UNIVERSITY OF CALIFORNIA
Los Angeles

Seeking Context:
Archaeological Practices Surrounding
the Reuse of Spatial Information

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Library and Information Science
in Information Studies

by

Deidre Alyse Whitmore

2016
The reuse of spatial information is essential to archaeological research. In this study, sixteen semi-structured interviews were conducted with archaeologists from the Cotsen Institute of Archaeology at UCLA to document current reuse practices. These archaeologists seek spatial information throughout their research projects and frequently access potential resources through colleagues and government institutions that hold archaeological materials and reports from a region. Respondents identified a wide range of resources that contain spatial information and described gathering and joining together multiple types of these resources. The archaeologists described publications and grey literature as essential to reuse and would often extract spatial information from the text, figures, and supplementary materials. The time to process and
difficulties joining disparate datasets were cited as the most significant factors impacting reuse. Additionally, lack of documentation, incompatible levels of granularity, and concerns for accuracy impacted researchers’ ability to reuse spatial datasets. This study illustrates the need to incorporate information science skills and information professionals in the archaeological research community.
The thesis of Deidre Alyse Whitmore is approved.

Jonathan Furner
Willemina Z. Wendrich
Christine L. Borgman, Committee Chair

University of California, Los Angeles
2016
# Table of contents

List of figures ................................................................................................................................. vi
List of tables ................................................................................................................................... vi
Acknowledgments ............................................................................................................................ vii
I. Introduction .................................................................................................................................. 1
II. Statement of significance ........................................................................................................... 2
III. Literature review ....................................................................................................................... 2
   A. Scholarly communications ................................................................................................................... 4
      1. Data dissemination ............................................................................................................................ 5
      2. Open access ....................................................................................................................................... 6
   B. Data reuse ............................................................................................................................................. 7
      1. Data reuse practices in non-archaeological domains ........................................................................ 8
      2. Data reuse practices in archaeology ................................................................................................ 10
   C. Spatial information ............................................................................................................................. 11
   D. Research methods in information studies ........................................................................................... 14
IV. Definitions of key terms and concepts .................................................................................... 15
V. Research questions ................................................................................................................... 19
VI. Research methods ................................................................................................................... 20
   A. Participants ......................................................................................................................................... 21
   B. Design ............................................................................................................................................... 22
VII. Results ................................................................................................................................... 24
   A. Seeking spatial information ................................................................................................................ 24
   B. Identifying sources of spatial information .......................................................................................... 26
   C. Accessing spatial information ........................................................................................................... 31
   D. Processing spatial information for reuse ............................................................................................ 35
   E. Reusing spatial information ................................................................................................................ 37
VIII. Discussion ............................................................................................................................ 42
   A. Context for reuse ................................................................................................................................ 42
   B. Processing and reuse .......................................................................................................................... 45
IX. Conclusion .................................................................................................................................. 47
X. Appendix: Interview script and questions ................................................................................ 48
XI. Glossary ................................................................................................................................... 49
XII. Bibliography .......................................................................................................................... 50
List of figures

Figure 1: Workflow for joining multiple datasets from the same original collector..............41

List of tables

Table 1: Conditions motivating researchers to seek spatial information.........................25
Table 2: Sources of spatial information...........................................................................27
Table 3: Resources with spatial information....................................................................29
Table 4: Factors influencing ability to access sources of spatial information...............32
Table 5: Techniques used to process spatial information.................................................31
Table 6: Types of reuse defined by the study participants...............................................36
Table 7: Types of issues encountered by the participants...............................................39
Acknowledgments

First and foremost, I would like to thank the Cotsen Institute of Archaeology for generously supporting my research and providing a collaborative environment to explore archaeological data. I especially appreciate the archaeologists that participated in this study— their openness and thoughtful responses were invaluable. I would like to extend my gratitude to the members of my thesis committee for their extremely helpful guidance and feedback. Christine Borgman introduced the complexities of data and gave me the tools to investigate solutions to the challenges surrounding their use and reuse. Jonathan Furner provided constructive input and expertise throughout. I would especially like to thank Willeke Wendrich for her mentorship, enthusiasm, and guidance—I would not be where I am today without her. My heartfelt appreciation goes to Libbie Stephenson for encouraging me to pursue information studies and to Michelle Caswell for clearly and comprehensively helping me develop a research proposal. I would like to thank Dominic Polito for listening and providing entertaining analogies as I verbally worked through each component of this study. I am especially grateful to my parents and family for their constant support and encouragement—their perseverance is inspiring.
I. Introduction

Scholarship within archaeology is increasingly dependent on digital technologies. As data collection practices evolve, the ways in which information is managed and shared are also altered. The shift from research data and publications being primarily disseminated through paper media to an increasing dependence on digital platforms calls for the reexamination of data practices. I propose that this reexamination takes place within an information studies framework. The theoretical and practical experience information professionals have developed while addressing information-related practices in a variety of disciplines applies to the issues arising during this transition in archaeology. Information institutions frequently deal with transitional situations requiring simultaneous support for legacy behaviors and developing practices.1

Within the field of archaeology this shift is occurring amidst a lack of common standards in data collection and attempts to synthesize disparate datasets.2 Consequently, adoption of online publishing and data repositories geared towards the field has been sporadic and localized within sections of the research community. However, as use of these services increase and emphasis on data sharing and reuse grows within the field, the need for a better understanding of the current practices becomes more critical. This study addresses this need by examining the way in which archaeologists currently seek, access, and reuse spatial information.

1 Zimmerman, “Not by Metadata Alone,” 2.

II. Statement of significance

While questions of data sharing and usability appear across disciplines, the significance of data dissemination and reuse extends beyond basic research funding and scholarly practices in archaeology. In many archaeological contexts the methods employed to collect data actually destroy the context and ability for subsequent researchers to investigate the same setting. Physical scientists struggle with similar difficulties replicating the exact situation in which data was collected but often are able to study comparable phenomena in slightly different contexts. Archaeological excavations on the other hand disrupt the stratigraphic deposits that provide insight into the data’s context. Methods of mitigating these disruptive techniques are employed through intensive survey and reconnaissance of an area prior to excavation, incorporation of experimental archaeological studies, and costly conservation practices of artifacts and architectural elements exposed during excavations. However, many of these techniques rely on or are used in conjunction with excavation data. This practice makes preserving the data—and enough metadata and documentation to ensure practical access and reuse—highly significant to the archaeological research community. While understanding the context of reuse for all archaeological data types is important, examining the practices surrounding the reuse of spatial data in particular provides insight into a type of data central to the domain. The findings presented in the results of this study can be applied in the ongoing efforts to create sustainable data preservation and dissemination services for archaeologists.

III. Literature review

Over the past two years the recent changes in the White House Office of Science and Technology research policies have spurred a collection of studies evaluating the sharing and
reuse of research data. Sarah Whitcher Kansa, the Executive Director of the Alexandria Archive Institute, and Eric Kansa, Lead Technology Developer for Open Context, have conducted a series of research studies on archaeological data reuse. In particular, Kansa; Kansa and Kansa; and Kansa, Kansa and Arbuckle have explored the impact of open access and data sharing on the field and archaeologists. While there is no literature specifically on the curation, dissemination, and reuse of spatial information in archaeology, researchers in other disciplines and at academic libraries have investigated and published on the reuse of spatial data. My study aims to bridge this gap between the current knowledge of best practices surrounding spatial datasets and contemporary research on data reuse in archaeology within an information science framework.

The literature review describes the publications resulting from the collection of studies on scholarly information, dissemination, and reuse. The first section explores changes in scholarly communications in archaeology, data dissemination, and how the recent emphasis on open access policies has led to research on ways in which these policies may impact archaeological data practices. The second section details the literature pertaining to data reuse, both within and outside of archaeological domains. The third section addresses scholarly practices surrounding spatial information and current works on the preservation standards and access infrastructure for this data type. The last section places the data collection methods to be employed in this study within the context of other information studies research methods.

3 Kansa, “Openness and Archaeology’s Information Ecosystem”; Kansa and Kansa, “We All Know That a 14 Is a Sheep”; Kansa, Kansa, and Arbuckle, “Publishing and Pushing.”
A. Scholarly communications

Research practices, communications, and access to resources increasingly operate within a digital scholarly infrastructure. Information is critical to this infrastructure and data are seen as outputs that can be reused in subsequent research. However, data and information are defined differently across disciplines, and within archaeology, data are collected from a complex network of sources, activities, and expertise. Similarly, attempts to define access to information introduce additional contextual variances and interpretations. In studies examining access to information, often through or within networked infrastructure, researchers either define access in broad terms covering information providers and services, locating, retrieving, and using the information or focus on specific aspects within the concept. One such aspect, usability, is particularly relevant to the reuse of spatial information within archaeology. Increasing emphasis on sharing information and providing open access to scholarly outputs—both publications and the associated data—is influencing the development of data services available to archaeologists.


