A Focus on the Future of our Tiny Piece of the Past: Digital Archiving of a Long-term Multi-participant Regional Project

Madry Scott[1]*, Jansen Gregory[2], Murray, Seth[3], Jones, Elizabeth[1], Willcoxon, Lia[3,] Alhashem, Ebtehal[4]

[1] Research Laboratories of Archaeology
The University of North Carolina at Chapel Hill, NC USA
[2] Advanced Information Collaboratory
The University of Maryland, College Park, MD, USA
[3] College of Humanities and Social Sciences
North Carolina State University, Raleigh, NC USA
[4] School of Information and Library Sciences
The University of North Carolina at Chapel Hill, NC USA

*Corresponding author
Correspondence: madrys@email.unc.edu

Abstract

This paper will consider the practical realities that have been encountered while seeking to create a usable Digital Archiving system of a long-term and multi-participant research project. The lead author has been involved in archaeological and landscape research in the Burgundy region of France for the past 45 years. This long-lived project has continued across several generations, institutions, continents, and disciplines, and began in the mid 1970's before many of our commonly used digital data types and capabilities even existed. Over the decades, many individual researchers, students, and local community members have participated in our broadly defined research activities, conducting field and laboratory research, and they have, cumulatively, woven a tapestry of knowledge regarding some

2,000 years of the interaction between peoples and their landscapes in our study area. Many project participants have moved on to other interests and some have passed away. Homes and personal archives have sadly burned, and offices and labs have been flooded. All while an analogue method of work has transitioned to a new digital paradigm that is completely unrecognizable from how we began our journey. As this project slowly winds down, the issues of both analog and digital data preservation and the means of providing continued access to other researchers who may be interested in accessing our vast repositories and datasets has become one of great interest to our group. How can we address the proper archiving and metadata of thousands of individual analog and digital records and datasets located in multiple institutions and attics? How can we even accurately know what we all have? How can these be properly archived and preserved? And most importantly, how can other researchers gain access to these for future use after we are no longer here to share them? This is the topic of our paper.

Keywords: Digital archiving, archaeology, Burgundy, France, Historical Ecology, Dataverse, Dryad

Kommentiert [DH1]: Add 1 sentence: what is it all about? (see comment below)

Kommentiert [DH2]: Delete

Introduction

Researchers have been conducting a long-term program of investigations in the southern portion of Burgundy, France since the mid 1970's (Crumley and Marquardt 1987, Madry et al. 2023). This unusually large and diverse project has amassed a very large and disparate amount of data in many formats over the decades. These data have included archaeological surveys, aerial surveys, historical documents, geophysical and geological data, aerial photos, remote sensing, maps, GIS, GPS, and other data, collected over a 45-year period by multiple researchers from many different institutions. Initially, these data were only 'archived' by the individual researcher who collected the data, often kept in personal or university facilities, file drawers, and map cases. In the early days of the project, little thought was given to the concept of permanent archiving of these data, and as the research world transitioned from an analog to digital paradigm, many of the original records and data were simply ignored.

Digital archiving has become an established field of study, primarily within schools of library and information science. It integrates traditional library and archival methods and theory, computer science, databases, the internet, and the various disciplinary traditions of those seeking to properly archive their data. Digital archiving began more as an end-of-career records repository when researchers became concerned about preserving the integrity of their research holdings and ensuring continuity of access. But it has evolved to be a tool provided to new professionals to be used throughout one's career, providing an easily accessed repository while work is underway, and creating a lasting digital repository when a project or career is completed. This has also been a subject of important discussion within the archaeology community and the Computer Applications in Archaeology community (Huvila 2008, Wright and Richards 2018, Richards et al. 2021).

As our project has begun to wind down, and more of our team are nearing or in retirement, we have begun the process of considering what life after this project will be like, and what should happen

Kommentiert [DH3]: Please add one more sentence to the whole project; what is the project all about? dating? Thematic frame? If this would be to complicated, then please add some information about why the topic is so diverse and complex...

