "legacy software cannot be obtained in the commercial marketplace" and "commercially-available rendering tools may not faithfully represent digital objects originally created in now-obsolete formats," is both necessary and permissible under fair use.[20]

We experienced a lot of system failures along the way. For example, the virtual machine software SheepShaver is an open-source Macintosh emulator that was not as well supported as the alternative emulation software Virtual Box. We used SheepShaver to emulate the Macintosh computers used to design and layout the book. But SheepShaver was not designed to support the operating system we needed, nor to run PageMaker. The virtual machine struggled to run those programs and often crashed in the middle of a process.

Despite these challenges we were able to build two successful emulations of environments that could open the *Necropoleis* files, described in the following sections.

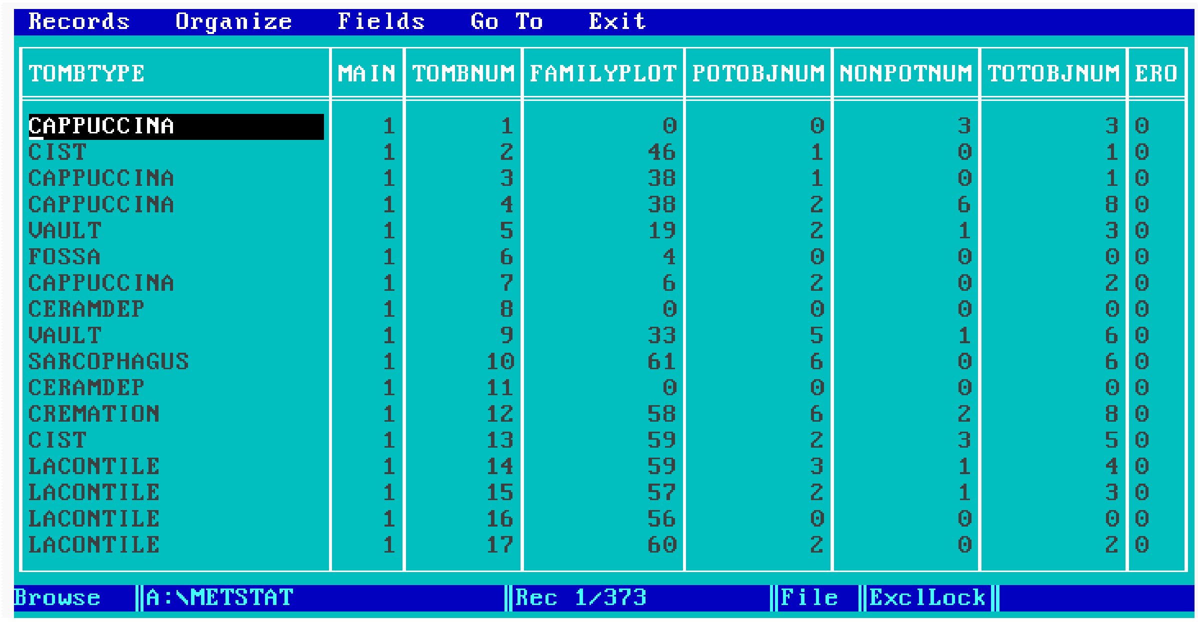
**DBase Emulation**

The first was an emulation of the environment the archeologists used to record excavation data. We knew they were using dBase because of the notes left by Carter and his team in the book, and we guessed that the operating system was Windows 3.1 based on the dates they were reportedly on-site at Pantanello.

Windows 3.1 ran on top of MS DOS 6.6, which we installed in a virtual machine in Oracle VirtualBox. We then installed an ET 4000 SVGA video driver, a SoundBlaster 16 sound driver, and a copy of Microsoft Office 4.3, which included Word, Excel, PowerPoint, and Access. Later on, we installed imaging software and file compression software to assist with transferring the exported data back to the host machine. We tried installing five different versions of dBase before we found one that worked, dBase IV, which was released in 1988.

To transfer the files to the emulation, we used WinImage to write the .dbf files as a disk image (.vfd) on the host machine. Then, in the virtual machine, we mounted the disk images in the virtualized "floppy drive" and were able to access the files in the Windows 3.1 environment. Later, we installed an early version of WinZip in the emulation and used that to compress the exported data in .dif format so they could then be written on the 1.44M virtual "floppy drive" and transferred back to the host machine as a disk image.[21]

Using this emulation, we were able to open the .dbf files in dBase. The floppy disks had twenty-nine different database files representing datasets collected at the site. One of the largest, METSTAT, which contained 373 records of exhumed burials at Pantanello, is depicted below [See Figure 5].


Records Organize Fields Go To Exit

| TOMBTYPE | MAIN | TOMBNUM | FAMILYPLOT | POTOBJNUM | NONPOTNUM | TOTOBJNUM | ERO |
|---|---|---|---|---|---|---|---|
| CAPPUCCINA | 1 | 1 | 0 | 0 | 3 | 3 | 0 |
| CIST | 1 | 2 | 46 | 1 | 0 | 1 | 0 |
| CAPPUCCINA | 1 | 3 | 38 | 1 | 0 | 1 | 0 |
| CAPPUCCINA | 1 | 4 | 38 | 2 | 6 | 8 | 0 |
| VAULT | 1 | 5 | 19 | 2 | 1 | 3 | 0 |
| FOSSA | 1 | 6 | 4 | 0 | 0 | 0 | 0 |
| CAPPUCCINA | 1 | 7 | 6 | 2 | 0 | 2 | 0 |
| CERAMDEP | 1 | 8 | 0 | 0 | 0 | 0 | 0 |
| VAULT | 1 | 9 | 33 | 5 | 1 | 6 | 0 |
| SARCOPHAGUS | 1 | 10 | 61 | 6 | 0 | 6 | 0 |
| CERAMDEP | 1 | 11 | 0 | 0 | 0 | 0 | 0 |
| CREMATION | 1 | 12 | 58 | 6 | 2 | 8 | 0 |
| CIST | 1 | 13 | 59 | 2 | 3 | 5 | 0 |
| LACONTILE | 1 | 14 | 59 | 3 | 1 | 4 | 0 |
| LACONTILE | 1 | 15 | 57 | 2 | 1 | 3 | 0 |
| LACONTILE | 1 | 16 | 56 | 0 | 0 | 0 | 0 |
| LACONTILE | 1 | 17 | 60 | 2 | 0 | 2 | 0 |

Browse ‖A:\METSTAT ‖Rec 1/373 ‖File ‖ExclLock‖

*Figure 5 METSTAT database opened in dBase IV on Windows 3.1*

It was interesting to navigate the data records in dBase through the emulation. As Jon Morter, field director, database manager, and co-editor of the *Necroploeis*, writes in Chapter 9, "The platform dictated the nature of the original data entry, and the structure of the files influenced the way tests and statistics were handled."[22] He goes on to explain how the data and analysis used in their scholarship was based on data structures and formats dictated by the software and computing systems that were set up in the 1980's. Since then, database programs have become more sophisticated and better suited for archeological record keeping. Relational databases, for example, can link multiple datasets, allowing researchers to represent more complex relationships between records. Considering the limitations of their tools when working at Pantanello we can better appreciate the significance of their research.

