

“Not for Casual Readers:” An Evaluation of Digital Data from Virginia Archaeological Websites

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Mark Antony Freeman

August 2015

Licensed as CC BY-NC-ND

Creative Commons: Attribution, Non-
Commercial, No Derivatives

<http://creativecommons.org/licenses/by-nc-nd/3.0/us/>

2015 by Mark Antony Freeman

Acknowledgements

I would like to thank my wife, Barbara Heath, for her encouragement, support and insight.

I'd also like to thank my committee for their expertise and considerable help.

Chair: Dr. Xiaohua Zhu

Dr. Suzie Allard

Dr. Carolyn Hank

Finally I'd like to thank the following people who assisted with the research: David Anderson, Gillian Galle, Fraser Neiman, Jolene Smith; and Eleanor Breen and Esther White whose initial study was part of the basis for this research.

Abstract

Archaeological data dissemination is complicated by the need to serve disparate audiences, each of which has different data needs. This study examined the websites of 148 Virginia institutions identified as having archaeological collections or data, and used content analysis to see how they supported characteristics of scholarly publishing, open data and public outreach. Archaeologists are increasingly looking for comparative data sets for research needs, with professional ethics and a desire for public engagement encouraging data sharing. However this analysis suggests that, while there are some exemplary websites, much of the archaeological record remains publicly inaccessible. The majority of websites examined provided no reference to archaeology and, of the remainder, a third did not provide archaeological data. These websites did provide many supporting characteristics for public outreach, but concerns about preservation, data “openness,” and limited datasets remain.

Table of Contents

Chapter 1 - Introduction	1
Overview	1
Definitions used in study.....	3
Conceptual Framework.....	4
Significance	8
Chapter 2 – Literature Review	12
Archaeological publishing and the curation crisis	12
Archaeological Data	18
Archaeological repositories and data publishing options.....	20
Scholarly publishing	27
Open data.....	30
Public outreach and audience.....	32
Summary	39
Chapter 3 - Methods	41
Content analysis	41
Population	42
Data Collection	43
Data Analysis	47
Chapter 4 – Results	49
“No Archaeology” web pages	49
Archaeology content, but not data.....	51
Public Outreach.....	51
Archaeological Data	54
Open Data	55
Scholarly publishing	56
Archaeological data sub-clusters	57
Chapter 5 – Discussions	58
Overview	58
Limitations.....	59
Scholarly Publishing	61
Open Data	65
Public Outreach.....	66
Findings by institutional type.....	68
Aggregated websites.....	69
Conclusions	71

References	75
Appendices.....	85
Appendix A – institutional websites holding archaeology collections.....	86
Appendix B – Website Coding System	91
Data Management	91
Institutional Information	91
Data Content.....	91
Scholarly Publishing	92
Open Data	92
Public outreach	92
Vita	94

List of Tables

Table 1. Data collection fields by section	46
Table 2. Clusters by presence of data	49
Table 3. No archaeology pages by organizational type	40
Table 4. Public outreach characteristics	52
Table 5. Number of public outreach characteristics	53
Table 6. Types of data	54
Table 7. Open data by type	55
Table 8. Five star rating	56
Table 9. Scholarly publishing	57

List of Figures

Figure 1. Conceptual Framework	5
Figure 2. Benefits of Open Data	9
Figure 3. Clusters by presence of data	49
Figure 4. No archaeology pages by institution type	50
Figure 5. Public outreach characteristics	52
Figure 6. Open data by type	55
Figure 7. Scholarly publishing characteristics	57

Chapter 1 - Introduction

Overview

An early tenet of archaeology was “preservation by record,” the idea that publishing the detailed record of an archaeological site served to preserve it. Data publication has always been integral to the practice of archaeology, a connection explored by Richards (2004) when referring to the position attributed to the 19th-century English archaeologist Augustus Henry Lane-Fox Pitt-Rivers that “publication provided an objective record” (Richards, 2004, section publication and archiving). Pitt-Rivers also said that his voluminous excavation volumes were “not intended for casual readers,” raising the still relevant question: who are the audiences for archaeological data (section: publication and archiving)?

How archaeological data are disseminated has changed over time, and the web has changed communications in most disciplines. A number of factors are important to this evolving environment. Federal grant requirement supporting the open data movement are pushing grantees, including archaeologists, to consider where their data will eventually reside, and how they will be accessible (National Science Foundation, 2013). This movement coincides with an increased emphasis on collaboration, data sharing and using existing collections. Kintigh et al. (2014), in “Grand challenges of archaeology,” noted the need for “far more comprehensive online access to thoroughly documented research data and to unpublished reports detailing the contextual information essential for the comparative analyses” (p.879). Finally the professional and ethical standards of archaeologists continue to stress the importance of public outreach; that archaeologists “strive to engage citizens in the research process and publicly disseminate the major findings of their research” (SHA 2003, section: Principle 7).

Digital technologies have allowed archaeologists to capture more information from archaeological sites. This information includes digitizing what had been paper records (field notes and

artifact catalogs); increased photography from digital images, supplemented by video data; digital spatial data; and laboratory data including databases, images and telemetry. In addition, the National Historic Preservation Act mandated archaeology for federal projects, leading to more archaeology being done, and increasing the extent, and amount of archaeological data. In the United States alone, current archaeological projects continue each year with “approximately 50,000 field projects involving archaeological resources” (McManamon, Kintigh, & Brin, 2010, section: The Need for Digital Archiving). There are questions regarding the extent to which artifact collections’ data are being archived and, in addition to new data being generated, a significant backlog of non-archived, pre-digital material exists. Like many other professions, the shift to digital data in the last three decades has made it difficult for repositories to keep up with both physical and digital collections (Watts, 2010, p. 1). The situation with physical material has meant that some of the archaeological work done in previous decades is now inaccessible to researchers. Missing field notes, artifacts that were never washed or cataloged, paper catalogs that were not preserved, or that are too lengthy to use efficiently, all mean that the evidential value of these archaeological sites is limited (Sullivan & Childs, 2003). These problems continued with digital data, making aspects of preservation and access ever more crucial in the dissemination of new archaeological digital data.

In examining current practices, how are archaeological data made accessible? If we examine a subset of dissemination, to what extent are websites presenting archaeological data aligning with the requirements of open data? Are there characteristics supporting aspects of scholarly publishing? Are there characteristics supporting aspects of public outreach? This study will attempt to answer these questions through a content analysis study of the websites of institutions identified as having archaeological collections or data in Virginia. Characteristics of open data, scholarly publishing and public outreach are identified, and the organizations’ websites systematically analyzed to see whether

these characteristics are present. The intent is not to provide a critique of individual sites, but to understand of how archaeological data from Virginia projects is being disseminated through the web.

Definitions used in study

For the purpose of this study the following terms and concepts are defined as follows:

Archaeology: is the study of human activity in the past. Through excavation and the study of material culture, archaeologists create many forms of data, and it is these data that form the basis of this study.

Websites: This study will use the website as the unit of analysis. If institutions have multiple websites, identified through unique website top-level domains, they will be treated separately. In cases where archaeological data are provided at another website, the institution will be considered to have disseminated data, provided there is an explicit link. Websites are dynamic, so this study by necessity only analyzes website data available from the limited time frame of the study (December 2014 to January 2015).

Digital archaeological data: The purpose of this study is to identify websites that include archaeological data, rather than summary reports. To make a more fine-grained analysis, these data will be categorized into: archaeological site reports; field records; artifact collections; spatial data (GIS); virtualizations; and 3D models. These categories are more fully explored in the research methods section. A news website reporting on an archaeology excavation will not be considered as having “data,” however a blog post by archaeologists that provides information on new findings will be included.

Re-usable data: This term is used in this study to refer to formatted digital data. Interoperable data, which is a term also used in this study, is a subset of re-usable data since they requires metadata allowing the data to be used with other datasets.

Virginia projects: The focus of the study is on data that originated from archaeological excavations conducted in Virginia. Some Virginia archaeologists work around the world, however this study is only concerned with archaeological projects undertaken in Virginia.