Kommentiert [DH4]: Please include the information from this sentence in the abstract (see comment above).

Kommentiert [DH5]: I'd suggest to rewrite the sentence, in order to acknowledge the various sources and methods which are mixed up in the actual sentence; maybe replace with something like that:

The data gathered over a span of 45 years by a plethora of researchers from diverse institutions encompasses a wide array of sources. This includes data from archaeological and aerial surveys, information derived from historical documents, geophysical and geological datasets, aerial imagery, remote sensing outputs, maps, Geographic Information System (GIS) resources, Global Positioning System (GPS) readings, among other data types.

Kommentiert [DH6]: internet technology

to the large amounts of data that have been amassed. Commonly, for such projects, files were simply placed in cardboard boxes and, hopefully, these would be retained in a university storage somewhere. But oftentimes, these unique records were simply pitched out, either by the researcher themselves when the 'downsized' or moved, or by family members who survived them, not knowing the importance of these records. Our project began to consider what the options might be, and began a tentative and, frankly, uninformed process of trying to learn about what digital archiving was and what our options might be. Initially, we did some web searches and discovered the Dspace system, maintained by the Massachusetts Institute of Technology (MIT). We made a tentative attempt to download the system and begin an archive of our data, but we were not successful, partly due to the unfamiliarity with the nature of digital archives, partly due to our lack of experience in the field, and partly due to the large number of records and files that we held. Our first attempt was a failure.

Later, in our evolving discovery process, we evaluated several currently available digital archiving environments and options, and we had the good fortune to meet and work with experts in the field, which enabled us to better understand the processes, strengths, and limitations of these various capabilities, and to comprehend how we and our particular data collections might fit into modern digital archiving environment. This paper presents the initial results of our work.

Methods

An earlier attempt at archiving our large and disparate data collections ourselves using DSpace was not successful, due to the large amount of data and, as archaeologists, our unfamiliarity with the digital archiving world. The pressing importance of creating a permanent repository for our data within a reasonable amount of time with the resources and skills available to us has led to a different approach, and recently we conducted a comparative analysis of different digital archiving environments to assess their suitability for the preservation and dissemination of our archaeological data.

Having enlisted the assistance of experts in the field, we selected two digital archiving environments for consideration, based on their popularity, functionality, and accessibility: Dataverse, and Dryad. We evaluated their technical features, such as metadata standards, file formats, storage capacity, interoperability, and access control. These two environments were selected because two members of our project maintain faculty status at two universities that maintain digital archive infrastructures for their faculty in these two environments.

Since two of our home institutions use Dataverse (UNC-CH) and Dryad (NCSU) for their digital repositories, these became obvious choices for us. Having an institutional repository, maintained by our universities and with skilled archivists available to support our activities was a clear improvement over our failed initial efforts. UNC-CH uses the Davaverse system for their digital data repository, originally developed by Harvard University (https://dataverse.org/) and in use by nearly 100 institutions around the world. This is managed by the UNC Odum Institute Data Archive (https://odum.unc.edu/archive/), which has over 50 years of data management experience. Their mission is to provide trusted long-term preservation and stewardship of research data assets to broaden scientific inquiry, promote research producibility, and foster data fluency now and into the future.

Dryad is used by NC State University, which provides free digital data archiving for its faculty and students. The data repository has integration features with Zenodo, a useful tool in archaeological publication and research. Zenodo is an open repository developed and operated by CERN in Europe

Kommentiert [DH7]: This would better fit the method section and is redundant since the paragraph is already mentioned there too. I'd suggest deleting this section.

Kommentiert [DH8]: Maybe rewrite for clarity:

Our project allowed us to gain a deeper insight into these various capabilities' procedures, advantages, and constraints. Furthermore, it helped us understand how our specific data collections could be incorporated into the contemporary digital archiving landscape.