To provide future researchers the means to study these datasets in their context of creation, we exported the entire emulation package, including all the software we had installed, the files from the floppy disks, and our documentation in an .ova file (open virtualization format archive). After installing Virtual Box and then importing the .ova file, users will be able to open the entire emulated environment that we built. We also exported the data to .dif files for easier access, although we cautioned the ICA that data could have been altered in the migration process and should be reviewed for errors before it was shared on the Texas Data Repository.

We experienced a few challenges that we couldn't resolve without more time and resources. The floppy disks also contained 300 .mss files, a few .doc files, some program files, and some files with unfamiliar file extensions. The .mss files were an unfamiliar file type. More research is needed to determine what program they were created in. We opened them in the emulation using Microsoft Word 4.3, but there was a significant amount of loss of information there. Instead, we opened them on a modern computer running Windows 10 and Microsoft Word 13, using Western European (IA5) encoding. We were able to see the content—Microsoft Word retained the characters and some of the formatting—and to prevent future loss we exported the files to .txt [See Figure 6] Text files are preferred archival formats because they are at lower risk of becoming obsolete, making them the most likely format that would remain accessible to future researchers if we were ultimately unable to open them in their context of creation.

The .mss files contained detailed descriptions of each excavation—each file corresponded with a different tomb number and includes their methods, what they found, conclusions, and notes for further work. This type of material was a good example of information the ICA used in the making of *The Necropoleis*, but did not necessarily publish with the book. Future researchers can return to these files and learn more about the conditions at the site, the archeologists' reactions to what they found, and other first-hand information. Efforts should be made to preserve this information as a record of the work at

Pantanello, and further effort should be made to discover what program was used to create them, to see if that program had any influenced on the way data was collected and analyzed.

CTLG215.txt - Notepad

File Edit Format View Help

```
Tomb 215
Sarcophagus
Nucleus: 10
Number of skels: 1
Orientation: 35
Wide Date Range: 0-0
Narrow Date Range:0-0
Basis for date:
Sex: Male
Age: 30-40

Tomb details:
Material :
L: 2.05
W: 0.755
H: 0.521 ??
Wall Thickness:


Vase summary: None

Other objects: None

Comments:

This was a sarcophagus with a carparo lid consisting of two
slabs, the S. slab measuring 0.93 x 0.8m (approx.), the N. one,
1.26 x 0.8m (actual widths are not given). The S. slab was
thicker than the N. slab (actual thicknesses are not given).

The skeleton was partially covered by soil-fill and badly
preserved because of the moist conditions. The remaining
fragments were widely scattered, probably the result of flooding,
and the skull was located at the S. end.
```

*Figure 6 .mss file converted to .txt on Windows 10*

**PageMaker Emulation:**

The second environment was an emulation of the computing system used to create the book which we used to open the files from the Zip Drives that we couldn’t access, and to migrate the PageMaker files to a .pdf version to be shared online.

We knew from the book appendices that the design team had used Mac computers and a combination of Microsoft Word, Adobe Photoshop, Illustrator, and PageMaker. Based on dates, we could guess that the operating system was a MAC 7, and the PageMaker version was 5.

We used SheepShaver to run our virtual machine, because VirtualBox did not support Mac operating systems. After some difficulty with SheepShaver, we installed Mac OS 7.5.3, which was released in 1994, and PageMaker 6.5, because PageMaker 5 was unavailable. We were able to open the PageMaker files, but because SheepShaver was not designed to support PageMaker, the emulation was unstable and frequently crashed. Also, PageMaker 6.5 in this system would not export to .pdf format without a specific Adobe driver that we were unable to find online.

In the interest of migrating the data to a .pdf, we built a new emulation with a Windows XP operating system in VirtualBox, where we discovered a work-around that would let us open the files in a different operating system. Using the software's built-in functionality, we could open the Mac PageMaker 5 files in Windows PageMaker 5, then re-save the files as Windows PageMaker 5. Next, we opened the Windows PageMaker 5 files in Windows PageMaker 6.5. The files could be converted in this way from one system file to another, and from one version to the next, but not both at the same time. Once we had opened the file in PageMaker 6.5, we had much greater control and were able to fix the things that had been lost during migration. In PageMaker 6.5 we were able to re-link the high resolution images embedded in the layout with where they were stored in the files.

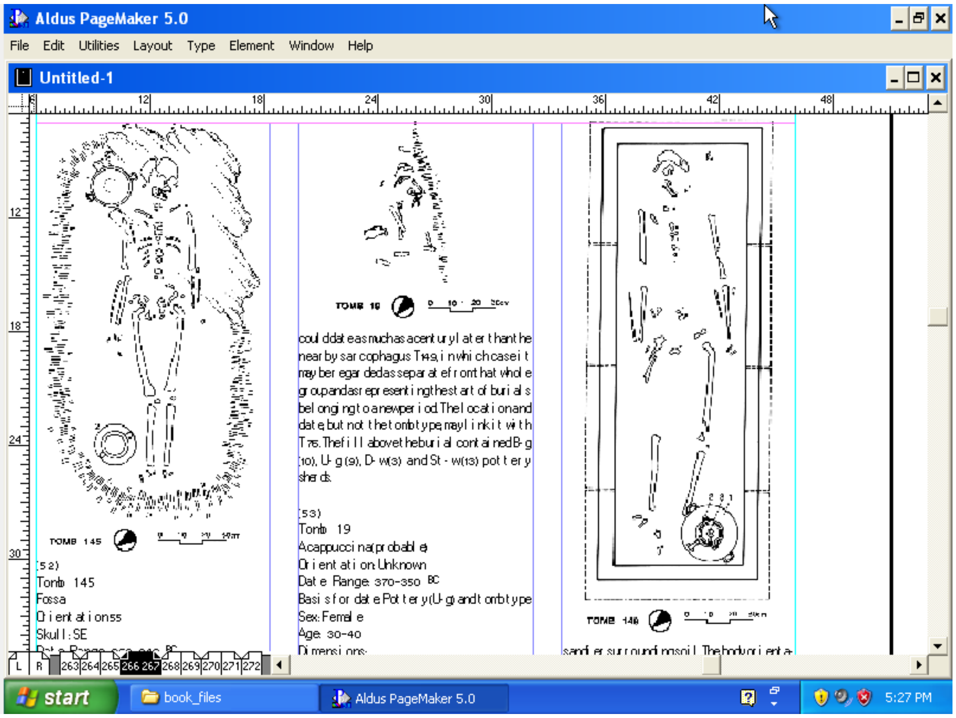



*Figure 7 Page 267, Chapter 7, (Burial Descriptions—Nucleus 4) in Windows PageMaker 5 on Windows XP*