6 Lynch and Office of Technology Assessment, Accessibility and Integrity of Networked Information Collections, 7; Borgman, From Gutenberg to the Global Information Infrastructure Access to Information in the Networked World, 57.

1. Data dissemination

The advantages and disadvantages of existing modes of data dissemination within archaeology are succinctly summarized in Table 1 in Kansa and Kansa.8 As key participants and founders of Open Context, the authors have an interest in the “Integrated Data Publishing” mode of dissemination. They illustrate the benefits offered by integrated data publication through a detailed description of the workflow and actions taken to make the data accessible, sustainable, and reusable. However, Borgman illustrates how the term “data publication” as more broadly used in scholarly communications, introduces a series of issues.9 While the datasets disseminated through Open Context undergo a custom form of peer review developed for data, datasets incorporated into or associated with journal articles and books rarely receive independent review.10 Additionally, data and scholarly publications are often only associated through the inclusion of a link to the archive or repository holding the dataset and researchers rely on the context of the written text to understand the data.11

In “Publishing and Pushing” Kansa et al argue that the incorporation of a data “push” approach, calling on the term used by software and code developers to describe the release of updated versions, better takes advantage of the opportunities surrounding digital data publication.12 Instead of the finality associated with a publication, they argue that incorporating

---

8 Kansa and Kansa, “We All Know That a 14 Is a Sheep,” 5.
9 Borgman, Big Data, Little Data, No Data, 48.
10 Ibid.
11 Ibid.
12 Kansa, Kansa, and Arbuckle, “Publishing and Pushing.”
push versioning into data sharing allows feedback from the actual reuse of data to enhance the value and applicability of the dataset.

2. Open access

Open access has been presented as a method to enhance and foster future research; however, the details of how increased sharing will accomplish these goals are not yet clear. Peter Suber asserts that all content, including data, can be open access and the ability to more quickly release content for testing and review benefits research.13 Kansa writes about the potential for open access to research data to alter the influences on career trajectories in academia via the incorporation of data publication metrics in professional evaluations.14 Borgman systematically describes the context of this proposition across academic domains and provides an explanation of the obstacles impeding this incorporation currently.15 However, by situating the discussion of open access within an archaeological context, Kansa calls on the unsustainable funding and reliance on public support inherent in archaeology and advocates for using these characteristics to spur the various stakeholders into uniting through open initiatives.16 He argues that this unification is dependent on archaeologists developing better digital information skills which facilitate the process of publishing functional data. Kansa briefly presents some examples of successes but does not delve into the specifics or actually define success. Subsequent articles by

13 Suber, Open Access, 98.
14 Kansa, “Openness and Archaeology’s Information Ecosystem.”
15 Borgman, Big Data, Little Data, No Data.
Kansa and colleagues rectify these oversights through more systematic examinations of current knowledge infrastructures.

**B. Data reuse**

This idea of reuse is key to data sharing and open access endeavors if the research potential sought by the recent policies changes is to be realized. As data archiving and sharing practices gain momentum in archaeology, researchers within this discipline are attempting to identify the processes and factors surrounding data reuse. Qualitative research of data reuse in disciplines outside of archaeology can provide valuable insights into the process of finding, assessing, and processing data for a secondary use. These types of investigations are not limited to a single domain and, expectedly, how use and, subsequently, reuse is defined varies across studies and disciplines. Borgman summarizes the struggle to define “information use” across research communities and how this lack of agreement impacts the ability to study data reuse practices. 17 In her study on data reuse in ecology, Zimmerman provides an explicit, functional definition of reuse in her research on ecologists’ reuse practices. 18 She defines reuse as “the use of data collected for one purpose to study a new problem” and includes the term “secondary use” to mean the same. 19 This definition provides a working framework for investigating reuse practices within archaeology.

---


18 Zimmerman, “New Knowledge from Old Data.”

19 Ibid., 4.
1. Data reuse practices in non-archaeological domains

Research on data reuse in non-archaeological domains such as ecology provides insight into common practices across communities. Zimmerman conducted semi-structured interviews with ecologists that were identified as having reused data and subsequently published articles incorporating results from this reuse. The participants were selected through the examination of two prominent ecology journals for recent articles fitting this criteria so that the data reuse experiences would be relatively recent—having taken place at most three years prior to being interviewed. Zimmerman investigated the process of evaluating and using data collected by others, as well as the influence of associated metadata and surrounding contextual information. Zimmerman’s findings reveal traits specific to this research community such as the reliance on field experience and veteran scientists in the selection and evaluation stages of data reuse. However, correlations in the field conditions and data collection methods between ecologists and archaeologists may produce similar correlations in data reuse practices. The research design and methods of data collection employed in Zimmerman’s study influenced the structure and methods employed in this study.

Similarly, Faniel and Jacobsen situated their research in detailed descriptions of the community studied—earthquake engineers—and acknowledged the importance of understanding the research community and their data practices in the quest for a better understanding of scientific data reuse. Faniel and Jacobsen also conducted semi-structured interviews and framed the questions to address how the data were found, evaluated, and processed. Contrary to the Zimmerman study however, Faniel and Jacobsen did not preselect researchers that were known

---

20 Faniel and Jacobsen, “Reusing Scientific Data.”
to have reused datasets but rather added that as an interview question. While the sample population was too small to extract information regarding the prevalence of data reuse in the field, the qualitative results displayed similarities in reuse strategies. The three sections identified from Faniel and Jacobsen’s research—identifying relevance of the data, ensuring understandability, and judging trustworthiness—influenced the categorization and structure of the interview questions employed in this study although adjustments were made to address the features of geospatial and archaeological data collection and context.

Recent publications pertaining to the barriers hindering sharing and reusing of archaeological data describe the complexities surrounding the issue. As seen in the studies on different domains and research communities, narrowing the focus of the research question to obtain detailed information about a particular subset of the knowledge infrastructure provides valuable insight into some of these complexities. A selection of analyses have focused on the reuse of certain types of archaeological datasets within artificially controlled settings. Atici et al conducted research on analytical processing of zooarchaeological data from the 1960-70s.\textsuperscript{21} The provenance and history of the dataset, due to its multiple migrations across media (from the original report to punch cards and later Excel spreadsheets), is vital to the context of this experiment. Their findings describe the influence these migrations had on the process each researcher chose to implement during analyses. In addition to the descriptive documentation of each step the researchers took while evaluating, processing, and analyzing the dataset, Atici et al recorded the discussions which took place after the work. This publication provides an indispensable demonstration of the potential for scientific replicability as datasets are reused

\textsuperscript{21} Atici et al., “Other Peoples Data.”
within archaeological contexts. The destructive nature of much field research in archaeology obstructs the ability to reproduce most investigations. However, as Atici et al illustrate, data reuse offers methods of comparative replicability. As the literature surrounding data reuse within archaeology accumulates and the process surrounding this reuse is better understood, guides and standards can be developed to assist in optimizing the potential for reusable data.

2. Data reuse practices in archaeology

Faniel et al have begun to address the development of these guides and standards through their research on the needs of archaeologists reusing data collected by others.\(^2^2\) This study is situated in the current issues surrounding the collection and management of archaeological data. As seen above in the publications investigating data sharing and access, data repositories and documentation standards have yet to be incorporated into the regular research practices of archaeologists throughout the field. While the general acceptance of the need for change is growing, the distributed, and often conflicting, interests and lack of cohesion across the domain have prevented widespread engagement. Because of this, it is difficult to implement standards for depositing the data, much less standards that address the needs of the potential data re-user. However, Faniel et al gathered informative descriptions of the issues researchers have faced while reusing data and explanations about the actions they took to address these issues. The frequent lack of context presented the most pressing issue for researchers and the lack of procedural documentation was found to be the second greatest limiting factor. Faniel et al discussed these findings in the context of current practices employed by archaeological data

repositories. While the standards enforced at the repository level address the needs of the researchers attempting to reuse datasets to some degree, Faniel et al argued that guides and standards need to be in place prior to the initial collection of the data instead of at the time of deposit. Faniel et al propose the incorporation of data management and information science skills into archaeological training in concordance with the proposal by Kansa to develop better digital data management skillsets within archaeology.23 These studies show that collecting informative records of the data throughout the entire data lifecycle and preserving that information in association with the data is key to higher quality datasets that will, in turn, facilitate reuse.