Kommentiert [DH9]: Please delete

Kommentiert [DH10]: Please add a reference for Dspace

Kommentiert [DH11R10]: You can further add here "(...) maintained by the Massachusetts Institute of Technology (MIT)." from the paragraph above

Kommentiert [DH12]: Is there a URI for Dryad available too?

and permits the deposit of a wide range of documents, data, reports, and more in a structured and recoverable format. Dryad is a non-profit organization that provides a curated repository for data underlying scientific publications (*Dryad | Good Data Practices*, n.d.). It supports a wide range of file formats, metadata standards, and identifiers. It also offers long-term preservation, data citation, and integration with journals and other repositories. Dataverse is generally comparable to Dryad in capability and features, but they are two separate systems.

To ensure the quality and reproducibility of our research, we follow the FAIR principles for data management promoted by both Dataverse and Dryad (*The FAIR Data Principles – FORCE11*, n.d.). FAIR stands for Findable, Accessible, Interoperable and Reusable. This means that our data will be assigned persistent identifiers, stored in a public repository with clear metadata and licenses, compatible with multiple formats and tools, and available for reuse by other researchers with minimal barriers. In order to efficiently apply these principles, we began by creating and using an informational Excel spreadsheet that remained consistent with data organization standards for spreadsheets (Broman & Woo, 2018).

Having made the choice of these two institutional repositories and their respective archiving environments, we began an initial data search and data preparation process to prepare our data for entry into the two archives. Upon completing the data preparation process, we will upload our datasets to our respective institutional repositories. During the upload process, we will assign appropriate metadata standards, file formats, and persistent identifiers to each dataset, following the recommendations of the respective platforms.

Kommentiert [DH13]: Could you mention the metadata standards?

By using Dataverse and Dryad as digital archiving environments for our archaeological data, we aimed to promote transparent, reproducible, and collaborative research practices, ultimately contributing to the advancement of the field. Our hope is that our data will prove useful to future researchers.

Process

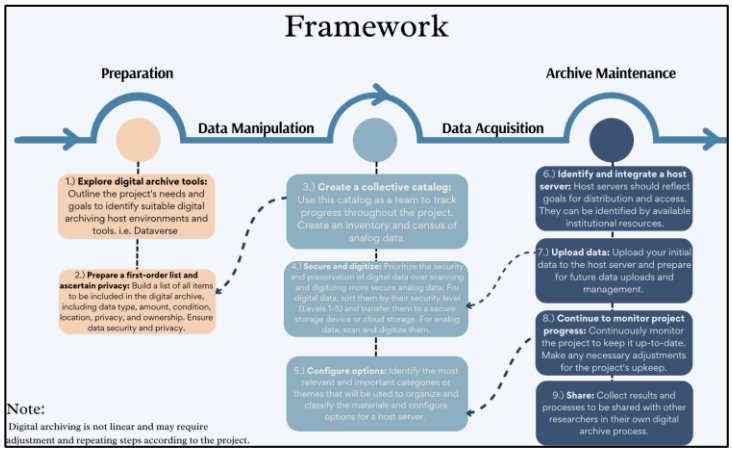


Figure 1 - A framework model for developing our digital archive.

Having decided on the use of these two institutional archiving systems, we have begun the initial analysis of our needs and goals, along with - given our inexperience - an initial investigation into what digital archiving is and is not, and what it, in its current state of evolution, can (and cannot) provide to our particular situation. It has been a steep learning curve for those of us without experience in this field, but we have been fortunate to have the assistance of subject matter experts who have eased our way. We have connected with some very experienced digital archivists at several institutions who have been extremely helpful and generous with their time and expertise to assist us.

Locating Data

As a practical first step, we have begun to amass the first order listing of all extant project data, in both analog and digital formats, from both current and former project participants. Over 40 people have worked on the project at some point, and several are now retired and, sadly, several are no longer with us. We have reached out to current and former participants in our project and have asked them if they hold any data from the time of their involvement in the project, and if they would be willing to share their data with us to include in the digital repository. Several have responded positively and have provided us with trays of 35 mm slides, personal papers and recordings, and digital documents. Two previous project participants very regrettably suffered major house fires at various times, and there was a flooding event at Dr. Madry's lab at Rutgers university, resulting in the loss of important project data, some of which had no backup or duplicate copies. This reinforces the importance of creating and uploading such data while the project is ongoing, rather than simply storing data in the closet.