Virginia was chosen for this study for a number of reasons. A study undertaken by White and Breen (2012) listed a wide range of institutional types in Virginia holding archaeological data. White and Breen focused on physical collections and they provided a comprehensive list of repositories, which was used as a starting point for this analysis. For historical archaeology, the middle Atlantic, including Virginia, is the most studied area in America (Society for Historical Archaeology, in press). V-CRIS, the state archaeology site database (http://www.dhr.virginia.gov/archives/archiv_vcrisHome.htm) has recently been updated and, beyond a number of university archaeology programs, there is a range of long-term archaeology projects in Virginia including those at Jamestown Rediscovery, Colonial Williamsburg, Monticello, Mount Vernon, Montpelier and Poplar Forest. The Digital Archive of Comparative Slavery (www.daacs.org) is based in Virginia, and A Comparative Archaeological Study of Colonial Chesapeake Culture (<http://www.chesapeakearchaeology.org/>), including many Virginia archaeological sites, is close by in Maryland. While archaeology in the United States is not a completely homogeneous activity, most archaeology is conducted by cultural resource management companies in response to federal regulations that, along with guidelines for federal grants, apply equally across the United States. Additionally, national conferences, memberships in regional, national and international professional societies, and professional ethics provide some commonality of practice.

Conceptual Framework

This study is situated in concepts from the open data movement, scholarly publishing and public outreach.

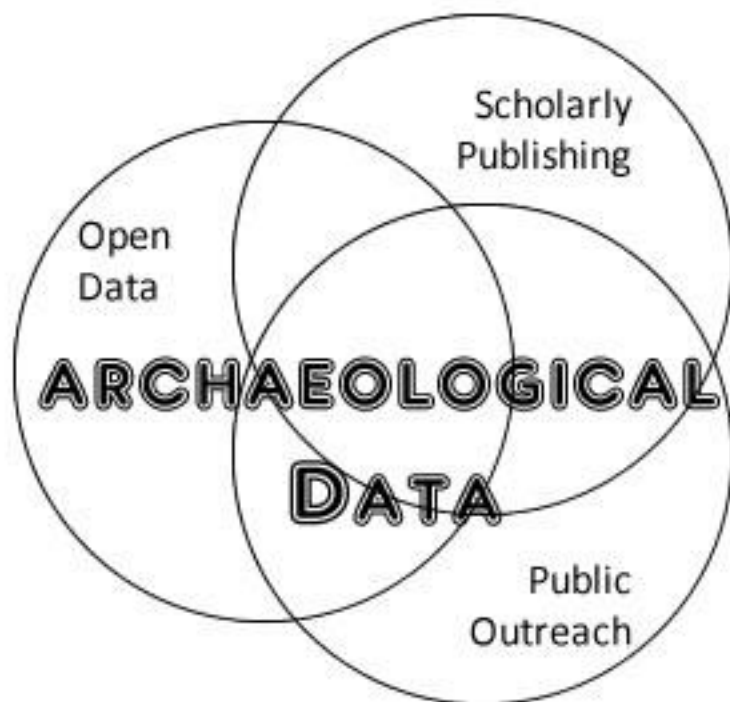


Figure 1: Conceptual framework

Scholarly access and public outreach are closely intertwined with open data concepts, but each maintains some distinctions in focus, and in the supporting literature. Figure 1 shows these three concepts laid on top of archaeological data. It is not suggested that there is a “sweet spot” where these ideas come together. For example, a website

intended for other archaeologists might provide little support for public

outreach. However websites are likely to be more effective for multiple audiences as they tend to the center.

Scholarly publishing: Unquestionably, the largest potential audience for archaeological data will be other archaeologists, and issues of scholarly publishing are pertinent to this study. Scholarly publishing here is considered a part of scholarly communication, concerned with data accessibility, standards and reuse. Accordingly this study will examine how these data are described, and whether they can be manipulated and reused. The licensing and citation of data are important factors in using data for scholarly research and these characteristics will be considered. Finally the archiving of these data will be considered; does the website provide information, or a provision, for these data to be available in the future.

Archaeological research questions are increasingly focused on comparative collections, with data from multiple sites offering the promise of new areas of research. As noted by McMillan, Hatch and Heath (2015) “The continued opportunity to reanalyze old collections and incorporate them into more current historical narratives has become even more significant in recent years with funding cuts in archaeology” (in press). But The Andrew Mellon Foundation stated in its 2008 annual report “there are no commonly agreed upon standards and little capacity for preserving and providing access to the databases, electronic field notes, digital images, and other materials that document archaeological findings” (Andrew W. Mellon Foundation, 2008, p. 24). While the environment has improved since 2008, issues of standards and capacity still exist. For professional use, accessible data means access, preservation, standards, interoperability, permanence and citable source. Borgman’s observation that “scholars in all fields are rewarded for publication but rarely for managing their information” (2010, p. 210), suggests the incentives support using other people’s data, rather than making data accessible. This point is addressed in part by granting agencies’ recent requirements for data management plans. Though National Science Foundation (NSF) doesn’t require public access to data, it does ask that awardees include information about what will happen to the data generated by their projects. In Virginia, as noted by White and Breen (2012), at present there is little capacity for a systematic and comprehensive study of archaeological collections, digital or otherwise.

Open data movement and archaeology: Archaeology, like other disciplines, exists in a changing data environment with a recent push for open data (Beck & Stott, 2013; Beck, Bevan, & Stott, 2012). Open in this sense is defined in a similar way to open knowledge: “Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness” (Open Definition, n.d.). The shift to opening access to data sets is part of the Open Source and Open Government movements (White House, 2013), and has been accompanied by granting agencies such as NSF asking grantees to identify where their data will eventually reside and how it will

be made accessible, as mentioned in the previous paragraph (National Science Foundation, 2013). Accordingly this study looks at the type of data on the website: text, structured data, non-proprietary structured data, and linked open data. Other considerations included whether access is free, and whether there is a need for browser plug-ins or proprietary software to access the data. Finally, does the website provide metadata through a data dictionary, thesaurus or other ontological tool?

Public outreach: Archaeological ethics strongly support communicating research to a broader public. The principles of archaeological ethics of the Society for American Archaeology ask that archaeologists “communicate archaeological interpretations of the past” (SAA, n.d. para. Public Education and Outreach). They define the audiences for these communications as “students and teachers; Native Americans and other ethnic, religious, and cultural groups who find in the archaeological record important to aspects of their cultural heritage; lawmakers and government officials; reporters, journalists, and others involved in the media; and the general public” (para. Public Education and Outreach).

Archaeologists have a long history of public engagement through museum exhibits, public talks, site tours and open houses, field schools, and engagement with the media. However, archaeologists are not the only voice in dealing with the past. There is a popular culture perception of archaeology that has recently been affected by film and cable television programming to include practices that may involve digging, whips, metal detectors and artifacts, but that are not archaeology. Archaeology outreach becomes, in the face of these subversions of archaeological practice, not just education, but advocacy. Archaeologists, ideally, need to explain not only the results of their excavation but the methods and questions of the discipline. To encourage preservation and to promote good stewardship of archaeological resources, archaeologists need to show what was learned, and also why archaeology, as practiced by professionals, is necessary. As Richards and Robinson (2000, section 1.3 Resource discovery and re-use) state “We no longer operate in a world where archaeological data are created by

archaeologists for archaeologists.” Archaeologists are careful to stress the nature of their excavations as distinct from “pot hunters” of archaeological sites. The Canadian Archaeological Associations position paper (Canadian Archaeological Association, 2014) decries the use of broadcast media showing excavations for pleasure or profit, after explaining the detailed process of archaeology. Integral to excavation process, they note, is the publication, formal or informal, of the work undertaken. A statement in the British PUNS report observes "the primary purpose of publication to be the provision of information to facilitate research, and the dissemination of knowledge for public benefit" (Jones et al, 2003, section: Overall conclusions: The functions of publication). It would be easy to assume that syntheses support public outreach, and data support the professional community, but much more work is needed to understand the relationship between the public and archaeology data (some of which is referenced in the literature review). For the purpose of this study, websites will be evaluated on the availability of archaeological site summaries (in addition to data); the presence of contextual information; the use of supporting media; and the use of social media. Also examined is whether the website provides information on community outreach, and related programming, and does it include educational content? It is beyond the scope of this study to evaluate the effectiveness of this public communication, though this is an interesting area for further investigation.