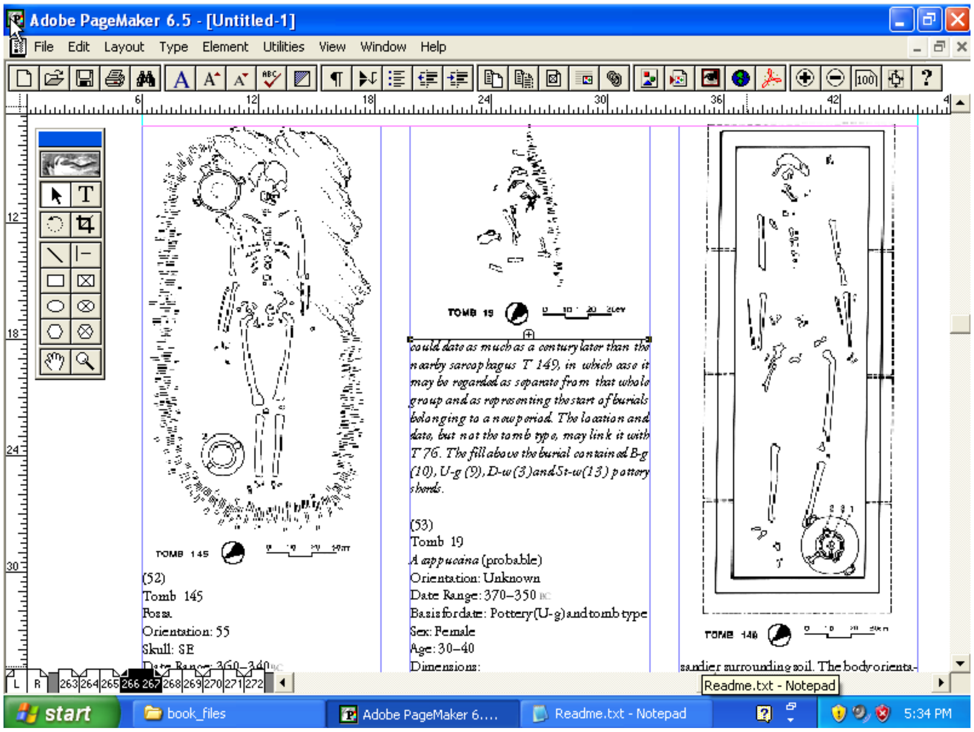



*Figure 8 Page 267, Chapter 7, (Burial Descriptions—Nucleus 4) in Windows PageMaker 6.5 on Windows XP*

As you can see from the two images of the same page in PageMaker 5 and 6.5 [see Figure 7 and Figure 8], the text in the first version did not render correctly because the original fonts had not been saved along with the PageMaker files. A font file is essentially a set of instructions that tells the operating system and the other software how to render text. The common font standard for both Mac and Windows systems since 1980's is called TrueType, which uses vector algorithms to describe the shape of the characters, enabling the computer to render text at a higher resolution than it would with a pixel-based font. Typically, an operating system comes with a set of fonts pre-installed, which are used by multiple different programs, including PageMaker.

Adobe Garamond was drawn by Robert Slimbach and released by Adobe in 1989.[23] The book was designed with the TrueType font AGaramond, which may or may not have come pre-installed on the Mac environment. When we tried to open the PageMaker files on our modern computers, the Windows 10 operating system substituted AGaramond with a similar, but slightly different version of that typeface called Garamond Pro. The differences were clearly visible in the numbers, especially 4 and 5, and the letters w, t, and c. The slight differences of proportions and spacing affected the size of the text, and in cases where the text is related to specific page numbers, and as a result some of the structure of the information in the book was lost.

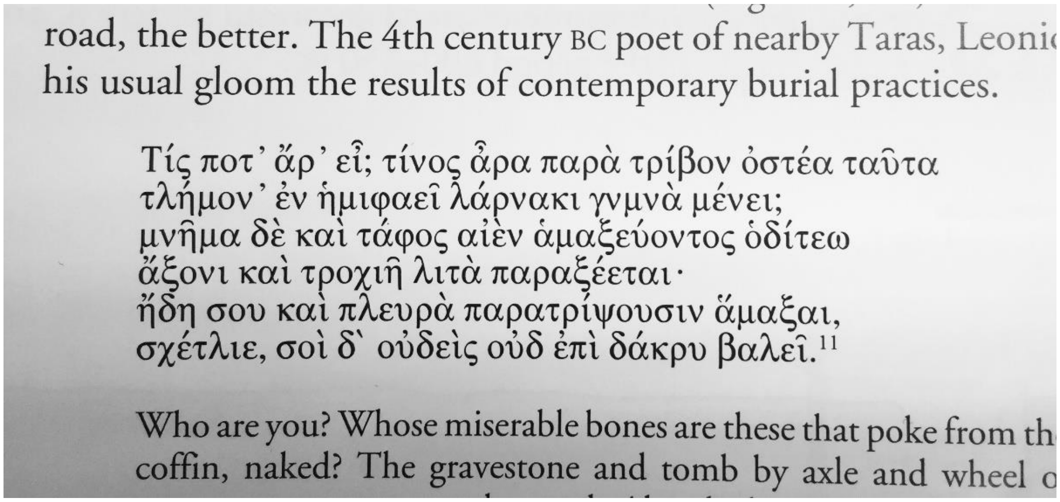
When we used the Windows XP emulated environment, we had to find and install a version of the typeface that looked exactly the same as AGaramond. What we found online was called Adobe Garamond, which we installed in the emulation, and substituted in pace of AGaramond in PageMaker. The resulting text was identical to that of the book.