C. Spatial information

The Library of Congress (LOC) format sustainability website in association with the Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP) researched and published information about file formats best suited for long-term preservation of specific data types.24 Spatial formats were added to the LOC sustainability website in 2012 to support the growing number of researchers across disciplines that are relying on spatial information. Hoebelheinrich details the strategies and methods the LOC employed during the evaluation of file formats, factors surrounding the spatial formats, and contextual information for the use of the website itself.25 The LOC format sustainability website provides an introduction to the complexities of archiving spatial data so that it remains functionally accessible for reuse in the future. However, Rob Kitchin describes the increasing tendency

23 Kansa, “Openness and Archaeology’s Information Ecosystem.”
25 Hoebelheinrich N.J, “An Aid to Analyzing the Sustainability of Commonly Used Geospatial Formats.”
among researchers to exhaustively collect and record spatial information, with minimal thought to maintenance or preservation of these datasets, much less the complexities surrounding future reuse.\footnote{Kitchin, \textit{The Data Revolution}, 44.} In some cases, these data are housed within a formal data infrastructure which enforces certain processing and structural attributes that facilitate reuse; however, the bulk of data, now and historically, have been stored in informal “data holdings.”\footnote{Ibid., 38.}

In addition to the LOC contributions, Shaon and Woolf reviewed the requirements for long-term sustainability of spatial environmental data.\footnote{Long-Term Preservation for Spatial Data Infrastructures.} This study is framed in relationship with Spatial Data Infrastructures (SDIs) and focuses on the ability of applications such as Infrastructure for Spatial Information in the European Community (INSPIRE) to continue to read file formats and metadata.\footnote{“INSPIRE > WELCOME TO INSPIRE.”} As archaeological data repositories develop and more efficient methods of ingesting spatial data are explored, these providers are likely to call upon the experiences of environmental science applications such as INSPIRE. Shaon and Woolf evaluated the models defined in INSPIRE against the Open Archival Information System (OAIS) Reference model.\footnote{Consultative Committee for Space Data Systems, \textit{Reference Model for an Open Archival Information System (OAIS)}.} The OAIS Reference model has become the gold standard for data repositories and established archaeological data repositories such as Archaeology Data Service (ADS) and tDAR (The Digital Archaeological Record) evaluate their procedures and
infrastructure against this model. While Shaon and Woolf describe a process in which SDI applications like INSPIRE can maintain the ability to access historical or dated spatial data through efficient documentation of digital migrations, little is written about the need for documentation and records of the original data collection despite its impact on decisions made during these migrations.

Wilson addresses this need for digital preservation to be incorporated into the practices of not only the preservation specialists and archivists but also the members of the research communities. His evaluation of the Preservation Metadata Maintenance Activity (PREMIS) data dictionary revealed the role preservation of contextual metadata, provenance tracking, and management or administrative metadata plays in the functional reuse of data over time, regardless of overlap among these metadata categories. Wilson argues that archivists need to work closely with the domain experts in order to ensure the integrity of the various categories of metadata to ensure the data preserved is accessible, usable, and authenticated. Borgman supplements this argument in a series of case studies pertaining to the humanities, social sciences, and physical science domains. These case studies highlight the role domain expertise plays in data practices from collection to preservation and reuse. The evaluation of these practices throughout their life cycles provide detailed, informative comparisons of the techniques.

31 The Digital Archaeological Record (tDAR), “Getting Familiar with tDAR - The Digital Archaeological Record - Digital Antiquity Wiki (Documentation)”; “Archaeology Data Service.”

32 Wilson, “How Much Is Enough.”

33 “PREMIS: Preservation Metadata Maintenance Activity (Library of Congress).”

34 Borgman, Big Data, Little Data, No Data.
and impact the research community has on the likelihood of producing and preserving a dataset that can be reused by other members of this community or scholars from other disciplines.

A variety of aggregations of geospatial data are stored in data catalogs housed at academic libraries. These catalogs serve as portals to the datasets which can be accessed by researchers from various disciplines. Kollen et al reported on the findings of the Spatial Data Subcommittee’s investigation of spatial data catalogs for the American Library Association (ALA) Map and Geospatial Information Round Table (MAGIRT) Geographic Technologies Committee in 2010. The subcommittee interviewed data catalog managers at academic libraries that had collections of geospatial data in their catalogs. The interviews were designed to obtain a broad overview of the tools and practices utilized in handling this type of data. Because the technical and human resources available at an academic library are likely to be comparable to the resources available to archaeological data repositories and publishing platforms the information gathered from these interviews can facilitate the development of similar tools and practices in archaeological repositories. Data archivists at these institutions work one-on-one with researchers depositing data or seeking to reuse data. The findings from this study support the personalization of tools and resources to specific contexts while attempting to alleviate the total extent of these one-on-one interactions.

D. Research methods in information studies

The primary method proposed to gather data on current practices in archaeology surrounding the reuse of geospatial data is based on prior studies conducted by information

---

35 Kollen et al., “Geospatial Data Catalogs”; “Map & Geospatial Information Round Table (MAGIRT) | Map & Geospatial Information Round Table (MAGIRT).”
studies scholars. While at the University of Michigan, School of Information, Ann Zimmerman conducted semi-structured, in-depth interviews for her research on data reuse by ecologists.\textsuperscript{36} Her interview questions covered the processes around data reuse from initial location of data to the decision to reuse. However, the work by Faniel et al and Atici et al influenced the decision to extend the interview in this study to investigate the processes surrounding the reuse itself in order to gain a more complete model of the current practices.\textsuperscript{37} The common themes among research into processes surrounding data reuse, whether data from an archaeological or other context, address issues that arise throughout the life cycle of the data. The qualitative studies above frequently found the need for better information management training and techniques at earlier stages in the research process. Spatial data is a significant source of information for archaeologists and, with the increasing prevalence of Global Positioning System (GPS) devices and geographic information system (GIS) software, a type of record likely to be shared and reused in subsequent research. This study addresses the need to better understand how archaeologists currently reuse these records so that standards and guides to facilitate the collection and preservation of spatial data can be developed.

IV. Definitions of key terms and concepts

Few studies formally define data and many of the definitions that do exist are circular or overly simplified.\textsuperscript{38} Borgman explores in depth the variety of definitions employed by

\textsuperscript{36} Zimmerman, “Not by Metadata Alone”; Zimmerman, “New Knowledge from Old Data.”


\textsuperscript{38} Borgman, \textit{Big Data, Little Data, No Data}. 
researchers when discussing data. After breaking these definitions out into a series of categories she presents a concrete, explicit statement defining data within that publication which I have chosen to apply in this study as well. Borgman defines data as “entities used as evidence of phenomena for the purposes of research or scholarship.” The term entities is especially important within this context because it implies a materiality to data, conveying the idea that data can be moved around, transferred and manipulated. Contextualizing their association within phenomena is particularly important to spatial data within archaeology because it captions the intrinsic temporal nature of spatial information. Geophysical locations change over time and any collected data varies greatly depending on the instruments, their calibration, and the user. How the term data is used varies within archaeology, therefore the broader term “information” was initially employed in interactions with the participants and during the interviews. Therefore, the term spatial information is used within this study to describe entities serving as evidence of geographic locations associated with phenomena.

Metadata within this study are characterized as information associated with data entities that provide the context of the phenomena, research, and any alterations that have occurred—a definition consolidated from the National Information Standards Organization (NISO) report on metadata, the Archaeology Data Service (ADS) Guides to Good Practice, and the Digital Archaeological Record (tDAR) documentation pages. The implementation of metadata varies greatly across archaeological projects. Archaeological data services have begun collaborating on

39 Ibid. pg 29

40 National Information Standards Organization (U.S.), Understanding Metadata; Archaeology Data Service (ADS), “Guides to Good Practice: CreateData_1-2”; The Digital Archaeological Record (tDAR), “Getting Familiar with tDAR - The Digital Archaeological Record - Digital Antiquity Wiki (Documentation).”
the construction of a metadata schema for archaeological datasets. The relationship of metadata standards created and employed by individuals and projects and the schema developed by these institutions will be investigated within this study. Although, schema provide technical methods of managing, searching, and displaying data on digital platforms these technological requirements most likely do not address researchers’ specific scholarly needs or capture the context of their research.