Cataloging our Data

Our second step was to create the first order cataloging of all extant project data, in both analog and digital formats. This catalog was created and designed to include information on general data type, amounts, condition, location, sensitivity, ownership, etc. An initial Excel spreadsheet on a secure, shared Google folder was created to catalog what and how much of each type of record has been located for use in this project.

Kommentiert [DH14]: Could you please clarify, for the record, if you indeed used an MS Excel spreadsheet in Google Drive? Thanks!

Kommentiert [DH15]: Do you mean Google Drive?

Data

Our project, with its many individual components, has generated significant amounts of disparate data. Fortunately, over the years, much of our existing analog data has already been scanned or is in the process of being scanned. These include large numbers of 35 mm slides (remember those?), paper field documents, field maps, aerial photographs, paper and digital historical documents and manuscripts, audio recordings, paper and scanned digital maps, and other field and lab documents and archival records. Most of these still do not have any item-level metadata, provenance, or keywords at this point, and this will be a vital component of our subsequent work. Our digital data include large amounts of scanned maps, documents, and photos, and also an extensive project GIS and remote sensing archive, databases, websites, presentations, posters, and more.

Kommentiert [DH16]: I unfortunately can definitely remember them 😅

Kommentiert [DH17]: Did you use specific metadata standards? If yes, which? If not: why not?

Kommentiert [DH18]: Did you use a specific controlled vocabulary for the keywords?

Kommentiert [DH19]: These listings of the various data are somewhat redundant; please clarify.

Work Program and Project Strategy

Having chosen our two institutional digital archiving software systems and repositories and having learned what institutional support will be available to our project members, we next had to map out a realistic work program to populate our archives at our two respective institutions. We then began the actual process of creating, annotating, populating, and managing our project digital archives.

A digital archiving strategy was developed that outlines how the materials are to be digitized, stored, and made accessible. This strategy included factors such as the format and size of the materials, the required storage capacity, the sensitivity of the data, and the accessibility needs of future users. The proposed framework for the Burgundy digital archiving process began by evaluating these current digital archiving environments (Dataverse and Dryad) in this context. This involved assessing factors such as modes of use, scalability, and potential for integration or interoperability with other tools and platforms such as those made available by our institutions.

Initial Data Catalog

An initial catalog of all existing data, in both analog and digital formats, from all current and former project participants was generated in an online Google Spreadsheet, as shown below in figure 1.

Product	Rationale	Origin	Producer(s)	Contributor	Medium	Format	Access	Repository	Owner
Maps	Maps of Burgundy	Multiple, IGN, BNF, AN, LOC	Multiple, IGN, BNF, AN, LOC		Paper	Analog		UNC Dataverse	S. Madry
Maps digital	Scanned maps	Multiple, IGN, BN	Multiple, IGN, BNF, AN, LOC		Digital	Digital		UNC Dataverse	S. Madry
Vertical aerials analog	1945 aerials	US DOD	US DOD			Analog		UNC Dataverse	S. Madry
Vertical aerials digital	1945 aerials	US DOD				Digital		UNC Dataverse	S. Madry
IGN aerials digital	IGN mapping pho	IGN				Digital		UNC Dataverse	S. Madry
Oblique aerials	aerail photos	Multiple				Analog		UNC Dataverse	S. Madry
Oblique aerials digital	aerail photos					Digital		UNC Dataverse	S. Madry
Remote Sensing data	Satellite imagery of various dates	Landsat, ESA, JAXA	US, ESA, Japan			Digital		UNC Dataverse	S. Madry
GPS files	GPS ground points		US GPS			Digital		UNC Dataverse	S. Madry
GIS files	GIS data		Multiple			Digital		UNC Dataverse	S. Madry
35mm slides	ground and aerial photos					Analog		NCSU/UNC	
35mm Scanned						Digital		NCSU/UNC	
Print photos	ground and aerial photos					Analog		NCSU/UNC	
Scanned photos						Digital		NCSU/UNC	
Videos and movies						Digital		UNC Dataverse	
GIS animations/flythroughs						Digital		UNC Dataverse	S. Madry
PowerPoint presentations						Digital		NCSU/UNC	
Poster papers						Digital		NCSU/UNC	