Significance

Websites with characteristics supporting open data, scholarly publishing and public outreach provide many benefits. The broader importance of access to digital data, and in particular digital collections, to archaeologists is reflected in a Society of Historical Archaeology survey (Society for Historical Archaeology, 2015). Over 50% of respondents rated the importance of digital archaeology collections as very or somewhat important. Similarly a survey of repositories in 2010 asked “How important is the collection/curation and access to digital data to your organization?,” with nearly 50% answering highly important (Watts, 2010, p. 7). Kintigh and Altschul (2010) note in their exploration of

the need for a trusted archival digital repository for archaeology, “As consumers, archaeologists genuinely desire effective, on-line access to the results of others’ work” (p. 270). Integrating different datasets, managing diverse vocabularies, and supporting emerging data forms (e.g. GIS, 3-D scans and models) remains a challenge for the profession.

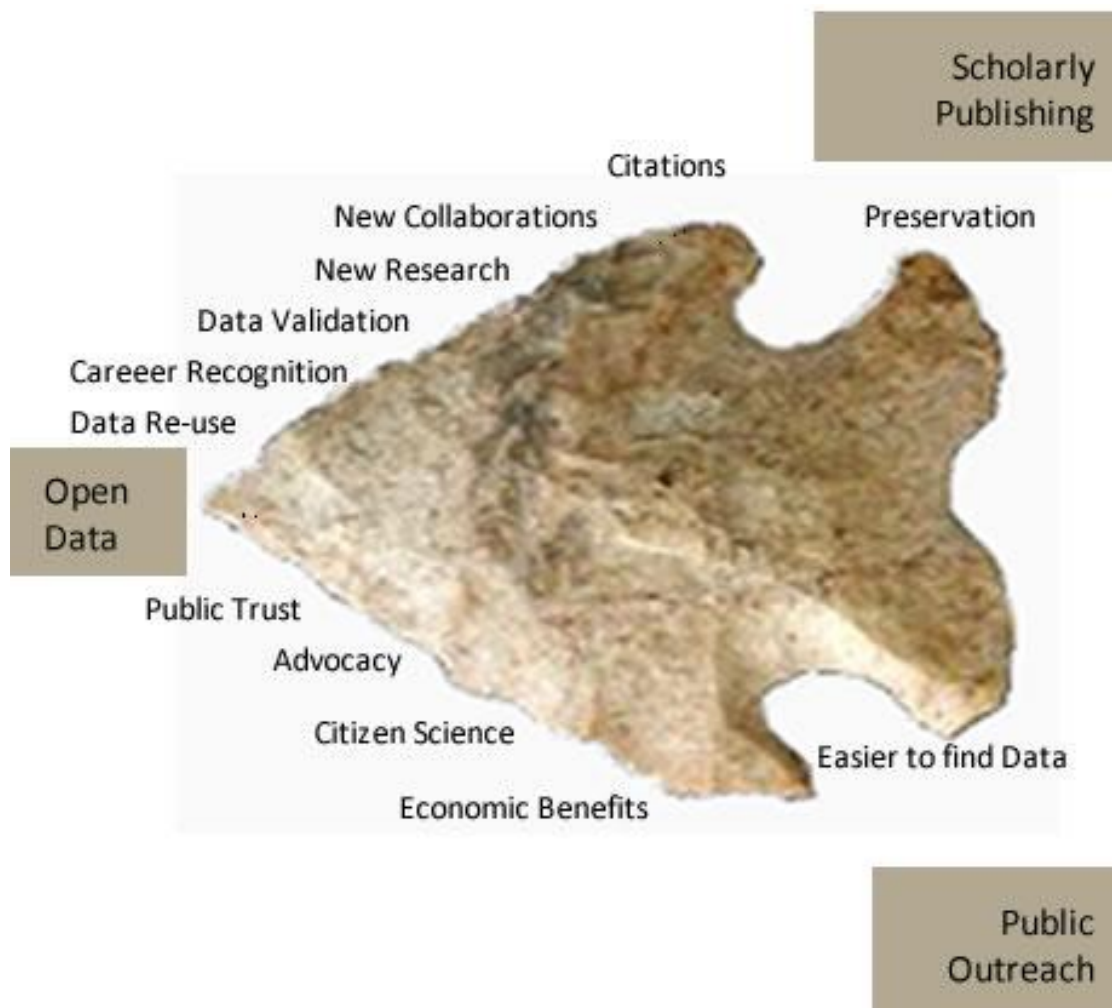


Figure 2 - Benefits of open data

Figure 2 is derived from the Journal of Open Archaeology Data, illustrating the benefits of accessible archaeological data (and replacing the cited image’s central figure of a single white male scientist with an Archaic point (JOAD, n.d.)) It shows some of the benefits of accessible archaeological

data for both scholarship and public interest. For scholars, making data available can mean more citations, the ability for new research and collaborations, the reuse of data and the validation of results. For the public there may be economic benefits – archaeological sites feature in development and tourism decisions – as well as the possibility for citizen science, and the need for public trust, since archaeological data are often created with public funds. The presentation of these data supports the practice of archaeology, differentiating it from pot-hunting. There is some justification in saying that even if the public never looks at archaeological data they should know that the data are available. Preserving data, often excavated at considerable cost, and making them accessible, thus serves many constituents.

The benefits of this study are two-fold. First, methodologically, it will be in identifying a set of criteria through which websites may be evaluated in meeting the overlapping demands of openness, scholarship and outreach. The study design, including coding system, may serve as a model for looking at websites in other areas, for new projects, and for dealing with previously excavated, but still inaccessible, materials. Second, this study's results provide a reference for where Virginia archaeological data currently resides, helping archaeologists with data discovery since, at present, these resources are diffuse. Finally the results of this study will show the degree to which current digital dissemination is serving multiple publics.

It is unlikely that any one website can meet the needs of multiple publics. Preservation, a major concern, considering the history and persistence of collections, suggests an emphasis on developing data format standards and strong metadata for housing archaeological collections within trusted digital repositories. For those most concerned with scholarly access, there is a choice between lightweight publication models that expose collections, and the development of multi-site comparable datasets, based on an open data philosophy. For presenting data to archaeology's many publics, there is the need to provide data in a rich contextual environment, while allowing for multi-vocality and engagement. This

study is an attempt to see what data are accessible, and how dissemination is balancing these characteristics of openness, scholarship and outreach.

Chapter 2 – Literature Review

Archaeological publishing and the curation crisis

The nature of archaeology publications has been affected by the increased archaeological excavations done since the 1960's, contributing also to a curation crisis for physical collections. Archaeologists (Marquardt, Montet-White & Scholtz, 1982; Sullivan & Childs 2003) noted over thirty years ago that some collections have remained unprocessed for decades, with the impact of legislation in the 1960's leading to an increase in material to process, and a change in where they were stored. This legislation shifted the focus of collections management from university departments and museums to state, federal and private repositories, and was part of a trend that saw collections being used less for research and teaching. This change in curation practices led to a decreasing emphasis on managing collections and many of these problems remain despite more recent attention (Sullivan & Childs 2003, p. 9; Childs & Kagan, 2008). Issues of potential digital dissemination of this material are thus rooted in problems of managing physical collections that have existed for a long time.

For archaeology done on federal land, the Archaeological Resources Protection Act, passed in 1979, stipulates that archaeological sites and objects are the property of the United States, and that the objects and associated records should be curated by a suitable repository. The Secretary of the Interior issued regulations about the exchange and disposition of archaeological collections, but it took another eleven years before 36 CFR Part 79 (Curation of Federally-Owned and Administered Archeological Collections - <http://www.nps.gov/archeology/TOOLS/36CFR79.HTM>), regulating the curation of federally owned and administered collections, was passed. These guidelines helped curators manage collections, but did not include language on digital, or electronic material, though admittedly this law was passed early in the digital revolution (National Park Service, n.d.). Addressing the use of collections, the legislation refers to scientific, educational and religious uses, with specific language regarding Indian tribes. The scientific and educational uses are constrained by language limiting access to the materials to

qualified professionals, though this term is broadly defined. It seems unclear whether digital dissemination would count under the broader exhibits and publications provision.