The concept of data reuse has many facets but is generally understood to involve the secondary use of data whereas the primary or initial use of data involves use by the original collector for the purposes of research or scholarship. For this study, following Zimmerman’s definition, data reuse is defined as any subsequent use of data in a context other than that of the original collection. This may involve the application of the data in a new research study within the same project or by entirely new set of individuals. Potential types of reuse may be hypothesized but are ultimately unknown at the time of collection and processing. Thus there is the possibility that the documentation and metadata associated with a dataset may not be applicable to the needs of the researchers reusing the data. During the interviews I did not impose this definition on the participant but rather attempted to extract their working definition of reuse in order to broaden our understanding of the impressions surrounding reuse in archaeology currently.

Methods of disseminating data are broadly described as either sharing or publishing data. These types of dissemination hold key differences, however, and the terms will be used within

---

41 Faniel and Jacobsen, “Reusing Scientific Data.”

42 Zimmerman, “New Knowledge from Old Data,” 634.
this study according to the following definitions. Data sharing involves the transfer of data from one individual to another. This transfer may be facilitated by a third party or service such as a repository or archive but does not require this intermediary. Sharing does not imply any alterations have been made to the data such as cleaning or review prior to the transfer. Data publication on the other hand, introduces actions taken prior to transfer including an editorial review of the content and a system of verification, whether the data are published within a written text or as standalone datasets online. By definition, data publication introduces a third party—the editor(s)—into the process. However, as discussed in the literature review, it may not be apparent to the researcher how that third party interacted and reviewed the data themselves. Explicitly defining these terms is linked to the importance of distinguishing between these two forms of dissemination and their use within archaeological research. The interviews conducted in this study aimed to gather descriptive information about both of these dissemination techniques, the participants’ perspectives on both, and any biases or leaning towards one technique over the other.

In addition to dissemination techniques describing the supply of information, terms associated with the demand for and retrieval of information are significant to interpreting the findings. Within this study, information seeking is defined as the “conscious effort to acquire information in response to a need or gap in knowledge” and that knowledge may be actively or passively gained.43 As discussed in the literature review, access is multifaceted and often applied to a variety of materials and contexts. Distinguishing between the participant’s ability to physically acquire an item containing spatial information and the usability of the item, its

content, or the information itself is significant to understanding the factors impacting reuse of spatial information in archaeology. For the purposes of this study, the term access is used to describe the researcher’s ability to obtain a resource containing spatial information. Following this, usability addresses that resource’s format, its usefulness, and the requisite knowledge and skills associated with the use of the resource’s content.⁴⁴

V. Research questions

The findings presented in this study will facilitate the development of a framework in which data archiving and publishing services can structure their metadata, policies, and procedures to support the preservation of data so that it can remain functional and reusable. The research questions guiding this study concern three aspects of data reuse: the need for information, the process of seeking and accessing the information, and the processing of the information.

1. When, and under what conditions, do archaeologists seek spatial information?
2. How do archaeologists currently identify potential sources of spatial information?
3. How do archaeologists access previously collected spatial information and what factors may impact their ability or inability to access the information?
4. What steps and processes do archaeologists employ to reuse spatial data and what challenges arise?

⁴⁴ Borgman, From Gutenberg to the Global Information Infrastructure Access to Information in the Networked World, 57.
Specifically, awareness of the current avenues of discovering and selecting data for reuse can inform services on appropriate approaches to discoverability and metadata prioritization. Identifying the circumstances surrounding failed searches or the inability to access information can illustrate areas of need. In addition, the detailed descriptions of reuse procedures can highlight what types of documentation and metadata are considered essential to reuse. While these avenues and procedures varied widely, this study extracted commonalities and themes among the interview data that can be applied to the creation of a set of standards and guides to best practices.

VI. Research methods

This study utilized a semi-structured interview format to address the research questions posed above. Previous reuse studies have employed a variety of techniques to enlist participants. Sixteen interview participants were enlisted from graduate students, postdoctoral fellows, and faculty at the Cotsen Institute of Archaeology at UCLA. Potential participants were contacted by email and given a brief background on the project aims and interview process. The interviews ranged from 40 minutes to 1 hour and were conducted in-person with the exception of one videoconferencing session. Each interview progressed through the questions listed in the Appendix while remaining open to additional information offered by the participant throughout the process. The interview questions were structured to gather detailed information on the methods archaeologists have recently used to find, access, assess, and utilize previously collected data. During the interview I aimed to extract each participant’s working definition of data, metadata, and data reuse. Specific details pertaining to any issues that arose during the reuse are incorporated throughout the interview. Each interview was recorded and saved in a digital
format. The recordings were transcribed and coded through NVivo, a qualitative data analysis application.

The data resulting from these interviews were analyzed according to Hsieh and Shannon’s “conventional content analysis” approach. Each interview was transcribed and preliminary themes were noted during transcription. I consolidated and organized these themes in relation to the four research questions and during the first coding cycle added additional codes to this framework. The first cycle consisted primarily of In Vivo coding and was highly granular. During the second coding cycle many categories were lumped together and larger portions of the transcription were tagged with more descriptive codes in order to re-contextualize the extracted concepts. The data were evaluated for similarities and differences across projects and participants, sequences in the codes, and correlations among the coded categories and subcategories.

A. Participants

The participant group was comprised of graduate students (n=6), post-doctoral and visiting scholars (n=4), and affiliated faculty (n=6). Each participant was asked to focus on a recent research project in which spatial information that was previously collected played a role in the research. Six of the participants included descriptions of a second—and in one case a third—project that influenced their perspective and processes surrounding spatial data. The resulting 23 projects discussed during the interviews cover archaeological sites in Africa, Asia, Europe, North

---

America, and South America. Of the archaeological sites investigated, all had previously been studied by other researchers to some extent.

B. Design

While archaeology is frequently practiced across international borders, political and economic factors surrounding the establishment and maintenance of these services have produced data publication and archiving platforms that are geared primarily towards their geographic constituents. This bias and practice is recognized and the findings from the interview interpretations are applied only within an American archaeological context and specifically within the research community at the Cotsen Institute of Archaeology.

This study targets archaeologists operating in an academic setting; however, contract archaeologists, government archaeologists, and cultural resources management scholars similarly seek, assess, and reuse spatial information. While it would be interesting to compare the data reuse practices of private sector and government archaeologists with academic researchers this is beyond the scope of this project therefore questions relating to this relationship have not been included in the interview outline.

This study focuses solely on the reuse of spatial information within archaeological research. These parameters have been established to concentrate the research methods on the collection of data pertaining to the reuse of a significant data type in archaeology. Spatial data is key to constructing physical as well as chronological context in archaeological studies and a vital

---

46 UK: Archaeology Data Service (ADS), Netherlands: Data Archiving and Networked Services (DANS), US: the Digital Archaeological Record (tDAR) and OpenContext, Italy: Metodologie Applicate alla Predittivita del potenziale archeologico (MAPPA)
component to most research projects. The evolution of recording, processing, and visualizing
techniques has left a range of spatial dataset file types and formats including advancements in
capturing and rendering three dimensional datasets. By focusing on the use and reuse of these
data this study gathers detailed perspectives on one of the most common and significant data
types in the research community. While the findings cannot be generalized to other types of
archaeological data or the reuse of spatial data in other disciplines they can directly influence the
development of dissemination platforms providing services for geospatial datasets.

Participants interviewed in this study consisted of English-speaking American
archaeologists affiliated with or studying at the Cotsen Institute of Archaeology at UCLA. This
is due to the sampling methods utilized in recruiting participants and in part to barriers to
sufficiently interpreting and understanding other languages. While this parameter narrows the
study to a subset of the research community similar parallels are found in the publishing and
repository services. Current platforms typically focus their resources and the efforts of their staff
on research data generated by national or regional archaeologists.47

Several biases can be associated with information gathered from interviews including
self-reporting biases by the participant. The ability of the researcher to recall the specifics of a
task they may perform routinely or infrequently could introduce loss of steps or actions in their
descriptions. The use of another colleague’s data may present sensitive situations and an
unwillingness to objectively describe the documentation and format of the data they received for

47 For example the Archaeology Data Service in the UK prioritizes data collections from UK
archaeologists: http://archaeologydataservice.ac.uk/advice/collectionsPolicy#section-collectionsPolicy-2.1.ScopeOfCollections
fear of harming a collaborator. All participants were assured of the confidentiality of their responses at the beginning of the interview and as needed throughout.

VII. Results

The results are organized by research question. The first section explores the initial reasons the researchers sought spatial information for the 23 projects described in the interviews. The second section presents the types of sources containing spatial information and where they are located, as well as the techniques employed by the participants to identify potential sources. In the third section I describe the ways in which the participants attempted to gain access to sources of spatial information and the variety of factors that impacted their successes and failures. The fourth section examines techniques that archaeologists employ to process spatial information and the final section delves into the issues that impacted each participant’s ability to reuse the spatial information they obtained.