Figure 2 - Spreadsheet of Digital Archive Catalog

This list includes information on the data type, rationale for archive, location. origin, producers, and contributors, medium, data format, current mode of access (analog or digital), privacy and activity, data size, geotagging, keywords, and physical or digital repository location. An Excel spreadsheet in a shared Google Docs folder, titled “Digital Asset Inventory”, was created to catalog the data, and track the progress of the project, so that all interested project participants had access and could provide input into the process, add additional data, and assist in the work. When cataloging the resources found in our inventory, special

Kommentiert [DH20]: This would be a great place to explain in 1-2 sentences why exactly you are using two systems, as outlined in the "project status" section. Does each of the two repositories contain the entire dataset? Or are some specific datasets archived in repository A while others are archived in repository B instead?

Kommentiert [DH21]: Since you mention the title of the "Digital Asset Inventory" too below, it would be good to mention the title fo this catalog ("Digital Archive Catalog"[?]) too.

Kommentiert [DH22]: These are nowadays called "Google Sheets" as far as I know

Kommentiert [DH23]: Do you mean figure 2?

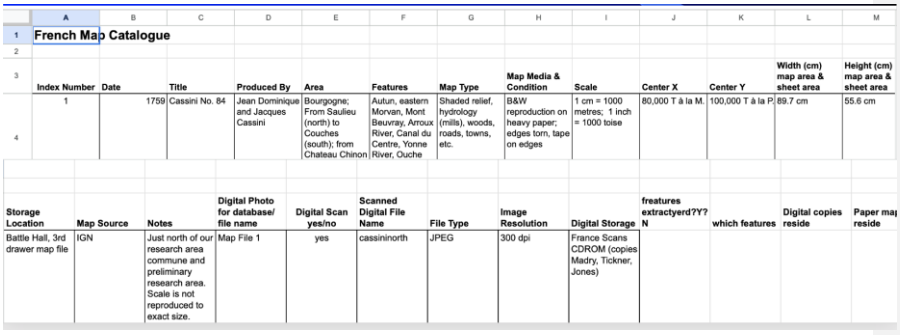
Kommentiert [DH24]: So you are using both Excel and Google spreadsheets? Do you mean a Google "Drive" folder?

care was taken to ensure that human subjects and sensitive archaeological site data were appropriately protected.

This spreadsheet was created, listing all of the types of data, owner, security levels, keywords, etc., as shown above in figure 1 above. After an initial review and much internal discussion and advice from our digital archiving experts, a total of 16 separate 'buckets' of archival data were decided upon, and individual spreadsheets were then created for each of these, which contained details specific to each data type: These 16 data buckets are:

1. Ethnographic interviews and recordings
2. MSWord, Excel, or PowerPoint documents (including Google docs and OpenOffice).
3. Digital GIS data in either raster or vector format
4. Remote Sensing Imagery from multiple satellite systems dating back to 1964
5. Scanned aerial photographs
6. Historical documents of many types
7. GPS locations and geotagged photo data
8. Academic papers and reports
9. Genealogy data
10. Html sites and html data
11. Videos and movies, including computer visualizations
12. Paper maps
13. Map inventory of the Virunga Volcano project, Rwanda
14. Scanned 9x9 inch 1945 aerial mapping photographs
15. Data backups
16. References, articles, and related documents

Each of these bucket spreadsheets contain unique categories of information relevant to that specific data. A portion of the current working spreadsheet for our collection of over 200 paper maps is shown in figure 3 below. This contains 26 categories of information for each map.