This solution was possible because Garamond is such a widely distributed and commonly used typeface, which has been distributed by a major vendor, Adobe, for decades. But Garamond wasn't the only font that was used in the book. In Chapter two, page 33, in a section about ancient roads, Carter quotes 4$^{th}$ century Greek poet Leonidas to illustrate the connection between roads and burial sites in the ancient world:

Τίς ποτ' ἄρ' εἶ; τίνος ἆρα παρὰ τρίβον ὀστέα ταῦτα τλήμον' ἐν ἡμιφαεῖ λάρνακι γυμνὰ μένει;
μνῆμα δὲ καὶ τάφος αἰὲν ἀμαξεύοντος ὁδίτεω
ἄξονι καὶ τροχιῇ λειτὰ παραξέεται.
ἤδη σευ καὶ πλευρὰ παρατρίψουσιν ἅμαξαι,
σχέτλιε, σοὶ δ' οὐδεὶς οὐδ' ἐπὶ δάκρυ βαλεῖ.

*Who are you? Whose miserable bones are these that poke from the half-opened coffin, naked?
The gravestones and tomb by axel and wheel of the passing wagoneer are ever ground smooth.
Already the wains rub your ribs. Wretch, and no one sheds a tear for you.*[24]

This quote had been rendered in an ancient Greek typeface called Kadmos-Sal+ and Kadmos-Sal produced by the independent foundry the GreekKeys Project. We searched for a font online that we could substitute for these typefaces, since neither the Apple nor Windows systems came pre-installed with a close approximation. In our search, we got in touch with the now-retired director of the GreekKeys Project, Dr. Mastonarde, who was happy to provide the original fonts for free, but the files we received were incompatible with our systems because they were not TrueType fonts. Without a way to convert these fonts to TrueType, we chose instead to install a modern Greek typeface and re-type the quote. Modern Greek and Ancient Greek are slightly different from each other. Replacing ancient Greek characters with modern ones, which have fewer accent and breathing marks, can potentially leave out information that the author intended to convey. The original as it appeared in the book can is seen below [Figure 9].


road, the better. The 4th century BC poet of nearby Taras, Leoni
his usual gloom the results of contemporary burial practices.

Τίς ποτ᾽ ἄρ᾽ εἶ; τίνος ἆρα παρὰ τρίβον ὀστέα ταῦτα
τλῆμον᾽ ἐν ἡμιφαεῖ λάρνακι γνμνὰ μένει;
μνῆμα δὲ καὶ τάφος αἰὲν ἁμαξεύοντος ὁδίτεω
ἄξονι καὶ τροχιῇ λιτὰ παραξέεται·
ἤδη σου καὶ πλευρὰ παρατρίψουσιν ἅμαξαι,
σχέτλιε, σοὶ δ᾽ οὐδεὶς οὐδ ἐπὶ δάκρυ βαλεῖ.[11]

Who are you? Whose miserable bones are these that poke from th
coffin, naked? The gravestone and tomb by axle and wheel o

*Figure 9 Ancient Greek text from page 33 of The Necropoleis*

Using our Windows XP emulated environment, PageMaker 5 and 6.5, the typefaces we found online, and the illustration files that were saved alongside the PageMaker files, we were finally able to render book layouts in a way that most closely resembled their original form. From here, we were also able to export the files to a .pdf, and transfer the migrated version to the host computer and give them to the ICA. However, the version we produced, at the time of this writing, was awaiting review by ICA staff to ensure accuracy and fidelity to the original book. Also, due to corrupted files that could not be recovered, one of the chapters will need to be re-typed and re-inserted into PageMaker. We were also able to export this emulation package along with detailed instructions on how to use it to view the PageMaker files.

# SECTION II: LESSONS LEARNED

This description of our project is intended to provide an example of successfully utilizing emulation tools for recovering digital files stored in legacy formats. But in keeping with the ideology of reflective practice encouraged by Dr. Galloway, this paper will also provide further analysis and critique, an important aspect of establishing intentionality and rigorous study in the expanding field of digital archiving. Our rapidly evolving technological landscape, and the steadily increasing number of different formats, standards, and iterations of software and hardware, means that archivists will have to work with a multitude of different tools during their careers. Specific challenges will be different with each new project as the environments change. Familiarity with fundamental concepts of digital preservation will help guide archivists when they encounter new and surprising challenges. We learn these concepts through reflective practice. Examining the work we did for the ICA reveals to what extent the materiality of digital objects constrained and determined the process of preservation.

The ICA and our student group had many aspirations for this project that can be summarized as two broad objectives. The first was to archive the files that were created and used during the excavation at Pantanello and in publishing the book. The other was to create access copies of the book and databases by recovering old digital files and migrating them to newer formats.

Examining our decision making process and the challenges we encountered has helped me frame our goals as two different approaches to digital archiving. The first operates on the basis of providing long term preservation of the working files in their original context, while the second focused on migrating the most significant information to new formats for preservation and increased access. Our treatment of these materials differed based on each approach. By preserving the working files in their original environment, we assumed the evidentiary value of the records, not just as evidence of their creation and use, but as social and cultural artifacts. When we migrated the files to new formats, we focused on the textual content of the material, or the information that the book and the excavation databases were intended to convey.

In practice, these two approaches were sometimes in conflict with each other. In order to preserve the working files in their entirety, we had to consider the materiality of the records; this was important for capturing the original bits without altering any of the code, and to correctly interpret them with the correct computing system. Migration of the textual content allowed us more flexibility, and we focused on the immaterial aspects of the records—conceptual information to be transferred from one format to another. These fundamental differences in regards to the materiality of digital information meant that pursuing one objective could potentially put the other at risk.

In the following sections, I will demonstrate how both these approaches guided our process and how they sometimes came into conflict with each other. And I will argue that fully understanding the differences between each approach helps archivists make informed decisions about the tools and processes they need to use in other, similar projects.

**Approach 1: Preserve the working files in the original environment**

Modern archival practice relies on the ideology that documents contain the memory of actions related to their creation and use. Modern diplomatics, as Luciana Duranti explains, is concerned with the study of the creation and use of records, and "their relationship with the facts represented in them and with their creator."[25] Archivists preserve records for their ability to provide evidence of the activities they represent.

A document's evidentiary value can, however, extend beyond the evidence of the specific activity for which it was intended to include signifiers of cultural, visual, or social significance. As Geoffrey Yeo puts it, "The memorial affordances of records are wide-ranging: records may support individual and collective memory, not only of the activities that gave rise to their creation, but also of innumerable other aspects of the world in which they were created, maintained, or used."[26] In this context, documents convey information other than textual information, and can communicate through contextual clues such as their presentation, formatting, medium, structure, and relationship to other surrounding documents.

By preserving the digital working files, we were treating them as evidence of the events surrounding their creation and use. These files were records of the ICA's research methods, the contributions of different team members, and the scale and scope of their work. The ICA considered themselves, and other Classical Greek scholars, the primary user groups for these records.