The majority of archaeology projects done outside of federal lands are federally mandated under Section 106 of The National Historic Preservation Act. These projects are carried out by cultural resource management (CRM) companies. Combined with federal projects, they accounts for 90% of the archaeology done in the United States. The reports and resulting data from these projects are rarely disseminated directly by the CRM companies, but managed through State Historic Preservation Offices, who act as the repositories for information on the projects (Advisory Council on Historic Preservation, 2009, p. 2). While most states maintain a database of archaeological sites, data structures, data types, and uses of numeric coding vary. These differing systems have grown over the last fifty years, often from card catalogs, and present very heterogeneous information (Anderson, Yerka & Kansa, 2013, p. 5). The Digital Index of North American Archaeology (DINAA) is an emerging attempt to address these issues by developing a national index of archaeological sites supported by data cleaning, an ontology tool to map terms, and added metadata. The DINAA project is in progress (the archive is published through Open Context), with fifteen states currently participating, but offers a strong model for presenting archaeological site level data across site boundaries. However, the various state repositories do not always manage other types of archaeological data. Collections are often just recorded at a summary level, and some research projects fall outside of the scope of the state repositories. For Virginia, like other states, the state database is a crucial component of digital dissemination and access, though the Virginia website, V-CRIS, is currently only accessible through a paid subscription.

Successful ways of publishing archaeological data need to consider preservation as well as access; the data from archaeology collections can still be of use decades later. One scholar quoted by Harley et al. (2010) suggested in reference to archaeology “I’d say, 80 to 90 percent of the data that’s being generated now would be lost” (p. 87). Richards (2004) explored this connection noting the

tradition attributed to Pitt-Rivers that the published report provided a complete and factual record, becoming an indispensable part of the excavation; forming the idea that it provides “preservation by record” (section: Publication and Archiving). This concept has changed over time, with the acceptance of the value of analysis and synthesis; and the idea that any report is a reflection of its authors, but the point remains that publication, in its broadest sense, is seen as integral to the process of archaeology. In some ways this idea mirrors the Jenkinson- Schellenberg debate in archival theory (Tschan, 2002). Jenkinson, a proponent of natural order, keen to keep a complete record without judgment from the archivist, reflects the older archaeological position, while Schellenberg’s more activist approach accepts the role of the archivist (or in this case archaeologist) as shaping the view of the past – a connection that was made by Mark Warner (2014) in a paper examining deaccessioning of archaeological collections presented at the annual meeting of the Society for Historical Archaeology.

The American Library Association (ALA) provides several definitions of digital preservation. ALA notes that “Digital preservation combines policies, strategies and actions that ensure access to digital content over time” (Association for Library Collections and Technical Services, 2009 para. Short Definition). Digital content is fragile, and needs management by adding value to content through metadata, and the “commitment to preserve digital content for future use; specify file formats to be preserved and the level of preservation to be provided; and ensure compliance with standards and best practices for responsible stewardship of digital information” (Association for Library Collections and Technical Services, 2009, section. Long Definition). Social science data, and particularly archaeological data, tend to have value over a long period, emphasizing the issues of preservation and access. It has been suggested that archaeology as a profession has not been overburdened with standards (Andrew W. Mellon Foundation, 2008), and the history of physical archaeological collections clearly shows the need for a more active preservation policy than has been the case in the past. Digital dissemination exacerbates these issues, bringing into sharper relief the difficulties of comparing data from different

sites. One reaction to these problems can be seen in the “Best Practices” for digital data that have emerged from Archaeology Data Service and tDAR (Archaeology Data Service & Center for Digital Antiquity, 2013). These recommendations for suggested metadata, supported file format standards, and preservation practices may have a positive effect across the discipline and, if well disseminated, could lead to better preservation and access.

From 1997 to 2008, Childs and Kagan (2008) undertook three studies of repositories having archaeological collections looking at all fifty states. Their interest was in understanding curation fees as part of a broader concern with maintaining collections, but the survey touches on aspects of digital access. While observing that space problems are increasing (p. 12), at the time of the survey these repositories were only just beginning to realize the new problems associated with digital data (p. 8), and recognizing the costs of archiving the data (p. 10). Access to the collections was not included as one of the top reasons for use of the fees (p. 9); a discouraging note in terms of the potential for digital access. Watts (Watts, 2011) addressed this question through an online survey and telephone interviews with staff of archaeological repositories in the United States to understand how they plan to “preserve and provide access to digital archaeological information” (p. 2). While collection and curation of digital data were as seen as very important (p. 7), Watts found repository managers “uncertain of what standards and procedures to apply” (p. 5) regarding guidelines for submission. Furthermore, only 20% of respondents provided remote access through a website to digital assets (p. 13), with access limited to the collections, in almost 70% of cases, to those with a professional affiliation or degree.

White and Breen (2012) conducted “A Survey of Archaeological Repositories in Virginia,” sponsored by the collections committee of the Council of Virginia Archaeologists in 2012. The study arose out of a desire to show the value of archaeological collections, in response to a perceived threat to financial support for collections care in the Commonwealth. The survey was able to show a wide range of institutional types that held archaeological collections, in addition to the state repository. The survey

found 78 repositories with archaeological collections, 65% of which explicitly responded that their collections were open to use by the public. It is interesting to note that access to these collections was limited, and use rare:

Of those whose collections are available to researchers or the public, an appointment is usually required. In breaking down the frequency with which collections are used, 12 said never, 20 said rarely (1 to 6 times a year), 6 said occasionally (once a month), and 9 said frequently (daily or weekly) (White & Breen, p. 9).

The limited actual access suggests either that these collections are not valued for research or, more likely, that they are hard to discover remotely.

The latter part of the White and Breen survey examined how the Virginia state historic repository recorded the location of archaeological collections. It found that for half of the archaeological sites surveyed, there was no information about where the collections were housed (many of these were from sites recorded prior to 1990). Of those that were recorded, over a third indicated that they were held at the cultural resource management company that excavated the archaeological site. These findings are problematic given that some of the companies are no longer in business and these companies “do not curate artifacts permanently” (White & Breen, 2012, p. 27). Furthermore 61% of field notes—vital contextual documentation—were not collected by the state. The survey notes that many respondents could not generate a simple finds list. These collections are not only digitally inaccessible; they are difficult to access on any level. The institutions included in the White & Breen study formed the starting point for identifying institutional websites for this study.

One source of archaeological collections mentioned in the Virginia report, and expounded upon by McManamon, Kintigh, and Brin (2010), are the collections at universities, usually held by professors. These collections, and associated documentation, often stay with the professor, though some get moved to an institutional repository. Using an illustration from ecologist William K. Michener, the

authors point out that the informational content of data, and metadata, disappear rapidly, and then precipitously, following retirement. Since the funding for these projects is has, in many cases, long ago been spent, as many archaeology professors retire, many associated archaeological data are in danger of effectively disappearing through lack of active management, even if the collections themselves have been archived (and this is by no means certain). White and Breen (2012) suggest that “professor and graduate student excavations have the potential to become “orphaned” with the department upon the retirement or death of the archaeologist” (p. 12). The recent development of university digital archives and repositories, which have the capacity to manage multiple data formats, may provide a home for this material. Finding this material, however, can be problematic. For while active projects are sometimes linked from departmental faculty pages, finding the archaeological projects of retired faculty can depend on knowing that the content exists.

Social media plays a role in dissemination, but more recently there has been some discussion about the extent to which emerging social media extends the archaeological record. Jeffery (2012) was perhaps over-reaching when he categorized the potential loss of social media data as “a new digital dark age” (p. 567), suggesting that digital content will be unavailable in the future. Nevertheless, he made the point that whether this content needs to be preserved should be a conscious decision. Social media has become a common mechanism for communicating with non-archaeological audiences. The Day of Archaeology project (Jeffrey, 2012, p. 566) is a good example of secondary data becoming primary data. The project asks for blog entries from around the world, detailing the activities of archaeologists on one day of the each year. Thankfully in this case the content is archived through the Archaeology Data Service (ADS), a data repository in the United Kingdom.

Blogging generally has become a means of public discourse for archaeologists and is included in considering public outreach characteristics. An informal review of Virginia archaeological blogs shows they are written and illustrated in a way suggesting they are trying to reach a broad audience.

Nevertheless they are also used for reporting recent discoveries and include data in the broadest sense (William and Mary 2015; Montpelier 2015), data that sometimes take a long while to come into the formal publishing arena. Hank's (2013) discussion of the extent to which blogs are "scholarship" is relevant here to the extent that archaeological blogs be considered data, rather than public outreach (Hank would suggest they can be both). Archaeological blogs are potentially serving multiple publics, including their professional peers.