French Map Catalogue

Index Number	Date	Title	Produced By	Area	Features	Map Type	Map Media & Condition	Scale	Center X	Center Y	Width (cm) map area & sheet area	Height (cm) map area & sheet area
1	1759	Cassini No. 84	Jean Dominique and Jacques Cassini	Bourgogne; From Saulieu (north) to Couches (south); from Chateau Chinon	Autun, eastern Morvan, Mont Beuvray, Arroux River, Canal du Centre, Yonne River, Ouche	Shaded relief, hydrology (mills), woods, roads, towns, etc.	B&W reproduction on heavy paper; edges torn, tape on edges	1 cm = 1000 metres; 1 inch = 1000 toise	80,000 T à la M.	100,000 T à la P.	89.7 cm	55.6 cm

Storage Location	Map Source	Notes	Digital Photo for database/ file name	Digital Scan yes/no	Scanned Digital File Name	File Type	Image Resolution	Digital Storage	freatures extractyerd?Y?N	which features	Digital copies reside	Paper map reside
Battle Hall, 3rd drawer map file	IGN	Just north of our research area commune and preliminary research area. Scale is not reproduced to exact size.	Map File 1	yes	cassininorth	JPEG	300 dpi	France Scans CDROM (copies Madry, Tickner, Jones)				

Figure 3 - Spreadsheet for Analog Cartographical data

Kommentiert [DH25]: This section is a bit confusing? Are there 2 spreadsheets or 1?

Kommentiert [DH26]: Do you mean fig. 1 or 2?

Kommentiert [DH27]: delete

Kommentiert [DH28]: Please add a space after MS

Kommentiert [DH29]: How do you archive these? As exports in MS Word file format?

Kommentiert [DH30]: I'd rewrite that part for clarity and because of the technical framework:

2. Documents in MS Word, Excel, or PowerPoint format, as well as those created with Google Docs Editors and OpenOffice.

Kommentiert [DH31]: Maybe set "Sensing Imaginery" lower case?

Kommentiert [DH32]: Rather use GNSS (Global Navigation Satellite System) as an umbrella term for clarity?

7. Data related to GNSS coordinates and geotagged photographs.

Kommentiert [DH33]: Please rewrite for clarity: 10. Websites and data in HTML format.

Kommentiert [DH34]: Lower case?

Data Preparation

The process of scanning and digitizing analog data, including 35 mm slides, field and aerial photographs, historical documents and manuscripts, paper maps, audio and video recordings, and other field and lab documents and archival records has been ongoing piecemeal for several years, conducted at both NCSU and UNC-CH. NCSU has, over several years, provided student workers who have assisted us in this process, and this significantly accelerated our work. This involved selecting the appropriate hardware and software tools to ensure that the resulting digital copies are of sufficient quality and met appropriate preservation standards. Fortunately, NCSU has a 35 mm slide scanner, which was used to scan hundreds of these slides. UNC-CH had for several years a high-resolution flatbed color scanner which was used to scan many maps and large drawings and our collection of 9x9 inch World War II aerial photos of the region.

Many paper maps, aerial photos, etc. had already been scanned over the years, but at low quality, and many of these were rescanned to meet modern archival standards. Additional searches were made for data, and we contacted many previous project participants, asking if they had data they would provide us. Several provided photos, 35 mm slides, and other files to us, which have been added to the workflow.

Data Formats

Given that our project has been going on for 45 years, we have had to deal with the fact that older project data existed in many different data formats, including many that are no longer commonly supported. These include 35 mm slides, CDs, VHS cassettes, Zip backup drives, Super 8 mm movie reels, audio cassettes, and more, as shown below in figure 3. Many sets of data have gone through several generations of formats over the decades, from analog to CD to Zip disks to external hard drives to the cloud. This is a problem common to many long-term research programs such as ours. A vital aspect of this work is the annotation of data with appropriate metadata and keywords such as the person who took the photo, the date, location, subject, etc. Sometimes, only one person knows these details, and some are dealing with data collected over 40 years ago, so this will be a lengthy process, but is vital to our project. Our intention is that individuals will be able to search the archives for specific locations, dates, subjects (mill ponds, forests, old roads, etc.) and individuals across all data types, and this will require detailed keywords and metadata, much of which does not yet exist. We recognize that this will be a prolonged activity.