A. Seeking spatial information

The conditions that initially prompted researchers to seek out spatial information about their sites were divided into five categories: new site, publication, further analysis, expanding study area, and synthesis. These categories were derived from common motivations the participants described as influencing their decision to seek out previously collected spatial information and in some cases, the participants described multiple overlapping motivations.

Most frequently the researchers sought spatial information about a site or collection of sites in which they were planning to conduct analyses that had not previously been performed (Table 1). Of the seventeen projects in the further analysis category, eight were initiated in order to study a type of material that had been collected during an excavation or survey but not examined.
The time between the initial collection of these materials and the participants’ analyses ranged from under a year to centuries later. Additionally, ten projects were utilizing new techniques or technologies in order to conduct analyses that weren’t previously possible. In three cases, the researcher was taking a new approach to the data analysis—either by applying a new theoretical framework in her study or by altering the unit of analysis.

Six of the projects were started in order to expand the study area to include geographic sections that had not previously been investigated while seven were synthesizing data from a variety of archaeological sites and studies. One researcher was establishing a new project at a site that had been surveyed but not formally published in any academic journals or books. Another researcher was seeking spatial information as part of an effort to analyze and publish work that had been undertaken by the original investigators over 30 years ago.

Table 1: Conditions motivating researchers to seek spatial information

<table>
<thead>
<tr>
<th>Condition</th>
<th>Definition</th>
<th>Count of projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Further analysis</td>
<td>Researcher is studying the site(s) in order to perform additional analyses or utilize new techniques and/or technologies in the analyses. Examples: looking at from a different perspective; employing new methods, techniques, and/or technologies; performing analyses on a material not previously studied</td>
<td>17</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Researcher is synthesizing information from more than one study area. Example: comparing ceramic vessels collected from multiple sites that had been excavated by multiple investigators.</td>
<td>7</td>
</tr>
<tr>
<td>Expanding study area</td>
<td>Researcher is extending the study area beyond what had previously been investigated (geographically). Example: excavating a cemetery associated with an ancient city that had previously been excavated.</td>
<td>6</td>
</tr>
<tr>
<td>New site/project</td>
<td>Research site is a new study area to the participant. Example: establishing an excavation project at a site that had previously been surveyed by another investigator.</td>
<td>1</td>
</tr>
<tr>
<td>Publish</td>
<td>Researcher is seeking spatial information about the site(s) for the purpose of publication. The results of previous work at the site(s) had never or only minimally been published by the original data collectors. Example: original excavator of a site unable to find funding to complete analyses and publish findings.</td>
<td>1</td>
</tr>
</tbody>
</table>
The interviews revealed that the seeking process continued throughout the life of the project, although efforts to track down spatial data were particularly focused at the start of the project and prior to any field work. For example, three of the researchers described seeking maps specifically to satisfy permit requirements associated with defining geographic areas of study in their applications. They explained that the government institutions that issue archaeological permits require the researcher to submit a map with the proposed study area outlined or indicated in some way. Spatial information from previous archaeological studies influenced where and how the participants determined their study areas.

Individuals that were brought onto projects in order to conduct further analyses frequently discussed being provided with a dataset or collection and then seeking out additional spatial information in order to reconstruct the context from which these data originated. These reconstructions covered both understanding the archaeological context in which the remains were uncovered and the other researchers’ survey or excavations methodologies.

**B. Identifying sources of spatial information**

Participants identified a number of sources they searched or contacted while seeking spatial information (Table 2). While the 23 projects span four continents and an assortment of geographic regions within each, every participant discussed regional practices concerning the deposit of materials and often reports to local government institutions. Policies concerning the deposit of materials and reports were often explicitly stated in the permits or contracts the archaeologist obtained in order to conduct their research. In addition to holding the material

---

collections and seasonal reports, some participants stated that these government institutions would also register sites and maintain a “site list” or registry of site records for previously identified archaeological sites.

Table 2: Sources of spatial information

<table>
<thead>
<tr>
<th>Type</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colleagues</td>
<td>15 / 19</td>
</tr>
<tr>
<td>Company</td>
<td>3 / 3</td>
</tr>
<tr>
<td>Data repositories</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Government institutions</td>
<td>10 / 11</td>
</tr>
<tr>
<td>Libraries</td>
<td>10 / 11</td>
</tr>
<tr>
<td>Personal libraries</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Private museums and archives</td>
<td>4 / 5</td>
</tr>
<tr>
<td>Project websites</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Research institutions</td>
<td>5 / 5</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 / 2</td>
</tr>
</tbody>
</table>

Nearly all of the participants contacted colleagues or organizations such as those government institutions in an effort to identify and obtain sources of spatial data. Frequently, participants described reaching out to colleagues for copies of publications, unpublished reports, and even the colleague’s field notes or datasets. Most researchers contacted colleagues they had worked with previously or knew on a professional level or in cases where this was not possible, often a facilitator or intermediary contacted the data holder on behalf of the seeker.

The researchers would also conduct extensive bibliographic searches to identify any previous studies or individuals that may have information about the site(s). Half of the researchers describing traveling to the government institutions holding archaeological collections in order to search and review the records. Four participants explained that this entailed examining the location that was written on individual bags or tags—often the only place spatial
information was recorded. These four participants explained that the policies in place at the time of the deposit impacted the format and state of the metadata associated with the collections and while older collections rarely had a comprehensive catalog or listing of materials, recent changes to the policies were improving the conditions. However, these participants were attempting to extract spatial information from collections that had been deposited prior to these changes and subsequently were faced with minimal metadata and the time consuming task of recreating records in a more usable format: “if you’re looking at a collection that’s done 30 years ago they’re not that detailed” (Interview 7, graduate student).

While many of the participants had heard of online data repositories and publishing services, only one researcher had actually searched one. In this particular case, the researcher was searching tDAR for data from his own project that had been deposited by other members of the project. One researcher in particular expressed interest in the potential for such repositories:

No, but I know their existence. I just have to educate myself more on these things. This whole thing that’s kind of interesting, where you release, and that you can submit your archaeological data and it is going to be peer reviewed and of course with controlled vocabulary and standards being applied and then it takes and it actually counts as a real publication (Interview 9, faculty).

Many stated that they “had been meaning to” or “I’ve heard a lot about but I haven’t done anything with it” (Interview 15, visiting scholar; Interview 6, graduate student). Faculty, visiting scholars, and graduate students all conducted online searches to locate relevant publications; however, graduate students were more likely to conduct online searches with the aim of finding datasets shared through project websites, government agencies or academic institutions.
An expansive variety of resources containing spatial information were described by the participants (Table 3). The resources utilized by the researchers ranged from conventional spatial publications such as topographic maps and digital elevation models to more generalized descriptions including ethnohistorical accounts of land divisions in the past and correspondence between an excavator and his funder. Publications were cited most often as a source of spatial information and in addition to reading through location information in the text, the participants often described extracting a variety of other types of spatial data including photographs, sketch maps, and GPS coordinates from these publications. In most cases the participants obtained a physical copy of the publication, scanned the page(s) with spatial information and the cropped out the image or map. The participant or an associate would then incorporate the cropped item into a geospatial database or set of digital reference files. In two cases the researchers found GPS coordinates in captions and were able to add these coordinates to their own spatial databases.
<table>
<thead>
<tr>
<th>Type</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerial photographs</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Artifact or feature records</td>
<td>6 / 9</td>
</tr>
<tr>
<td>Bags and tags</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Catalogs</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Correspondence</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Cross sections</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Daybooks</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Descriptions</td>
<td>9 / 11</td>
</tr>
<tr>
<td>Digital elevation models</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Drawings</td>
<td>6 / 7</td>
</tr>
<tr>
<td>Ethnohistorical accounts</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Field notes</td>
<td>8 / 10</td>
</tr>
<tr>
<td>GPS coordinates</td>
<td>6 / 6</td>
</tr>
<tr>
<td>Grey literature and reports</td>
<td>11 / 11</td>
</tr>
<tr>
<td>Grid system</td>
<td>3 / 5</td>
</tr>
<tr>
<td>Level or depth records</td>
<td>4 / 5</td>
</tr>
<tr>
<td>Photographs</td>
<td>9 / 10</td>
</tr>
<tr>
<td>Plans</td>
<td>4 / 5</td>
</tr>
<tr>
<td>Preliminary articles</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Publications</td>
<td>14 / 19</td>
</tr>
<tr>
<td>Records of sale</td>
<td>2 / 3</td>
</tr>
<tr>
<td>Satellite imagery</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Shapefiles</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Site lists</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Site names</td>
<td>8 / 9</td>
</tr>
<tr>
<td>Site records</td>
<td>3 / 3</td>
</tr>
<tr>
<td>Sketch maps</td>
<td>14 / 17</td>
</tr>
<tr>
<td>Topographic maps</td>
<td>6 / 6</td>
</tr>
<tr>
<td>Travel accounts and books</td>
<td>2 / 3</td>
</tr>
<tr>
<td>Videos</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>2 / 2</td>
</tr>
</tbody>
</table>
C. Accessing spatial information

Participants’ overall ability to physically obtain the sources of spatial information they identified was mixed. A variety of factors influenced the accessibility of the resources and data sources themselves, as well as the usability of the information within the resources (Table 4). While many of the researchers noted that efforts were being made to digitize and share catalogs or collection materials from archives and museums, eleven of the researchers had traveled in order to access materials. This was due in part to staff and funding limitations which prevented the institution from being able to search and send records to the researcher, and part because reviewing and interpreting the content of these collections required specialized knowledge.