Crowdsourcing can be seen as another form of social media, utilizing the "crowd," or an external group, to add meta-data to existing content. Dunn and Hedges (2012) presented a typology of crowdsourcing in their report for the English Arts and Humanities Research Council. Their classifications include different process types, and output types, and several archaeological projects are using crowdsourcing to generate primary content. The ACCORD project (Accord, 2014) uses different local groups to gather 3-D modeling data of monuments and places of memory. A Virginia-based project, Culture Embossed, is a database of wine bottle seals, along with its partners Culture Impressed (pipes) and Culture in Stone (points). "Culture Embossed, Culture Impressed, and Culture in Stone are crowdsourced databases of material culture from a wide array of time periods and locations" (Culture Embossed, 2012). They use a crowdsourcing technique, allowing anyone to contribute images and locations of wine bottle seals, pipes or points to the website in order to study broad cultural trends. The website collects and presents primary data, though in a way that is outside of the traditional norms, and these projects blur the distinction between public outreach and scholarly publishing.

Archaeological Data

Archaeological data are tremendously varied and have been strongly impacted and expanded by the digital age. Archaeologists were early innovators in using computers in the humanities; for example see Chenall (1971) showcasing archaeologists' early use of databases. Digitization has opened up new methods of data collection and increased the amounts of data collected. Digital data include reports, GIS

and remote sensing data (for example LIDAR, satellite imagery, ground penetrating radar (GPR), magnetometry), 3-D models, archaeometric data (such as the results of carbon 14, x-ray fluorescence (XRF) and soil chemistry testing), artifact and field record databases, environmental data, field notes, digitized maps, images, historical documents and photographs, oral history, and historic maps (Watts, 2010, p. 10). Since most archaeology is a destructive process, many of these data cannot be recreated, making digital curation—preservation and access—of vital importance to the discipline. While some of these data are born digital—images, spatial and some scientific data—some are digital surrogates, for example collection records describing artifacts. Access can mean remote analysis of these data, or simply sufficient metadata and tools to be able to identify specific artifacts for further physical analysis at the holding repository.

Archaeological data are highly contextual, so these data sets are most meaningful when integrated. An artifact's meaning is tied to many factors including the other artifacts found with it; the quadrat from which it was excavated; the features and broader archaeological site in which this quadrat exists; and the methods used in excavation. White and Breen (2102), whose report was a conscious effort to deal with archaeological data as a connected whole, noted a problem with how existing collections in Virginia are sometimes maintained:

Virginia's archaeological collections, and with this we refer not only to artifacts, but also to ecofacts, field notes, photographs, maps, reports, special samples –everything that provides data about an archaeological site – are thought of as individual, isolated entities which are dealt with in a variety of ways (p. 3) .

From a user's perspective, meaningful access to these data means maintaining, at least in part, how they relate to each other, so that potential users of the data can contextualize as much of the total dataset as possible. Otherwise, users have to re-establish these relationships themselves, cross-referencing field notes with artifacts and other data. Moreover, the different forms of archaeological

data can range from simple to complex digital objects, including relational databases, media files, text and PDF documents, GIS and shape files, and other scientific data forms. The contextual, complex nature of digital archaeological data can provide challenges for digital archival repositories, and therefore approaches need to consider the long-term archival preservation of all of these data to ensure that they are available for future research. Archaeology data, with some exceptions, are about the many, not the one. Though there are some artifacts that provide unique or particular information, archaeological data are usually quantitative with researchers looking at assemblages and the totality of the collections from an archaeological site.

Sullivan and Childs (p. 103) pose the question “If we don’t know what is in a collection, and therefore cannot use it effectively or efficiently, why keep it at all?” Despite the progress made in recent years, large parts of the archaeological record remain effectively inaccessible. Archaeological excavations are not repeatable experiments. If the data from these excavations are lost, they are not recoverable. The problems with the curation of existing collections – such as storage concerns and incomplete documentation – potentially foreshadow problems with digital data. The increasing amount of digital data being collected will not be preserved unless they are stored in archival formats and maintained with appropriate descriptive metadata.

Archaeological repositories and data publishing options

Aware of the limited options for the access and preservation of archaeological data, Kintigh (2006) called for a new infrastructure to “foster the development of a new paradigm of integrative and synthetic research.” He further presented the tools that this new infrastructure should support.

- (1) work at scales not currently possible to answer pressing questions that cannot now be addressed because of a lack of effective access to existing data; (2) foster the development of a new paradigm of integrative and synthetic research; (3) scale and integrate archaeological data so that they can be used to address compelling questions in other disciplines; and (4)

sustain the scientific utility of existing digital data that are critically endangered by media degradation, software obsolescence, and inadequate data documentation (meta-data) (2006, p. 567).

Kintigh provided a clear rationale for needed improvements: (1) make data available; (2) make them available for archaeological research questions, (3) for other disciplines' research questions, and (4) for preservation. His call to action helped lead to a significant change to the digital curation landscape with the development in 2008, of an archaeological domain-specific digital repository, the Digital Archival Record (tDAR), opened in 2009 by Digital Antiquity. Preservation is clearly a major concern for future access to archaeological data, and one that has been well discussed in the literature (Childs & Kagan, 2008; Kintigh 2006; Kintigh & Altschul 2010; McManamon, Kintigh & Brin, 2010; Sullivan & Childs 2003; Watts, 2010;). The goals of tDAR are to increase access and preservation. tDAR "encompasses digital documents and data derived from ongoing archaeological research, as well as legacy data and documents collected through more than a century of archaeological research" (McManamon, Kintigh, & Brin, 2010, para. The Digital Archaeological Record). tDAR fulfills a crucial need, magnified by the noted shift by federal granting agencies towards data accountability, and answered some of the "intertwined problems of data access, and preservation" (McManamon, Kintigh, & Brin, 2010, para. The Need for Digital Archiving). The repository is based on the Open Archival Information System (OAIS) Reference Model, and is moving towards trusted repository status. As a relatively new addition to the data landscape, the reach of tDAR is growing, but the expectations for funding digital curation are new to archaeologists and perhaps slow to develop.

While access is a primary goal of tDAR, its main audience is other archaeologists. The website serves as an index to a huge number of archaeological reports, as well as a smaller number of collections datasets, images, and related data. In this way it is helping to address the legacy of grey literature, informally published reports that had been difficult to access. As a research tool, the website's

digital objects have been enhanced with metadata to improve discovery, and there is a map interface that allows query by area, and faceted search queries through resource, collection, institution and person. While the website is likely to develop, at present it offers limited options for the non-professional; for tDAR's outreach the "emphasis at this time is on the college and graduate students" (tDAR, 2014). While growing, at the time of this writing, tDAR holds only a fraction of Virginia archaeological sites.

For archaeology, there are only two repositories currently directly referenced by the National Science Foundation (NSF) (National Science Foundation, n.d.). One is tDAR and the other is Open Context (opencontext.org) which provides access tools and archive options through a cooperative agreement with the University of California, California Digital Library. Open Context is strongly tied to open data standards and publishes its data within a Linked Open Data environment. Its ingestion process aligns data to broad standards, enabling searching across archaeological sites, and it is developing ontologies to manage archaeology's varied descriptive language. Open Context offers researcher the chance to look at archaeological data across national and international boundaries, rather than in the varied silos of specific archaeological sites. Its North American collections are focused in the Southeast and include Virginia data, through the aforementioned DINAA project (Heritage Bytes, 2012). At present it includes very little collections data.

Both tDAR and Open Context are looking at a number of funding models that include charging for the archiving of data. Though current grants are more likely to encourage funding for digital preservation, archaeology has been slow to move to support funding for digital preservation. Funding options for all projects, but particularly for existing collections, many not digitized, seem to be a future challenge and a further impediment to resource sharing.

Kansa and Kansa (2011) conducted a survey of websites as part of their white paper on "Enhancing Humanities Research Productivity." The scope of their study was worldwide, and included

few specific American references, but they included 25 websites sharing culturally-related digital heritage collections. Their critique, perhaps reflecting their advocacy for open data, noted a need for better web architecture, machine interaction and data aggregation (p. 30), while minimizing issues of “appearance and front-end design” (p. 2).