Project Status

In the summer of 2023, the data will undergo a final quality assurance process and creation of missing keywords and metadata. We will then begin to undertake the process of entering each of the 16 buckets of data into the Dataverse system at UNC-CH and the Dryad system at NCSU, depending on whose data it is. As we are going to use both of these two institutional data repositories, a dedicated project archival access website is also under development, so that people can find project data regardless of its institutional archival location. This will be constructed using standard online finding aid structures. Metadata and keywords remain to be added to much of the data, in order to allow searches by location, date, type of data, etc. Final decisions will be made on data access and permissions before the archival 'system' goes live in the fall of 2023. Additionally, the existing document archives of Prof. Carole Crumley, already archived at UNC-CH, will be linked where possible. One outstanding concern is that the Odum Institute's archive only accepts digital data, and we do not know what will happen to all of the

paper maps, slides, and other data that we hold. We are seeking a proper repository for these, perhaps in partnership with the historical collections at one of our institutions' libraries.

Figure 4 - Data from our project in its many, outdated formats such as 35 mm slides, floppy drives, audio cassettes, Super 8 mm movie film, VHS videos, CD-ROMs, and Zip disks, none of which are supported today.

Results

This is an ongoing project and we do not yet have final results. But our situation is likely mirrored throughout the archaeological community. In our case, we followed a strategy based on digital archiving principles to guide our project steps, but we found it essential to adapt these standards to the specific challenges and needs of our extensive archive. We learned that digital archiving is not a linear process, but rather a cyclical and iterative process that requires constant evaluation and re-adjustment. Some of the challenges we faced included finding and using appropriate digital archiving tools, gaining access to some analog or digital data, and ensuring the appropriate security and privacy of our data. We crafted our steps in a non-rigid process that began with locating data, researching digital archive tools, preparing a first-order list, ascertaining privacy, and creating a collective catalog of our project. We anticipate that our next steps will include further securing and digitizing our data, uploading our data to the two servers, monitoring project progress, and disseminating our results to other researchers and community members. We are hopeful that these steps will contribute to the preservation and accessibility of our project in the long-term, understanding that this project will take time to further establish and maintain.

Large landscape projects, including archaeological, historical ecology, and related activities, often consist of multiple researchers from many different disciplines, institutions, and academic perspectives. Each participant brings with them large amounts of disparate raw, intermediate, and finished data in both analog and digital formats. Such projects can be very long-lived, continuing for several decades,

with both people, technologies, data formats, media, and archival perspectives coming and going over time.

Such projects can generate massive amounts of data, both digital and analog, which should be properly conserved and archived, and these should also be made available as a matter of course to the largest possible number of researchers, both within the project as well as beyond, after initial publication of results. Such archiving should be a part of all phases of work, including the initial planning and analysis work. Proper care must be taken for human subjects and sensitive archaeological site data, even after the specific project is ended.

Discussion

Digital archiving and related tools now exist that can be incorporated into new projects directly as they begin, so that they become another tool for the use of the researchers, but this requires specific knowledge and expertise which is traditionally outside of our fields, which is not always available to archaeological projects. Our project has struggled to keep track of the data held by various participants over the years, and significant data has also been lost. Data retained by former participants is rarely scanned or cataloged, nor does it contain metadata or is it accessible using keywords or geotags.

Kommentiert [DH35]: Maybe use “under certain circumstances” or something similar

Our strong advice to all is to please take digital archiving into consideration AT THE BEGINNING of your projects, and to incorporate digital archiving technologies and specialists in your work. Indeed, long-term data management strategies and plans are often now required components for many successful grant proposals at various public and private agencies. Key concepts such as digital data fixity, redundant backups, paradata, metadata, and appropriate keywords should become a part of all of our standard professional workflow, no matter what our particular discipline may be.