[Data] really needs to be shared better. Even to do all the stuff I’ve been doing I’ve needed to go to [location] on four separate occasions and actually sit in the archives with a list of things I absolutely needed to find and only I could find. And that’s with all the, they’ve put so much effort towards making it accessible online and digitizing everything and they still haven’t scratched the surface of it (Interview 11, visiting scholar).
Table 4: Factors influencing ability to access sources of spatial information

<table>
<thead>
<tr>
<th>Factor</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collections lacking organization</td>
<td>7 / 7</td>
</tr>
<tr>
<td>Collections full or too full</td>
<td>6 / 6</td>
</tr>
<tr>
<td>Collections well organized</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Disagreements</td>
<td>2 / 2</td>
</tr>
<tr>
<td>ILL</td>
<td>5 / 5</td>
</tr>
<tr>
<td>No longer working in archaeology</td>
<td>5 / 5</td>
</tr>
<tr>
<td>Obscure publication</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Professional relationship</td>
<td>16 / 20</td>
</tr>
<tr>
<td>Purchasing data</td>
<td>3 / 3</td>
</tr>
<tr>
<td>Shared online</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Subordinate role</td>
<td>3 / 3</td>
</tr>
<tr>
<td>Travel</td>
<td>11 / 11</td>
</tr>
<tr>
<td>Within a publication</td>
<td>13 / 15</td>
</tr>
</tbody>
</table>

Additionally, participants explained that the level of organization at these institutions repeatedly impacted the chances for retrieval. Researchers often described the archaeological repositories managed by government institutions as lacking the funds, personnel, and skillset to manage the collections: “the museum isn’t in great shape in terms of organization, organizational skills” (Interview 1, graduate student). The participants explained that collections deposited in these repositories are rarely cataloged or further processed by the staff and subsequent retrieval was dependent on the depositor’s organization and documentation. On top of these limitations, the same institutions were repeatedly described as full and researchers had been encouraged to find alternate locations to store their materials. In these cases, instead of searching a single repository, the participants described having to consult colleagues in order to create a list of potential institutions but stated that they were often unsure where materials ended up. Additionally, one researcher described how the lack of storage space at the repository in her region has caused materials within collections to be separated from one another.
All of them will be in that one museum but because they are so filled, their storage rooms are completely filled from wall to wall and floor to ceiling... what they do now, is they go through the collections, pick out materials that they want to keep and they will keep that and then the rest they give back to the archaeologists and it’s your responsibility to find a better storage for them (Interview 14, graduate student).

The participants frequently described textual documents, both published and unpublished, as central to locating sites and spatial information. Grey literature, most often reports submitted to permitting and funding institutions at the end of a field season in archaeology, were referenced in eleven of the interviews. Researchers would contact institutions, agencies, and even the individuals that had participated on the original research project, in attempts to obtain a copy of the report. One researcher indicated that he had been fairly successful at getting copies of these reports held in personal libraries around the world: “There is a lot of grey literature that is really hard to find but we find it and track it down... find it in some person’s library somewhere” (Interview 13, faculty). Another researcher currently has a Graduate Student Researcher working to track down publications and grey literature, frequently through Interlibrary Loan (ILL) requests. He explained that due to regional policies copies of these reports were frequently deposited in a series of libraries and research institutions.

As seen in the techniques to identify potential sources of spatial information, many of the participants also utilized professional relationships with colleagues and other scholars to obtain publications and datasets such as field notes, site coordinates, and shapefiles. However, disagreements among collaborators and differences in perceived roles on a project limited the ability of participants, and in particular graduate students, to follow-up with the original data collector. The graduate students interviewed frequently explained that requests for spatial data
were made by their faculty advisor or a senior member of the project and that they were hesitant to approach the data collector for more information.

In one case, the participant traveled to different countries in order to meet with scholars in-person, intending to develop a professional relationship with those individuals. He described needing to build this relationship in order to request access to their personal libraries and records. He described having to find the “one person” that would know about an obscure site or publication and even in some cases getting them to sketch a map or locate a site on a printout. Five of the researchers described situations in which they were unable to contact the original data collector because they had either passed away or were not working in archaeology any longer.

While the overall perception of sharing publications and grey literature among researchers was favorable, the participants had mixed experiences concerning sharing data files themselves. Half of the participants described the general attitude about sharing data in terms of the regional context, with an even division between “pro-sharing” and “anti-sharing” attitudes. In the regions described as pro-sharing, the participants felt that the majority of the researchers working in the region were willing to give their data to others if asked. On the other hand, in the regions the participants called anti-sharing, researchers guarded their data more closely and it was not common to ask for someone’s data. One researcher described the researchers in his region as fairly pro-sharing and cited the reason as being due to governmental mandates surrounding data submissions and access. Another researcher contrasted two regions she had worked in previously and referenced the amount of potential work in a region as a possible reason for the differences.
An additional facet to access that was revealed during the interviews was the length of time that may pass between identifying a source, requesting access, and actually obtaining the spatial data. One researcher described having to wait years while an archive was renovated before being able to view its contents. Another researcher had to wait months before a colleague that had left archaeology to work in another field was able to return to the storage space where he had deposited his notes and photographs.

D. Processing spatial information for reuse

The archaeologists interviewed described a variety of methods and techniques they employ to reuse spatial data (Table 5). Frequently, “quick and dirty” methods of checking for errors and estimating the accuracy of a dataset are initially conducted—most often through free mapping software such as Google Earth or cell phone applications. Nine of the participants conducted field checks on spatial data they obtained from other sources. These field checks would include attempts to relocate a site or to record the data necessary to “ground truth” larger datasets. The most common approach was to synthesize multiple types and sources of spatial data in order to determine accuracy and in an effort to identify limitations associated with these data. By combining spatial information from field notes, descriptions found in publications, photographs, and maps the participants gained a better understanding of the context in which the data were captured and processed. In two cases, this combining approach allowed the researchers to identify changes in personnel and recording systems on the original projects that were leading to confusion during reuse attempts.

49 Field checking involves physically going to the location to check the data. Ground truthing is a method of calibrating or adjusting an entire dataset based on field checking a subset of the data.
Table 5: Techniques used to process spatial information

<table>
<thead>
<tr>
<th>Processing technique</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotating</td>
<td>5 / 5</td>
</tr>
<tr>
<td>Approximating</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Crowdsourcing</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Entering into ArcGIS</td>
<td>9 / 12</td>
</tr>
<tr>
<td>Field checking</td>
<td>9 / 10</td>
</tr>
<tr>
<td>Following up for more details</td>
<td>5 / 6</td>
</tr>
<tr>
<td>Georeferencing</td>
<td>6 / 7</td>
</tr>
<tr>
<td>Ground truthing</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Google Earth</td>
<td>7 / 8</td>
</tr>
<tr>
<td>Phone app</td>
<td>3 / 3</td>
</tr>
<tr>
<td>Reconstruct context</td>
<td>6 / 7</td>
</tr>
<tr>
<td>Rubbersheeting</td>
<td>1 / 1</td>
</tr>
<tr>
<td>Scanning</td>
<td>10 / 11</td>
</tr>
<tr>
<td>Stylistic evaluations</td>
<td>6 / 7</td>
</tr>
<tr>
<td>Synthesizing and joining data</td>
<td>13 / 19</td>
</tr>
<tr>
<td>Transcribing</td>
<td>4 / 5</td>
</tr>
</tbody>
</table>

Additionally, as seen in the example in Figure 1, joining together multiple datasets allowed one archaeologist to conduct analyses based on spatial information she would not have been able to otherwise. However, this example also illustrates one of the most costly issues associated with processing spatial data for reuse—time. The amount of time, effort, and expertise required to process these datasets so that they can be reused is significant. Twelve of the participants noted time as being an issue associated with processing previously collected spatial information.
Nine of the archaeologists interviewed incorporate the geospatial platform ArcGIS (Geographic Information System) into their workflow. Six of these researchers described georeferencing maps, drawings, and aerial photography while four described entering point data gathered from other sources. The participant who described the need for better documentation established a system to systematically record and track the metadata and documentation she is able to find about each dataset she incorporates in her GIS files.