Several websites provide potential models for future projects and evolving data models. Gabii Goes Digital (<http://gabiiserver.adsroot.itcs.umich.edu/gabiigoesdigital>) is a website based on archaeological excavations in central Italy. The website features peer reviewed publication of digital 3-D models and complex, interactive datasets. In a review of the site, Nebbia (2014) notes that Gabii tries to reach both a scholarly and public audience featuring a “peer-reviewed framework for 3-D data and general greater use of digital data in archaeology” (section: A new standard for archaeological publications). In trying to engage (and help train) a community of peer reviewers, the website is seeking to allow critical review of its findings. The project itself implicitly presents the difficulties of challenging the presented narrative, at this and other websites, since it has felt the need to create training and education content for potential reviewers. Archaeologists will increasingly have to deal with these, and other, forms of born-digital data content, with related issues of data-sharing and preservation.

The regional solution provided by The Digital Archaeological Archive of Comparative Slavery (DAACS) shows the value of a project website. DAACS has become a gold standard for data sharing in historical archaeology, and received the Society for Historical Archaeology Award of Merit in 2007. The website (daacs.org) went live in 2004 and offers access to the collections of archaeological sites from the Chesapeake (Virginia, Maryland), the Southeast (South Carolina, Tennessee), and the Caribbean (Jamaica, Nevis, St. Kitts). It has received major support from the Mellon Foundation and the National Endowment for the Humanities (NEH), as well as other funding agencies. Compatibility is achieved through rigorous cataloging protocols and controlled vocabularies, with all of the archaeological collections cataloged, or partially re-cataloged, at the DAACS curation lab. Some issues of archaeological

site comparability remain, due to diverse excavation methodologies, but this “site metadata” is available to researchers when visiting the website. For archaeologists, the website offers an unparalleled set of data for studying issues of slavery. The impetus behind DAACS was to enable sophisticated quantitative analysis of site assemblages – something that can be challenging even to other archaeologists. Recently, however, object galleries have been added with more visual data - images and summary reports – and a revamped website design that makes the content accessible to a wider audience. To understand colonoware (a form of locally made coarse earthenware), a visitor can now look at <http://www.daacs.org/galleries/colonoware/> which provides examples, images, and a summary statement on this type artifact. In this way the website has broadened its reach, acting as both a repository and a gatekeeper, offering multiple views of the primary material.

DAACS is now launching the DAACS Research Consortium (DRC). This initiative will allow institutions to catalog their own data, using DAACS protocols, into a new cloud-based web interface. The first training sessions were undertaken in May 2014. While the initiative is currently limited to invited partners, the possibility exists that the results of the DRC pilot project could provide a significant new option for historical archaeology sites.

The Comparative Archaeological Study of Colonial Chesapeake Culture (chesapeakearchaeology.org/) has provided a similar approach with providing data pertaining to a number of archaeological sites within a tight regional cluster. Comparable collections information is provided at a more basic level than DAACS, but the datasets (artifacts and field records) can be queried across all archaeological sites, and are also available for individual download. Like DAACS, the primary audience is other archaeologists, though archaeological site summaries and artifact galleries provide visual examples of the material culture found, with site maps also available in number of downloadable formats. A current initiative funded by NEH will result in the doubling of the included archaeological sites.

The National Park Service supports its parks through a web catalog (museum.nps.gov), where parks can share their collections (and of which archaeological material is a substantial part). Currently there are six Virginia parks represented, including over 2,000 archaeological objects. The approach here has been at the object, rather than archaeological site level, so unusually artifacts are accessible, but not the related site files.

In Virginia the largest single repository is the Virginia state file system managed by the Department of Historic Resources. The Virginia Cultural Resource Information System (V-CRIS) allows submission (and access) to site level files on a subscription basis. In addition, there are university programs, with some material in university digital archives, and a number of historic house museums with active archaeology programs, as well as museums and other non-profits like Jamestown Rediscovery and Colonial Williamsburg that have extensive archaeological collections, and that provide varying forms of web access to their archaeological material. Some web projects have taken a data, rather than archaeological site-centric approach. For example, 3-D scans of artifacts create new data formats and The Virtual Curation Museum (<http://virtualcurationmuseum.wordpress.com>), at Virginia Commonwealth University uses a blog to showcase its scans created by the Virtual Curation Laboratory (<http://vcuarchaeology3d.wordpress.com>). It seeks to create an “extensive virtual catalog of American Indian and historic artifacts” (the Virtual Creation Laboratory, n.d.). Project participants are also active in pushing these blogs out to other social media platforms like Facebook. Limp et al. (2011) discuss the issues with 3-D objects, placing the data in what they describe as a digital ecosystem. Their comprehensive paper discusses the recording, re-use and citation of 3-D models. While the technical specifications for 3-D scanning and viewing are likely to change, the paper outlines how application of standards, and the use of accessible digital archives, can lead to “digital objects [that] are well documented, readily accessible, and available for use and reuse by others” (section: Archive and Dissemination). While the paper applies to 3-D models, this statement surely applies to all forms of

digital data. The paper references Kansa (2005) sharing his quote that “Without effective dissemination, claims and counter claims become much less meaningful, since all refer to essentially hidden, inaccessible, and controlled evidence” (Kansa, 2005, p. 100).

Morrison, Thomas and Gosden (2014) examine the use of GIS aggregate data from many projects in looking at a region in England. They suggest going beyond the use of shape files to delineate site boundaries (required in Virginia’s state database V-CRIS) to include georeferencing of all archaeological site plans, and plans of site boundaries and features along with dating information (section: Methodology). The resulting website becomes a tool for planning and research, highlighting areas of inactivity in the landscape. Presenting the absence of evidence, which is part of archaeological understanding, is always a difficult task.

Another type of non-site specific dissemination is through websites that focus on defining and illustrating different types of artifacts (types). Examples include:

- Jefferson Patterson Park and Museum: Diagnostic Artifacts in Maryland
(<http://www.jefpat.org/diagnostic/index.htm>)
- The Florida Museum of Natural history (http://www.flmnh.ufl.edu/histarch/gallery_types/)
- The Virginia Department of Historic Resources: Native American Comparative Collections
(http://www.dhr.virginia.gov/arch_DHR/archaeo_lpc.htm)

While these websites show examples of artifact types, rather than the entire primary record, they are still evidentiary. The emphasis on object images makes the websites helpful for archaeologists, as well as other groups interested in material culture. However the data in these websites are not easily re-usable. The examples listed are not “open,” so their potential for data reuse is currently limited. Using linked open data principles at these websites would enable other creators of digital materials to use the content as references.

The aforementioned Culture Embossed database uses crowdsourcing for contributions to create a database of seals across archaeological sites. The collection, focusing on a specific artifact type, allows people to contribute data with defined metadata and builds a dynamic typology. However, while an accessible resource, the reuse of the data (both from a legal perspective and through open data standards) still seems under development.

In part, the methods for data presentation and for access depend upon the audience and questions that these designated communities ask of the data. Multiple methods might well be needed to support multiple publics, though there are strong arguments for open data allowing data reuse for all constituents. Non-archaeologists might be better served by archaeological gatekeepers helping to provide local, heavily contextualized data, rather than the broader aggregated websites that can answer sweeping archaeological and anthropological questions. Scholars may require richer, open datasets with comparable metadata.

To move towards understanding the varying requirements of different publics, the websites of organization supplying archaeological data may be examined. While there is overlap, open data, scholarly publishing, and public outreach can be defined, in part, by certain characteristics that may, or may not be present on a website.

Scholarly publishing

Scholarly publishing here is seen as a subset of scholarly communication. This study is interested in those aspects of scholarship relating to licensing, citation, accessibility and re-use.

The primary audience for archaeological data is obviously other archaeologists. While researchers are increasingly interested in using data from other projects, there are disincentives to sharing data. In looking at scholarly communication, Borgman (2010, p. 193) observed that the incentives are more in data use than in data sharing, noting that “those whose data collections and analysis were the least automated and most labor intensive were most likely to guard their data” (p.

189). Archaeologists can wish to hold onto hard earned data until they feel they have gained all the benefits of their labor. Richards (2014) discussed the “carrot and sticks” of data publishing concluding that sticks – in the form of legislative requirements through grants – were the most effective tool for encouraging data sharing. The creation of digital resources is not valued in the academic environment, in comparison to other scholarly works, as noted by Harley et al (2010, p. 40) who say that “Developing and maintaining databases or resource websites is considered a “research technique” or “services to scholarship.” So there are disincentives to the creation and sharing of such resources.