The floppy disks, for example, contained the databases and other types of records worth preserving as evidence. Earlier I mentioned the written descriptions of each excavation, but we also found things like official correspondence, program files, and other documentation of their methodology [See Figure 10].

The book files were rich with evidentiary value as well. Many of the illustrations in the book were digitally altered for saturation, tone, contrast, removal of the backgrounds, and in the case of the object photography, resizing to achieve a consistent scale for comparison. [Figure 11] The drawings were created in Illustrator rather than drawn by hand, allowing the designers more flexibility and control in layout. Designing and assembling the book demonstrated proficient use of digital technology, and marked the point where the physical labor of archeology and the intellectual work of writing scholarship were brought together.

(You should get 422; mid-date is therefore 434).

Notes

Obviously, if all objects in a tomb have the same date, there is no need to do the weighting process.

An example of a problem which the weighting procedure helps to solve:

Tomb 69

| | |
|---|---|
| 1) Lebes | 420-380 |
| 2) Sq. lek. | 420-400 |
| 3) Skyphos | 440-420 |
| 4) Pelike | 430-400 |

If we are to believe in these dates, the only moment when the burial could have taken place is 420. We could subjectively try to resolve the matter by supposing that the date for the skyphos should extend perhaps a decade later; and if a single person with expertise (i.e. Maria) were prepared to go through each tomb and consistently apply her subjective opinion in each case, this would probably be the best way of arriving at a judicious narrow date range. But that will probably never happen, and so the weighting procecure provides us with a satisfactory way of resolving the situation with an element of objectivity.

In this case, we come up with a narrow range of 428-403, and a mid-date of 416. This is convenient in many ways, since it DOES extend the date of the skyphos just as we had wanted to do subjectively....

*Figure 10 Description of Dating Methods from the excavation records*

*Figure 11 Example of digital editing for the Necropoleis illustrations*

But these materials could also act as what Geoffrey C. Bowker and Susan Leigh Star have called *boundary objects*—“phenomena of interest to many communities.”[27] Besides classical Greek scholars, other types of researchers might also want to study the same files for different reasons. For instance, the functionality and presentation of the now-discontinued Adobe PageMaker could be of interest to historians of computing or graphic design. The initial development of PageMaker in the 1980’s helped solidify Apple’s popularity among artists and designers.[28] Today it is no longer in use but has been replaced with Adobe InDesign; the similarities between their interfaces are striking and it is easy to imagine how someone might study the evolution of desktop publishing software through an emulation of the vintage PageMaker software. We can compare the toolbar on the left hand side of both the PageMaker 6.5 and the InDesign CC 2014, for example, and find many of the same tools are still in use over twenty years later.  [Figure 12]

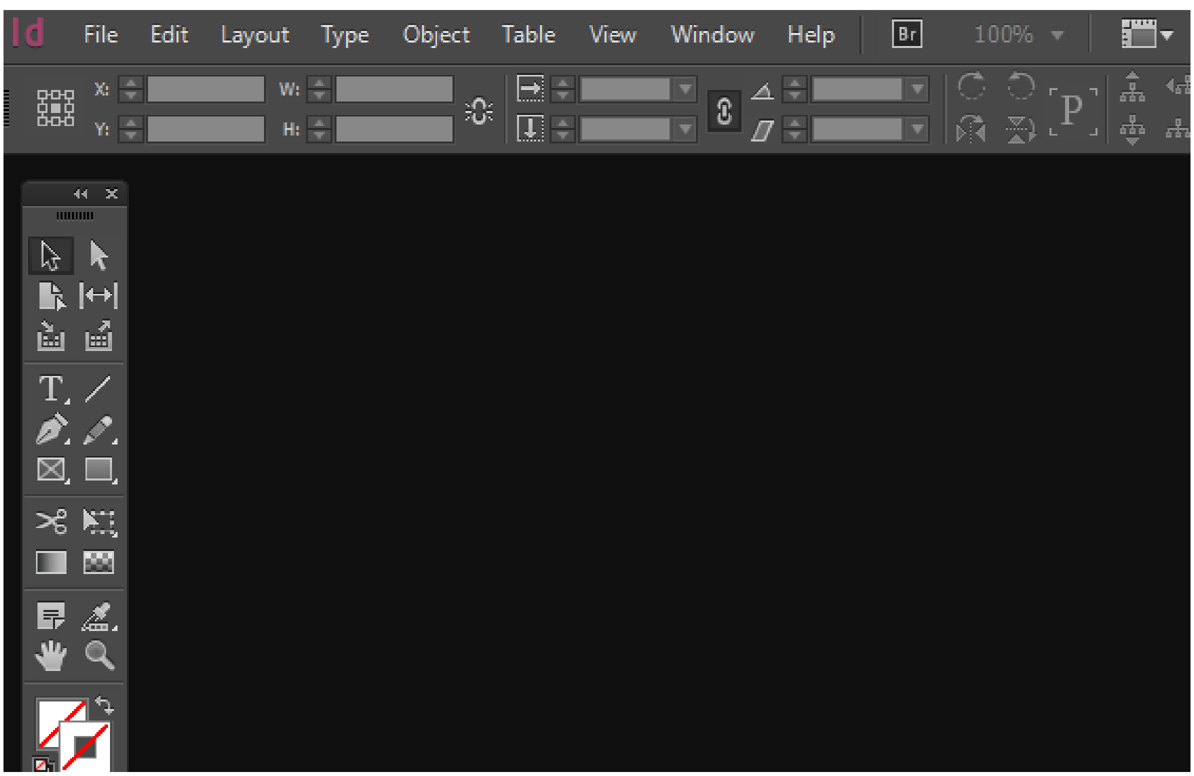



*Figure 12 Adobe InDesign CC 2014 (empty workstation) on Windows 10*

This comparison of legacy and contemporary software, and the evidence of the ICA's work in creating the *Necropoleis*, would have been impossible without a way to access the files in their original context. And, as Rothenberg explains, using the original context ensured that the files were rendered accurately by the software translating the physical bits to information on the screen. Using migration to open the files in different programs lead to errors in the text, as we saw with Figure 1, or other potential errors that were undistinguishable to the human eye. Plus, opening the files in the wrong program caused us to lose other aspects of the document, such as presentation, formatting, and structure.

Emulation also provided a way to study the impact of the original environment on the creation and use of the records. *The Necropoleis,* for example, was unique because it was designed and put together by the ICA. Rather than contract the design of the book to a third party, the ICA asked one of their team members, Anne Parmly Toxey, art director, designer, draftsperson, and co-editor, to do the work in-house on computers and software provided by the University of Texas School of Liberal Arts. This arrangement offered the editors and authors more flexibility, but had drawbacks as well in terms of costs and labor.