Once our data are uploaded and the archives ‘go live’, the project will continue to be monitored and adjusted as necessary, with the goal of ensuring that the data remains available for a broad scope of researchers over time. Finally, information on the process and results of the digital archiving project will be collected and distributed to other researchers to benefit their own research and digital archive development.

We hope that our experiences in seeking, well after the fact, to incorporate this important new capability into our work will be useful not only for our project and researchers in our ongoing work, but also for those who may come after us working in this region. We also hope that our experiences will be helpful to the larger community, to assist others to incorporate these tools into their future work as a matter of course.

Acknowledgements

We wish to acknowledge the participation of multiple people who have worked with us over the years. This includes Professor Carole Crumley, the numerous colleagues who collaborated with us on the project over the 45 years of its existence, and all of the collaborators in this work, past and present. We also acknowledge the North Carolina State University undergraduate students Hannah Taylor, Duncan Anderson, Angel Ngo, Madeleine Guyant, Daniela Agostini, Charlotte Bulkeley, and Kaitlyn Bohn, who assisted us in scanning and documenting much of the data that will be archived. We also recognize and thank the many people of our French study area who so graciously supported and assisted us over the

decades, and who shared their homes, fields, and archives with us. We also express our gratitude to the digital archiving experts who have so graciously assisted us, and the digital archiving staffs at UNC-CH and NCSU.

Data, scripts, code, and supplementary information availability

Data will be made available online at the University of North Carolina at Chapel Hill's Odum Institute Data Repository: https://odum.unc.edu/archive and at North Carolina State University's Dryad Data Repository at https://www.lib.ncsu.edu/do/data-management/dryad

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial or other conflicts of interest in relation to the content of the article.

Funding

This digital archiving work is primarily unfunded and is being conducted on our own. NC State University has provided funding for several undergraduate and graduate students in support of this project.

References

Aczel B, Szaszi B, Holcombe AO (2021) A billion-dollar donation: estimating the cost of researchers' time spent on peer review. *Research Integrity and Peer Review*, **6**, 14. https://doi.org/10.1186/s41073-021-00118-2

Broman, K. W., & Woo, K. H. (2018). Data Organization in Spreadsheets. *American Statistician*, *72*(1), 2–10. https://doi.org/10.1080/00031305.2017.1375989

Crumley, C. & B. Marquardt 1987 *Regional Dynamics: Burgundian Landscapes in Historical Perspective*. San Diego: Academic Press.

Copiello S (2020) Business as Usual with Article Processing Charges in the Transition towards OA Publishing: A Case Study Based on Elsevier. *Publications*, **8**, 3. https://doi.org/10.3390/publications8010003

Doyle H, Gass A, Kennison R (2004) Who Pays for Open Access? *PLoS Biology*, **2**, e105. https://doi.org/10.1371/journal.pbio.0020105

Dryad | Good data practices. (n.d.). Retrieved April 15, 2023, from https://datadryad.org/stash/best_practices

The FAIR Data Principles – FORCE11. (n.d.). Retrieved April 15, 2023, from https://force11.org/info/the-fair-data-principles/

Huvila, I. "Participatory archive: towards decentralized curation, radical user orientation, and broader contextualisation of records management". *Arch Sci* **8**, 15–36 (2008). https://doi.org/10.1007/s10502-008-9071-0

Madry, S., A. Westin, E. Jones editors. "Methods, Techniques, and Theory in Historical Ecology: The Collection, Analysis, and Integration of Complex Regional Data". 2023 Accepted for publication by the Swedish Diversity Centre Press, Uppsala, Sweden

Richards, J.D., Jakobsson, U., Novák, D., Štular, B. and Wright, H. 2021 “Digital Archiving in Archaeology: The State of the Art”. *Internet Archaeology* 58. https://doi.org/10.11141/ia.58.23

Wright, H. and Richards, J.D. 2018 “Reflections on collaborative archaeology and large-scale online research infrastructures”, *Journal of Field Archaeology 43*, supp1., S60-S67 https://doi.org/10.1080/00934690.2018.1511960