Standards and interoperability are crucial for data re-use and comparative study, but, as has been noted, these are still limited within the profession. Even within a tight-knit group, working on the same sub-discipline in the same area, issues of data integration remain complex (Kansa & Kansa, 2011). Looking at zooarchaeological collections within Turkey, Kansa, Kansa and Arbuckle (2014) tried to aggregate data from fourteen researchers. The use of codes, different data capture techniques, and language (even within a well-defined sub-discipline), resulted in multiple challenges in developing synthetic research from the aggregated collections.

When forming interpretations based on the use of comparative datasets, archaeologists have to take into consideration many issues, including archaeological site excavation methodologies, which are rarely transparent in the data. The inclusion of how the data were captured is part of the data provenance showing the transformations that the data take in arriving in their accessible state, their lineage and derivation (Simmhan, Plale & Gannon, 2005). The issue of understanding archaeological site metadata is perhaps best reflected in the finding of Faniel, Kansa, Kansa, Barrera-Gomez, and Yakel (2013). Their survey showed that knowledge of who created the data, and the reputation and scholarly affiliation of the creators, was important in the decision to use, or not use, particular archaeology datasets. These factors likely reflect not just trust in their work, but possibly a previous understanding of their methods. This difficulty in judging fieldwork can be referred to as “differentiating” (Meho & Tibbo,

2003, p. 581). Meho and Tibbo's study looked at the information seeking of social scientists, building on the work of Ellis, and helped frame how potential data models might be used. As well as differentiating, they noted the importance of networking in research (p. 582). For archaeologists, where many archaeological sites are only informally published as grey literature, potential datasets are only known through professional contacts created through conferences and other professional gatherings. In Virginia this means membership in the Council of Virginia Archaeologists, and attendance at local conferences such as the annual meeting of the Archeological Society of Virginia, the Middle Atlantic Archaeology Conference, and national conferences including the annual meetings of the Society for Historical Archaeology and the Society for American Archaeology.

Falkingham (2005) provides an extensive review of grey literature proposing a model using XML markup to aid discovery. He argues the issues with this type of material including interoperability (section 2.5), noting a lack of common vocabulary as well as a lack of a common structure or style (section 1.5). These grey literature reports are written in-house and therefore are not up to the standard of published, reviewed documents. They are also written for other archaeologists, limiting their value for broader audiences. The National Park Service archaeology program created an on-line index of grey literature reports, but this no longer seems to be supported (<http://www.cast.uark.edu/other/nps/nadb>), though many of the reports are available at tDAR. These reports, it should be noted, often contain collections data in the form of tables which could be translated into more open data formats, as well as field notes, drawings and photographs.

The Internet Archaeology magazine (<http://intarch.ac.uk/>), has completed two 'LEAP' projects that allow readers the "opportunity to 'drill down' seamlessly from the publication into the archive to test interpretations and develop their own conclusions" (Winters, 2006). Internet Archaeology is peer-reviewed and now mostly open access. It has continued the LEAP approach to data and synthesis with data papers in the most recent issues featuring "open access, short, peer-reviewed publications

designed to make other researchers aware of a dataset” (Winters 2013 section: Editorial). These hybrid publications provide both synthesis and data.

The Journal of Open Archaeology Data provides another outlet for archaeology, though one that is not currently utilized in Virginia. The website presents peer-reviewed data papers with archived open data, citable and tractable (<http://openarchaeologydata.metajnl.com/about>). While touting the benefits of open data, the website appears geared to scholarship and, though a useful outlet for data, lacks the structure to allow analysis across data sets.

Open data

“Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).” (Open Definition, n.d.). Tim Berners – Lee’s call for “raw data now” (Berners-Lee, 2010) speaks to the promise of shared merged data sets opening up new avenues of understanding across and within disciplinary boundaries. Berners-Lee’s (2010) five-star rating schema provides a way of evaluating the degree of openness of website data (LATC, 2012, para: By Example).

- One star means the content is available on an open license.
- Two stars means the content is provided as structured data.
- Three stars means the content is provided as structured data in a non-propriety format.
- Four stars means the website uses URIs (Uniform Resource Identifiers).
- Five-star means the site uses other linked open data.

For archaeology, a four-star rating would mean that data can be more easily shared, and be used on comparative research projects, with the expectation of persistent availability.

The impact of this movement for archaeologists can be seen in new federal requirements stressing preservation and access for projects involving federal funds. A memo from the Office of Science and Technology Policy (OSTP) notes, “digitally formatted scientific data... should be stored and

publicly accessible to search, retrieve, and analyze.” And more directly for the goals of this research, it adds that agencies should formulate plans for “a strategy for improving the public’s ability to locate and access digital data resulting from federally funded scientific research;” (White House 2013a). The requirements have taken the form of required data management plans, seen in both the requirements for the National Science Foundation (National Science Foundation, 2013) and the National Endowment for the Humanities (National Endowment for the Humanities, 2011). Grant recipients are asked how they will manage, preserve and provide access to their data, with allowance for the norms in their discipline.

While referring to Open Government data, the Open Government Data group (Open Government Data, n.d.) provides more specific guidelines of what it means to be open, listing eight major and seven minor traits. The principles are: complete, primary, timely, accessible, machine-processable, non-discriminatory, non-propriety, and license free. Seven additional principles are included: online and free, permanent, trusted, a presumption of openness, documented, safe to open and designed with public input.

Some of these have more resonance than others within the context of archaeological data. “Complete” refers in a governmental sense to all data be made available – not selections. In an archaeological domain, the importance of complete can be seen in reference to context. The meaning of archaeological data lies in the information context within which these data exist. It seems a good guiding principle that all data sets be provided for openness – artifacts, ecofacts, field records and maps. Of the other principles, primary asks that the data be provided at the highest level of granularity, rather than in summary reports. Timely reflects on scholarly communication and disciplinary norm between discovery and data publishing. Machine-processable is included in this study through looking for open formats and associated documentation (database metadata, or even a supporting ontology). The requirement for non-discriminatory, non-propriety, and license-free all pertain to licensing statements regarding how

data can be re-used. These principles were used as guides to identifying characteristics of openness in the websites reported in this study.

Accessible is a more complicated characteristic. Content may be technically accessible, but unmanageable because of the sheer quantity of the data; or because of the knowledge needed to process the datasets. Anne Trefethen, in a video interview speaks to the general issues and promise of open data. She sees value in allowing the public “to engage with science,” though noting the need for “tools and facilities to access the data in ways that are understandable” (Bland, 2011). Simply providing data does not make it accessible.

For some professionals, the principle of open data is seen as part of professional ethics. Kansa (2012) states that since “technologies make open data feasible, the discipline should not continue to tolerate the personal, self-aggrandizing appropriation of cultural heritage that comes with data hoarding” (p. 507). For Kansa, it is not that data can technically be shared, but that professional ethics require that they should be shared, and that “data withholding represents a clear threat to preserving the archaeological record” (p. 507). It is a point mirrored by Harley et al (2010, p. 88) regarding some scholars engaging in “radical data sharing.” Open data here, and in other aspects, connects, directly to scholarly publishing and public outreach.

Public outreach and audience

Potential and existing publication types, formal and informal, for archaeological data cannot be understood without reference to designated communities. So who are the audiences for these data, and what are their needs?

Moyer (2006) was discussing physical collections when she suggested breaking the audience into two groups. Those familiar with the value of collections— such as archaeologists and “employees and volunteers of historical societies and museums, collections managers, and administrators” (p. 3) — form the first group. Her second group included “educators, college and university professors,

researchers from various fields, interpreters, artists, local communities and descendant groups” (p. 3). Moyer did not address the virtual representation of artifacts, nor did she make any comment on the issues related to using archaeological data, but her study included several case studies showing how archaeological collections are used at five different institutions by different groups, noting that the value of collections is “frequently underestimated.” While there are differences in the use of physical and virtual collections, her strong support of the educational value, and her illustrative case studies, clearly show how archaeology can reach not just other archaeologists, but the many publics she described.

Between archaeologists and a “lay” audience are professionals in related fields. Curators, historians, and digital humanities researchers, for example, may have an interest and understanding of the material culture of archaeology without necessarily being experts in the field. Haley (2007) has made the call for more user studies to understand the how and why of using digital humanities data, as well as a need to “to measure its impact and its outcomes” (section: Value and sustain ability). Her research on professors and undergraduates showed that the resources have to “mesh with faculty members’ pedagogies” (section: conclusions). It is difficult for complex datasets to be assimilated with other materials.