The database files were another example of the importance of the context of creation. The database structure influenced the way research was conducted. The early dBase databases are also quite different from the relational databases that are widely available today. As such, these original environments could provide historical insight into the field of archeology by showing us how these experts adapted their methods to take full advantage of the technology that was available to them.

The challenges we encountered in building the emulations demonstrated the necessity of saving as much of the original environment as possible, from the program used to create the document to other things that might be overlooked. The designer of the book recognized the importance of the image files in relation to the way PageMaker functions, and saved them in their original file structure along with the PageMaker files. This meant that we could easily re-link the images in our emulation to produce a digital version of the book with hi-resolution images. However, the fonts had not been similarly preserved, and we had to substitute them with versions we acquired online. While fonts are readily available, other things might have been more specific or original to the ICA's work, and had they not been saved by ICA staff, they could have been lost forever.

This taught us to consider the digital document in terms of the material bits rather than what is displayed on the screen. Through our process of tracking down and testing many different components of the original environment, including the fonts, we started to understand that information in binary form is created, saved, and run through processes invisible to the user. Much of the document exists at the bit level and relies on coding and processes that users are unaware of. The graphic designer, in this case, had the foresight to save the images, but she did not anticipate having to save the fonts as well. With our limited knowledge of what goes on at the bit level, what types of things may be overlooked next time?

Despite some loss of the original parts of the files, we were able to find acceptable substitutes online, and build an emulation that closely replicated the original environment that was used to design the book over twenty years ago. We were lucky that PageMaker worked on both Mac and Windows, and that the version available to us was very close to the version the designer used. With more time, resources, and collaboration with computing experts and programmers, we might even be able to repeat the process with a virtual machine for Apple computers and a copy of PageMaker 5.0 for Mac.

Archivists using this approach should fastidiously preserve all the systems and software that were used in the original creation of the materials, including things like fonts, which operate the same way as software. Access to legacy software is now more possible than ever due to advances in intellectual property law regarding archival, cultural heritage work.[29] Most importantly, while we may not be familiar with all operating systems, we can learn and become familiar with the basic underlying structure of computing systems, how data is stored, and how computers translate that data into information on the screen. This awareness helps us save all the things necessary to preserve the working files for their evidentiary value.

**Approach 2: Migrating files for preservation and increased access**

This approach relied on migration for increased access and as part of preservation. By migrating the book and the databases the ICA could share them with a wider audience, which served the Institute's mission of dissemination of knowledge based on their "firm belief that archaeology is an international endeavor and that information, resources, and technical expertise should be shared."[30] The ICA had a significant impetus for focusing on increased accessibility to the *Necropoleis*—the copyright of the book had just reached the twenty-year mark, upon which it was transferred from the University of Texas Press back to the author, Joseph Coleman Carter, who wanted to make the book accessible online for free.

Creating a .pdf was also a form of preservation through migration, which focused on what the ICA determined to be the most significant parts—the textual information—and rejected other elements such as presentation, medium, and structure. Many digital preservationists have relied on migration as means of protecting information from technological obsolescence. Yeo writes, "The structure, appearance and functionality of objects can all be at risk during migration, but even if some differences are introduced—or so the argument runs—this will be inconsequential if migration is carried out in such a way that no 'significant' loss occurs." Yeo also describes how migration can serve preservation by creating access copies of the material that reduces the risk to the originals by decreasing handling.[31]

It happened that migrating the PageMaker files in the emulated environment was the most effective means of migration. More common methods, such as digitization of the book itself, or using text recognition software to re-create the book's contents, would have been impossible for the *Necropoleis* because of the size and format of the book, and because the ICA wanted to preserve the quality of the illustrations. These methods of converting media to digital formats, however, are frequently used as a means of preservation in other situations, and can provide stability and security for digital materials.

Likewise, in the interest of preserving only the most "significant properties"[32] of the excavation files, we could have migrated the databases directly from their legacy file formats to new file formats on our modern computers, without the help of the emulation. This produced some errors, as we could see in Figure 2, but those errors could possibly have been corrected upon later review by someone more familiar with the material.

It's true that migration can result in encoding errors that are indiscernible to the human eye, but these errors can be mitigated in future preservation projects by having a careful migration plan in place early on. Encoding errors can be minimized with periodic migration of the files each time a new operating system or version of the authoring software is released. Information can also be migrated to file formats that are seen as "archival" because of their ubiquity and lower risk of becoming obsolete.[33]

Archivists approaching digital material this way should clearly identify the most significant properties of their records before beginning work. It was obvious to the ICA that the book should be migrated to a digital format, but they also recognized the significance of the high-resolution images and illustrations in the book, which guided the decision to use emulation to produce the pdf rather than digitization. Archivists can work together with their clients to identify and communicate priorities like this to achieve their goals more efficiently.

The ICA considered migrating the book and databases the most important goal for this project. While the working files had value, providing the community of Greek scholars access to the book and databases was perhaps more significant in terms of fulfilling their mission. Additionally, because of the challenges of emulation and the time and resources required, our efforts to create a perfect emulated environment stopped short once the book and databases were migrated. Using a combination of emulation and migration to preserve the textual information of their materials was the most effective and satisfactory approach for the *Necropoleis* project based on their mission, resources, and timeline.

**Competing approaches:**

How did these two approaches conflict with one another? First, although they appeared to be similar and in many instances overlapped, the process of preserving the files and their environment was still significantly different from the process of migrating the text and data. Each approach called for different tools and methods—with limited resources, we sometimes had to choose one objective over the other in terms of allocating resources. For example, focusing on building a perfect emulation distracted us from finding simpler, faster ways to migrate the content. At some point, we decided to stop building the emulation in SheepShaver and devoted the rest of our time to finding a different way to produce the book as a .pdf from the PageMaker files. Our solution was to use VirtualBox to build an emulation of Windows XP, which was close to the original environment, but not exactly the same.

Secondly, migration tactics sometimes ignore the original bit stream, putting information at risk. Migration as a means of preservation captures the most significant properties and transfers them to a more stable file format, generates more copies of the records to increase their odds of survival, and creates access copies that can protect the original record from the risk of frequent and unsafe handling. But, as Yeo points out, these methods "are often motivated not so much by long-term preservation requirements as by desires to make archival resources more immediately available to wider audiences."[34] Sometimes, migration gives the impression that the files are safe, when in reality, there will be (or already has been) significant loss at the bit stream level.

The ICA engaged in a type of migration when they backed up the files on ZIP drives and when they transferred the files to the new server. In the process, they lost parts of the bit stream that contained possibly valuable information such as time and date stamps, the identity of the authors and contributors, draft version numbers, and other metadata generated by the authoring program.