So what I do is, I go through all of the files and I have an Excel sheet and I essentially try to put together a quick glance at metadata. So I put down the name of the file, the location of it in the computer and all that, but then I also put down when the file was made or the year the data is referring to, and the source that it’s from, the level of detail that it’s on, what exactly it described, who published it specifically. (Interview 6, graduate student)

E. Reusing spatial information

The participants identified a variety of ways in which they reused the spatial data they were able to access (Table 6). Often these types of reuse correlated with how the researcher designed his or her research project and the particular conditions under which work was begun. Reusing spatial data for administrative and strategic purposes was associated with projects that were extending the study area or beginning to work at a new site:

“when we were choosing our concession area to think about the terrain” (Interview 1, graduate student)

“if we’re trying to apply for a permit they ask for a site map” (Interview 7, graduate student)

50 “ArcGIS | Main”, Managing Data and ArcGIS.
Similarly, researchers working on projects synthesizing datasets from multiple sites would reuse spatial data to identify the location of each site in order to conduct distribution and comparative analyses across these sites. Additionally, when investigating particular topics or materials from a site, researchers would often reuse spatial data in order to reconstruct the context these materials had originated from within the site. Five of the researchers described their reuse of spatial information as primarily a means of visualizing the study area or as imagery in presentations so that they could illustrate the context in which the original and subsequent research took place.

Table 6: Types of reuse defined by the study participants

<table>
<thead>
<tr>
<th>Type of reuse</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative and strategic planning</td>
<td>5 / 5</td>
</tr>
<tr>
<td>Comparative analyses</td>
<td>8 / 10</td>
</tr>
<tr>
<td>Distribution analyses</td>
<td>8 / 13</td>
</tr>
<tr>
<td>Locating sites</td>
<td>7 / 8</td>
</tr>
<tr>
<td>Presentations</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Reconstructions</td>
<td>5 / 6</td>
</tr>
<tr>
<td>Visualizations</td>
<td>5 / 5</td>
</tr>
</tbody>
</table>

The participants identified a few factors that facilitated processing and reusing the spatial data they had gathered. For written documents, one researcher called out the importance of the index in a thousand page manuscript she examined for much of her dissertation: “it had a really nice index, every house, what page it was mentioned, and by topic of features... I mean for a type written manuscript from 1975—it was never actually published—it is beautifully put together” (Interview 11, visiting scholar). Another researcher was pleased to have a word document with the transcribed field notes from a previous survey and the ability to easily compare the site description and coordinates since they were directly associated with each other on the page. The
site descriptions allowed her to locate the sites in the field even though she described the coordinates as being “a little off.”

However, more often participants identified issues that hindered or prevented them from being able to reuse the spatial data (Table 7). The participants interviewed were rarely able to describe datasets that had detailed, or in most cases, even any documentation associated with them. The methods described in the publication—if it existed and could be accessed—were often the only source of information about how the spatial data was originally collected. One participant described the need to document not only the details of the collection technique but also the processing and parameters utilized on the raw data: “That would be great, just to have, basically documentation of the process they go through with each of the files and how they analyzed it or processed it” (Interview 6, graduate student).

Table 7: Types of issues encountered by the participants

<table>
<thead>
<tr>
<th>Issue</th>
<th>Count (Archaeologist/Project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>8 / 11</td>
</tr>
<tr>
<td>Age of data</td>
<td>7 / 8</td>
</tr>
<tr>
<td>Changing personnel and systems</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Disassociated datasets</td>
<td>9 / 12</td>
</tr>
<tr>
<td>Found errors</td>
<td>7 / 8</td>
</tr>
<tr>
<td>Granularity or resolution</td>
<td>9 / 11</td>
</tr>
<tr>
<td>Missing or sparse documentation</td>
<td>8 / 11</td>
</tr>
<tr>
<td>Ownership of data</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Time</td>
<td>12 / 16</td>
</tr>
</tbody>
</table>

In eleven cases, the researchers expressed doubts about the accuracy of the spatial data and in eight cases the researchers found errors as they were processing the data. For two of the researchers, any kind of spatial information was rare or if recorded, the information was highly
generalized due to the lack of provenience for many of the artifacts they were studying. The age of data was frequently cited as an indicator of potential issues, although the acceptable ranges varied depending on the type of analyses the participant aimed to conduct. Changing technologies, recording techniques, and acceptable practices were often pointed to as increasing the potential usability of spatial datasets. However, one researcher provided a detailed example of an excavation that incorporated more recent geospatial technologies into their project without fully adjusting their workflow, and the amount of processing that he had to take on in order to rectify the spatial data.

One of the issues the participants faced when attempting to process and reuse spatial data dealt with the discrepancies between the granularity of the original data collection and the granularity needed by the participant. The level of granularity necessary to conduct analyses varies according to the type of research. While one participant studying stylistic variations over large geographic areas was pleased to get an approximate location for a site, another researcher was attempting to track down which burial an artifact was associated with in a tomb with seven other individuals struggled with the lack of detailed information. Nine of the sixteen researchers expressed difficulties due to the original data collection occurring on a higher level than they required for their projects.

Additionally, on twelve of the projects described, the participants had to deal with disparate datasets and inability or difficulty joining these data. In one particular case, the researcher was able to track down a large collection of photographs, detailed logs for every artifact photographed on a site, and field notes (daybooks) recording each excavated trench and the depths they were excavated to throughout the day. She was also able to view the artifacts with their associated tags, on which an artifact identifier was recorded. However, these datasets
didn’t share a common identifying field on which they could be easily joined. In order to
determine which trench and at what depth an artifact in the photograph was from, the researcher
had to match the photograph to the physical artifact based on appearance, then use the artifact
identifier recorded on the associated tag to match it to the record on the photograph log—from
which she could identify the trench (Figure 1). In some cases she was able to identify which
daybook described that trench and if that find was described in the notes, she would be able to
determine the depth from which it was collected. Other researchers described similar scenarios in
which they had to conduct time-consuming, multistep processes in order to join data so that they
could tie spatial locations to their unit of analysis.

Figure 1: Workflow for joining multiple datasets from the same original collector

Similarly, two researchers ran into issues when attempting to join together plans or sketch
maps of individual features at a site. While the level of detail of the plan or sketch map was ideal
for their analyses, they were unable to find any data about the proximity of one feature to another. For example, Participant 3 showed multiple scans of house plans he had obtained in which every stone in a wall was mapped and drawn in; however, without a comprehensive site map, he could only approximately place them within the site.

VIII. Discussion

The discussion is divided into two sections. The first section examines recurring themes associated with the context in which the participants seek, identify, and access spatial information. The second section delves into the challenges associated with processing and reuse of spatial information.

A. Context for reuse

The archaeologists interviewed begin seeking spatial data during the initial design of their research project and continue to identify and gather spatial information from a wide range of sources throughout the life span of their projects. The participants discussed data with the expectation that they would be and should be reused and reusable within archaeological research. That expectation supports the view of data as inputs and outputs within scholarship meriting infrastructure and scholarly practices that support data curation, dissemination, and preservation. However, the extent of resources utilized by these participants suggests the array of challenges associated with finding and accessing spatial information (Table 3).