Similarly, once the files were moved to the ZIP drives, they were separated from the font files that were necessary to tell the computer how to render the ancient Greek font. Fonts might seem like small presentation details, but differences in accent marks can sometimes alter the meaning of the word. The presentation and formatting of digital information is key to preserving the intellectual content of the files, and can only be preserved by saving the original bit stream.

# CONCLUSION:

This paper describes the use of emulation for recovering files in legacy formats, but it also critically examines our process as a means of identifying the significant skills we acquired and the concepts we learned. This reflective practice is an important step for all professionals, but especially so for those working with digital and electronic material. As Galloway explains, working with technology requires problem solving and research skills that are best learned through direct application, as well as learning theories and concepts behind the methodology and adapting them to real-world examples.

Our student group learned technical skills by working with legacy systems. The early, less sophisticated versions of computing environments taught us the basics of software and hardware functionality that we can later apply to other, more sophisticated models. But learning digital archiving also requires familiarity with fundamental theories and concepts and knowing how to apply them to different situations. Rapidly evolving technology, and the variation between different proprietary, commercial distributions of that technology, means that archivists should be prepared to encounter and address a "whole set of uncertainties."[35] As Schön says, professionalism is embodied in knowing how to reflect on our actions and "use this capacity to cope with the unique, uncertain, and conflicted situations of practice."[36] The reflective practice allows us to learn from our project not just in terms of technical skills, but in concepts.

With reflection, we identified basic concepts at work within the different approaches to digital archiving, and how those concepts sometimes came into conflict with each other. The specific conflicts we encountered demonstrate how each approach conceptualized the records as either evidence or textual content and how these conceptualizations determined the way we treated the records during recovery and preservation. Understanding these differences and clearly differentiating between each approach can make preservation efforts more efficient.

Our challenge was to use emulation to recover decades-old files in legacy formats, and migrate the files to new, digital versions to be shared online. At first glance, it seemed as though all our objectives could be achieved simultaneously by emulating older computer systems to open legacy files. However, it became apparent that preserving the working files in their original environment, and migrating them to new formats, required different methods which could come into conflict.

Preserving the working files meant saving the original bit stream and providing future researchers a way to look at the files in their original environment through emulation. This approach considers the evidentiary value that can be conveyed through the context of creation, including non-textual information such as the presentation, formatting, medium, and structure of a document. Migration of the book and data meant increasing the accessibility and preserving the most significant properties of the documents, the textual information.

The archivist should be able to articulate the differences between these two frameworks and how to achieve each objective. They should also be able to identify which framework is best suited for the project at hand, depending on the materials, the resources available, and the stakeholder's specific needs and goals. Every institution is different. For some, migrating the most important files to new formats will be enough. Emulation requires more resources in terms of expertise, time, and access to software. While migration can result in formatting errors, these can be managed with quality control checks and a thoughtful migration plan. Migrating files, especially making material newly available as digital resources, increases accessibility, which can be a powerful motivating principle for some stakeholders. Cultural heritage and research institutions in particular may be more inclined to share their information quickly and simply.

On the other hand, using emulation to recreate the original environment preserves the records in their entirety. This preserves the collection of working files as evidence of the institutions' labor, and as social and cultural artifacts. It considers things like stylistic elements, presentation, formatting, file structure and organization, the functionality of the creating software, and file format attributes, as equally important as the textual information. To the primary user group, these elements may not be significant enough to be migrated, but they might be crucially important to other researchers later on. As Rothenberg says, "Without knowing what elements of the document are important to the future audiences, we cannot know what parts of the documents are significant enough to translate to the new migrated formats.[37]

Yeo uses, as an example, textual critics studying the archival documents of Emily Dickinson and Jane Austen who find meaning in the spacing of text, paragraphs indentions, and writing in the margins, These elements might have been lost if someone had re-typed the textual content of the documents and dismissed these details as not significant enough for preservation. He points out the similarity to the formatting and presentation elements of digital documents that can be lost during migration and says, "When textual critics begin to take an interest in digital archives of the twenty-first century, some might argue that [these elements] constitute significant evidence of intentionality or meaningfulness,"[38] and that "Just as textual critics' views of which properties are significant in Dickinson's correspondence have changed over time, a feature or features that seem unimportant to one generation of researchers in any user community may be perceived as highly valuable by their successors."[39]

In addition to stylistic elements, digital materials depend on structural elements that can be important to future researchers and which might be lost in migration. The structure and functionality of the databases as they existed in dBase IV on an early Windows computer had a direct impact on the way data was collected and manipulated, and may be valuable for historians studying the development of archeology as a discipline. Likewise, the PageMaker files in their original environment provided insight on the functionality of vintage computer programs, which might be useful to historians of computing and graphic design. But the structure and functionality is lost when migrating the data or the layouts to new formats to be opened in more modern programs. An archivist should be able to help stakeholders understand what can be saved, and what might be lost when using different methods of preservation and help them identify their archival goals based on their resources and needs.

Finally, an archivist should also make it clear to the stakeholder that while these two approaches have different merits and can be valuable in different ways, they are based on contradictory concepts of digital information. Saving the working files in their original environment focuses on saving the bits in their entirety and the computing system (hardware, software, and operating system) that will correctly translate them. Migrating just the most significant parts, or the content, ignores the original bit stream, putting the records at risk. The bit stream contains information that tells the computer how to translate those bits and render the information on the screen. Loss of information at the bit level can cause irrevocable loss of encoding that is "meaningful only to the software that created them,"[40] but which are vital to translating the bits into meaningful information. Since the user may have no idea what is happening at the bit stream level, it is impossible to make an objective decision on what should be preserved and how. When weighing the value of migration and emulation as preservation strategies, it is important to keep in mind that there is always a risk to the content when you fail to consider the digital material as physical bits on a hard drive. That original bit stream should always be preserved, regardless of what other steps the archivist takes.

Drawing from the concepts we see at work in our project, we can approach future projects of this nature with a few guiding principles. Using emulation to preserve the Institution's working files, and migrating the book and data bases to new formats for preservation and increased access, were two distinct approaches to preservation that treated information either as evidentiary records or as textual content. As such, these two approaches required different tools, processes and methods when handling the material. These methods can conflict with each other, but that is not to say that they are entirely antithetical. Both can be applied in the same archival project, with awareness of the differences and careful planning.

In order to help archivists allocate resources, it is important to identify which of the two objectives takes precedent in the project. With the *Necropoleis* materials, we built two emulated environments to open their working files, and then exported the emulators so that future researchers can access them as well. We also migrated the book and databases to new formats. Along the way, we had to sacrifice our search for a perfectly emulated environment for the sake of migrating the book. When the SheepShaver virtual machine failed, we moved to VirtualBox and used a Windows XP environment instead of the original Mac. We also substituted a Modern Greek font instead of spending more time and resources trying to convert the Ancient Greek to a useable format. While they may not be perfect replicas of the original contexts of creation, these emulated environments are still useful to future researchers because they provide a way to access the legacy files and preserve as much of the information inherent to the original files as possible. Understanding the differences and clearly articulating objectives allows the archivist to select and implement the right tools and processes for each unique situation.

# Works Cited:


[1] Joseph Coleman Carter, “The Pantanello Necropolis 1982-1989: An Interim Report” The University of Texas at Austin, 1990.

[2] Lauren Jackson, email correspondence with the author, April 1, 2019.

[3] “The Chora of Metaponto: A Digital Companion to the Publication Series, Pantanello The Necropoleis,” (https://metaponto.la.utexas.edu/pantanello/necropoleis/) accessed March 2019.

[4] Patricia Galloway, “Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice,” *Archivaria, The Journal of the Association of Canadian Archivists*, 72 (Fall 2011): 169–196.

[5] Galloway, “Educating,” 174.

[6] Josh Conrad, Kathleen Forrest, Birch Griesse, and Nicole Lumpkins, “Greek to Me: Using Emulation to Digitally Re-publish the *Chora of Metaponto: The Necropoleis.”* Project Report for INF392K Digital Archiving, Spring 2018, University of Texas at Austin School of Information

[7] Joseph Coleman Carter, *The Chora of Metaponto, The Necropoleis* (University of Texas Press: Austin, Texas 1998), xii-xiii.

[8] Kenneth Thibodeau, “Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years,” Council on Library and Information Resources, last updated July 10, 2002, accessed April 13, 2019, (https://www.clir.org/pubs/reports/pub107/thibodeau/).

[9] David M. Levy, “Where’s Waldo? Reflections on copies and authenticity in a digital environment,” *Council on Library and Information Resources, Authenticity in a digital environment,* accessed April 13, 2019, (https://www.clir.org/pubs/reports/pub92/levy/).

[10] Luciana Duranti, “Concepts, Principles, and Methods for the Management of Electronic Records,” *The Information Society: An International Journal*, 17 no. 4 (2001), 271-279; pg 272.

[11] Matthew G. Kirschenbaum, “Digital forensics and born-digital content in cultural heritage collections,” *Council on Library and Information Resources*, (2010); 25.

[12] BitCurator website, accessed April 13, 2019, (https://bitcurator.net/bitcurator/).

[13] Jackson, email, April 1, 2019

[14] Jean-François Blanchette, "A Material History of Bits." J*ournal of the American Society for Information Science and Technology* 62, no. 6 (2011): 1042-1057.

[15] Ciaran B. Trace, "Beyond the Magic to the Mechanism: Computers, Materiality and What it Means for Records to Be 'Born Digital,'" *Archivaria* 72 (Fall 2011): 5-27.

[16] Jeff Rothenberg, "Ensuring the Longevity of Digital Information." *International Journal of Legal Information* 26, no. 1-3 (1998): 1-22.

[17] Rothenberg defined migration as the translation of digital information into new formats. See Rothenberg, "Ensuing the Longevity," 11.

media (and possibly translating it into new formats

[18] Carter, *The Necropoleis,* 475

[19] A device driver is a special program required to make some hardware device accessible to a given computer System, see Rothenberg, "Ensuring the Longevity," 2.

[20] Patricia Aufderheide et al., "Code of Best Practices in Fair Use for Software Preservation," (Coordinated by the Association of Research Libraries, the Center for Media & Social Impact, and the Program on Information Justice and Intellectual Property, 2018).

[21] Detailed instructions on how to build a Windows 3.1 emulation were written and provided for free by Joshua Lovering, who included links to places where I could download software and drivers, and who kindly answered my questions through email when I got stuck. Joshua publishes these instructions, and other helpful information on vintage computing, on his blog; see Joshua Lovering, Socket 3: Bits & Bytes On Vintage Computing, "INSTALL & CONFIGURE MS-DOS 6.22 & WINDOWS 3.1 USING ORACLE VIRTUALBOX", first accessed April 10, 2018 (https://socket3.wordpress.com/2016/08/25/install-configure-ms-dos-6-22-and-windows-3-1-using-oracle-virtualbox/)

[22] Carter, *The Necropoleis*, 455-456.

[23] Monotype Imaging, Inc. "Illuminating Letters, Number One: Garamond," 2008

[24] Carter, *The Necropoleis*, 33

[25] Luciana Duranti, 1998, *Diplomatics: New Uses for an Old Science* (Lanham, MD: Scarecrow Press); 45.

[26] Geoffrey Yeo, "'Nothing is the Same as Something Else': Significant Properties and Notions of Identity and Originality." *Archival Science* 10, no. 2 (2010): 85-116, 104.

[27] Yeo, "Nothing is the Same," 98.

[28] Peter C.S. Adams, "PageMaker Past, Present, and Future," PageMaker for Desktop Publishers, Last modified March 16, 2004, archived on the Wayback Machine, accessed April 19, 2019 (https://web.archive.org/web/20070709042131/http://www.makingpages.org/pagemaker/history/).

[29] Aufderheide, "Code of Best Practices"

[30] University of Texas at Austin Institute of Classical Archeology website, "About," accessed April 29, 2019, (https://liberalarts.utexas.edu/ica/about/mission.php).

[31] Yeo, "Nothing is the Same," 87.

[32] Significant properties can extend beyond the content to be saved to include things like structure, format, and presentation. The Cedars Project Report, "Cedars Guide to: Digital Preservation Strategies," April 2, 2002, accessed April 29, 2019, (http://www.imaginar.org/taller/dppd/DPPD/146%20pp%20Digital%20Preservation%20Strategies.pdf)

[33] Library of Congress, "Recommended Formats Statement 2018-2019," accessed April 14, 2019, (https://www.loc.gov/preservation/resources/rfs/RFS%202018-2019.pdf)

[34] Yeo, "Nothing is the Same," 87.

[35] Galloway, "Teaching," 171.

[36] Donald A. Schön, *The Reflective Practitioner: How Professionals Think in Action* (New York: Basic Books, 1983) preface, viii.

[37] Rothenberg, "Ensuring the Longevity," 13.

[38] Yeo, "Nothing is the Same," 97.

[39] Yeo, "Nothing is the Same," 103.

[40] Rothenberg, "Ensuring the Longevity," 18.