The interdisciplinary nature of archaeological research is reflected in the conditions surrounding the projects described by the participants. The majority of the researchers in this

---

51 Borgman, Scholarship in the Digital Age Information, Infrastructure, and the Internet, 115–16.
study were gathering spatial information in order to conduct analyses on a specific type of material previously collected from sites. However, the discrepancies in the level of granularity collected by the original investigators and the granularity needed to conduct further analyses highlight how knowledge of a specialty or sub-specialty within the domain influences the ability to reuse spatial information. As Faniel and Jacobsen, Wilson, and Borgman all noted, domain expertise is vital to documenting and curating usable datasets.\(^5\) Additionally, the findings from this study support the argument by Faniel et al that data management and information science skills need to be incorporated into archaeological training so that research design and data collection methodologies facilitate subsequent reuse by a variety of specialists.\(^6\)

Comparing data reuse practices within archaeology to other fields exposes factors unique to archaeology that impact data reuse. The literature review presents a series of studies on data reuse in ecology, a field which—similar to archaeology—aims to understand interactions and context by examining individual elements within an area. Zimmerman’s study of data reuse among ecologists showed that field experience played a significant role in selecting and evaluating previously collected data.\(^5\) Similarly, the archaeologists interviewed in this study frequently conducted field checks or incorporated spatial information into spatial datasets they had collected in the field. However, unlike the findings from Zimmerman’s study of ecologists, the archaeologists interviewed did not mention prioritizing or evaluating datasets based on their knowledge of the original data collector’s amount of experience. While the researchers that were


\(^{6}\) Faniel et al., “The Challenges of Digging Data.”

\(^{5}\) Zimmerman, “New Knowledge from Old Data.”
pulling from centuries old datasets emphasized that the original collectors were not trained archaeologists, they still incorporated the spatial information into their own research. This discrepancy between archaeological and ecological data reuse practices may be explained by the inability of archaeologists to recollect this information.

While many of the participants were familiar with established archaeological data repositories and services, none had searched these services while seeking data for their own projects—analogous to the archaeologists interviewed by Faniel et al.\textsuperscript{55} Instead, data were commonly obtained by this study group through formal academic channels such as publications and written reports, and informal channels including correspondence and communications among colleagues and collaborators. In accordance with scholarly traditions, nearly every participant described publications and grey literature as essential to their efforts to obtain and understand spatial information. Frequently, textual descriptions, site names, photographs, and maps would be extracted from the publications and incorporated into the researcher’s own knowledge system. These practices illustrate that data are not standalone entities—one of the issues Borgman describes as being associated with the push for data publication.\textsuperscript{56} The role of publications within the practices described by the participants does however support the proposal for integrated digital data publications.\textsuperscript{57} As we consider options to increase the dissemination of data in archaeology, the current scholarly communications infrastructure and practices can provide a framework on which to build new methodologies and standards.

\textsuperscript{55} Faniel et al., “The Challenges of Digging Data.”

\textsuperscript{56} Borgman, \textit{Big Data, Little Data, No Data}, 47.

\textsuperscript{57} Kansa and Kansa, “We All Know That a 14 Is a Sheep.”
B. Processing and reuse

The findings from this study illustrate how significant processing is to data reuse in archaeology. In the study on reuse of archaeological data, Faniel et al found that the two most pressing issues for researchers were the lack of context and the lack of procedural documentation. While lack of documentation was frequently cited as a problem by participants in this study, the most significant issue surrounding processing was time. Nearly every interview contained descriptions of the amount of time required to process these data into a usable format. Additionally, the researchers described the difficulties associated with disparate datasets and their inability to join spatial data to the objects they were studying. In these accounts it became clear that joining these datasets was often difficult, time-consuming, and sometimes not even possible. Even reconnecting datasets from the same project—which was described as well-documented—proved challenging. As Faniel et al propose, the inclusion of information professionals and information science skills into archaeological practices can reduce the prevalence of disassociated datasets and issues surrounding reuse. In addition, the findings from this study support the need for this inclusion to occur throughout the life cycle of a project.

In addition to data rarely having what the researcher would have deemed adequate documentation, the interviews revealed a disconnect between the person obtaining the data and the person processing. Whereas the person processing the data was aware of what kinds of documentation and metadata would have best supported reuse, they were not in a position to interact or clarify with the original data collector. Within this study, the graduate students and

---

58 Faniel et al., “The Challenges of Digging Data.”

59 Ibid., 302.
post-doctoral or visiting scholars were more commonly the processors while the faculty
participants tended to be the person obtaining the data. It would be valuable to further investigate
this discrepancy with a larger and broader study population. In addition, while the agreement to
maintain the confidentiality of the participants prevented publishing analyses based on
differences among regions or specializations, future studies can investigate these differences by
conducting research with a population that agrees to be identifiable.

The perception of technology associated with data collection, management, and
dissemination varied among the participants. Researchers who had dealt with particularly
difficult data wrangling issues expressed hope that newer technology would mitigate these issues
in the future. However, the repeated accounts of lists of GPS coordinates with only a site name
associated and folder of unsourced digital elevation models reveal that reusing spatial data
collected by recent technologies remains vulnerable to similar challenges.

During the interviews, many of the researchers spoke about the need for better data
sharing practices and the importance of reusable data to the discipline. In particular, two of the
researchers described in depth the amount of detail and effort they go to now when depositing
their own materials at these institutions in order to make them as usable as possible.
Furthermore, the influence of data reuse issues on data practices, as described by the
archaeologists in this study, extends beyond depositing practices. Participants repeatedly stated
their plans for sharing or publishing their own data so that others would be able to access and
reuse them in the future. The inclusion of information professionals in archaeology and the
development of systems that align with the practices identified in this study—in particular the
merging of datasets and centrality of publications—can facilitate these participants’ data plans.
IX. Conclusion

The results of this study reveal that archaeologists across regions and specializations frequently attempt to find, access, and reuse spatial data in order to conduct further analyses on archaeological remains collected by other researchers. Regardless of the motivations for seeking spatial information, their efforts continue throughout their research project and attempts to obtain and process data may extend over years. While the level of granularity necessary for analyses varied across projects, the findings show a common desire for documentation of the original collection methods and data processing. Additionally, the inability or difficulty associated with joining disparate datasets revealed a pressing need for a wider investment in understanding data collection and management practices going forward. The practices surrounding reuse of spatial information reveal a pressing need for information science professionals and skills within the field of archaeology.
X. Appendix: Interview script and questions

After collecting consent form:

Thank you for agreeing to speak with me today. The purpose of this interview is to get an understanding of how you reuse spatial data in your work. Spatial information is central to archaeology and frequently, projects incorporate or build on spatial datasets that were previously collected. We’ll georeference maps of excavated trenches, try to trace down the location of a site through narrative descriptions in field notes, or even share 3D models of architectural features. In this interview I want to get a better understanding of how you find and reuse spatial data in your own work. I’d like to focus this conversation around your recent publication [title] and, in particular, the spatial data that are associated with it.

1. Could you describe the types of spatial data that you gathered from other sources? [If clear from publication, use example of reused spatial data]

2. Did you purposely seek out these data with the intention to reuse?

3. How did you search for them? [If more prompting is needed- Requests to colleagues? Publications? Repository?]

   Have you used this/these source(s) previously?

4. Once you found these data, how did you access or obtain them?

5. How did you determine if the data was in fact reusable in your context?

6. How did you envision using these data? Why did you decide to reuse instead of collecting the data yourself?

7. Can you walk me through what you did with the data after receiving them?

   If not clear from description:

   Were the data incorporated into your own datasets? Did they remain separate?

   How was the source tagged or indicated in your data management system?

   What information do you wish had been included with the data?

8. Would you consider the reuse a success?

   If yes: What would have made the process easier?

   If no: What could have made the process a success?

9. Is there anything you can think of that I’ve forgotten to ask or anything you’d like to share?
XI. Glossary

Access  Researcher’s ability to obtain a resource containing spatial information.

Artifact  Any portable object used, modified, or made by humans.

Data  Entities used as evidence of phenomena for the purposes of research or scholarship.

Data sharing  Transfer of data from one individual to another. This transfer may be facilitated by a third party or service such as a repository or archive but does not require this intermediary.

Data publication  Data are published within a written text or as standalone datasets online in a data repository, archive, or dissemination service; introduces actions taken prior to transfer including an editorial review of the content and a system of verification.

Granularity  The scale or level of detail present in a set of data; pertaining to the unit of analysis.

Grey literature  Documents such as reports and working papers produced by researchers and organizations outside of the traditional commercial or academic publishing and distribution channels.

Information seeking  Conscious effort to acquire information in response to a need or gap in knowledge; that knowledge may be actively or passively gained.

Initial use  Use by the original collector for the purposes of research or scholarship.

Metadata  Information associated with data entities that provide the context of the phenomena, research, and any alterations that have occurred.

Relocate  To locate again; find an object, feature, or site that had previously been located.

Reuse  Any subsequent use of data in a context other than that of the original collection.

Spatial information  Entities serving as evidence of geographic locations associated with phenomena.

Usability  A resource’s format, its usefulness, and the requisite knowledge and skills associated with the use of the resource’s content.
XII. Bibliography