There is some literature in museum studies and archaeology that tries to understand the extent of the audience for archaeological data. Specifically referencing digital archaeological objects, Breen talks of an audience of “everyone from my mom to academic researchers” (Breen 2011, para: From Dig Diaries to Digitized Databases). It would be useful to explore the evidence that this non-archaeological audience exists. Are there users of archaeological data outside of the profession? One insight into the public’s general interest in archaeology: “We may need point no further than the television set to confirm that an archaeological engagement with things and everyday heritage has captured public imagination” (Mullins, 2013). Archaeologists hold field schools and public archaeology events; they publish in a variety of formats including popular magazines; they create blogs and websites. But is there

any evidence that the public's interest in archaeology extends to archaeological data? How important is it that potential data reach these broader publics?

A 2003 paper reporting on *The Publication of Archaeological Projects: A User Needs Survey* (PUNS), while dated in terms of the type of digital material now available, is still informative in terms of attitudes to archaeology (Jones et al. 2003). The PUNS survey was broad based including museum curators, researchers in cognate disciplines, conservators, teachers and students (section: The Survey Methodology). The report looked at existing publication models, at this time still heavily print-based, and broadly contrasted field reports with more synthetic material. Perhaps unsurprisingly it found that "40% of respondents also favored greater integration of description and interpretation and more synthesis" (section: The structure of fieldwork publications). However in examining "specific categories of data, museum archaeologists, postgraduates, artefact/ecofact specialists, and university staff make the most such use, with over 70% (over 90% in the case of specialists) 'always' or 'usually' using them for the purpose" (section: The Survey: Use of the components of fieldwork reports). It may be hypothesized that a layered approach providing synthesis with supporting linked data would have been very appealing to these audiences.

A survey of the United Kingdom's Archaeological Data Service (ADS) (Beagrie & Houghton, 2013), which archives a wide range of archaeological data, including site reports, provided a limited comparison, but does suggest some potential for broad-based access. Seven percent of respondents from the non-educational sector fell into the curator category, though just 4% came from related disciplines, while 3% described themselves as metal detectorists. From the educational sector, a scant 1% was from non-higher education. However, when asked what the research was being used for, 25% said general interest, 20% heritage management, 19% learning and skills development, and 15% teaching, suggesting that a broader audience for these data exists.

There have been a number of museum studies that have looked at audience segmentation (Sarraf, 1999; Peacock & Brownhill, 2007; Fantoni, Stein & Bowman, 2012). This author's study of users of the National Park Service web catalog (Freeman, 2013) attempted a more detailed breakdown, asking visitors to self-identify according to their interest in the website and, for researchers, the nature of their research. The web catalog is an online catalog showcasing some of the objects (of which 84% are archaeological) maintained by the National Park Service. Sixty-two percent of website visitors classified themselves as researchers; of this group 50% were historians/archivists, while 21% were archaeologists. Interestingly, when the objects examined by the survey recipients were analyzed, archaeology made up a similar 19% of the records examined. Looking more closely at the results showed that the archaeology material was not being looked at by archaeologists, but by historians, with the archaeologists looking at historical objects; supporting the idea of other researchers being interested in archaeological data. There is clearly a strong public interest in archaeology. The question seems to be determining the level of granularity at which archaeological data should be presented, such as whether summary archaeological site information with illustrative artifacts is sufficient. More studies are needed, but there is some evidence that there is a utility for archaeological data outside of the profession.

To understand the needs of potential audiences of non-archaeologists, it seems useful to start with the observations of Kenneth Hamma (2004). "There is no such thing as a general visitor, no such thing as someone just browsing through the on-line collections" (Hamma, 2004, section: Introduction). While Hamma was referencing museum collections, the point is equally true for archaeology. In examining collections accessibility, there needs to be a clear understanding of the designated communities and their required use, and capacity for use, of primary data. From the earlier references to audience, students and teachers in K12 are likely to have different needs than historians, planners, community, and Native American groups.

McDavid's early paper (1999) on the Levi Jordan website noted how the web changed the nature of archaeological publication and identified some of the web elements that could serve a wider audience. The goals for this website included the desire to "incorporate elements of reflexivity, multivocality, interactivity and contextuality" (section: Overview of the website). The web's support for readers to engage in a conversation – now referred to as web 2.0 – allows for the possibility of a democratization of knowledge, allowing other voices (though the website text may be seen to have a louder voice than a contributed comment). The web isn't the transfer of text to the internet, but a new medium that opens new opportunities. Social media has expanded the ways to provide engagement on the web. Blogs provide the possibility of directed dialogue, while use of social media such as Facebook, Twitter, and Tumblr provide mechanisms for exposure to materials, as well as allowing for users to comment, and even provide additional user-generated metadata for resources through the implementation of a folksonomy.

Another aspect of effective public engagement is in providing surrounding contextual material with the archaeological data. An object focus is shown in an emerging new framework at the Midden website at Mount Vernon. This website explicitly sets out to create "nuanced and holistic archaeological collections information [that] is presented to the general public" (Breen, 2011). Breen explains that the new Midden website uses an "anthropologically-informed approach to digitizing [archaeological] collections," part of a " 'comprehensive information framework for the objects (Freeman, 2011)' that will include stratigraphic, temporal, documentary, and theoretical context" (Breen, 2011, section: Mount Vernon's Archaeological Collections Online Project). The website's intention to explicitly engage audiences in primary data is unusual, as is its ability to marry historical, museum collections and archaeological data together. It also challenges the assumptions of aggregated collections, at least in terms of access, suggesting that these data need a tight thematic and regional context to be useful to non-archaeological audiences.

The extent to which open data are valuable to broader publics is explored by Beale (2012) who makes extensive reference to the open data movement in an examination of its effect on community archaeology. She stated “Open data can act as a conduit for reflexive approaches” and suggested that open data actually challenges the current practice of community archaeology (Beale, 2012, p. 626). Open data, Beale stated, are more than informational: because the data extend beyond the scope of projects; because they can be commented on and added to outside the boundaries of the traditional authority structure, they actually change the notion of community archaeology. As a contrary notion Beale also observes that, when involved in community projects, archaeologists should be aware that communities may not share open data ideals, partial community control in a project setting could actually be a barrier to releasing open data” (p.622). Similarly Kansa (2012, p. 509) quotes Christen in saying “Indigenous stakeholders in archaeology, many of whom have had difficult histories and conflicts with archaeology, may regard open data as a form of cultural appropriation.” It is possible open data can serve, subvert, or even be in conflict with the goals of its audience.

Examining the overlap between open data and public outreach raises the question of who benefits from open data. These issues are explored by Gurstein (2011), who considered access as it relates to the broader issues of open data for poor and marginalized communities. He outlined seven parts of a successful model in response to an anticipated data divide that helps people make “effective use” of open data. Two of his points seem particularly relevant to the potential audience for open archaeological data: *Interpretation/Sense making*, which he defined as “sufficient knowledge and skill to see what data uses make sense”; and *Advocacy*, by which he meant “supportive individual or community resources sufficient for translating data into activities” (Gurstein, 2011, section: A model for effective data use). If open data are to be effective, and in this case if archaeological data are to be meaningful, then the audience needs the tools to understand the data, and a means of translating them into “use” at

the local level: perhaps data to support action to preserve a battlefield, or to be applied to local, or national, education standards.

These points were further explored by Cole (2012), who also examined problems with open data. His reference to “sources of error” seems to reflect the difficulty of understanding the problems inherent in any data set, which include domain level assumptions. All professional archaeologists have an understanding of the strengths and limitations of their data – accuracy, sample sizes, comparative analysis, and absence of data - that comes from their professional training. For those outside of archaeology who lack formal training, these data can be misunderstood, or hard to analyze. Cole’s point about the “Resources Divide” can be seen at a basic level – lack of access to computers, or the internet – or at a more complicated level, with users of the data lacking the mental “tool set” to interpret datasets. For public and scholarly use, the work of Purdue University Libraries Data Curation Profiles may be helpful (<http://datacurationprofiles.org/>). The profiles provide extensive descriptive metadata including describing the research goals of projects.

More work is required to understand the needs of this “professionally-related” community regarding archaeological data, but a white paper conducted by Kansa and Kansa as part of an NEH study is very helpful start (Kansa & Kansa, 2011). They looked at how their target group worked through their professional goals, and the problem they had using humanities datasets. They concluded that multiple solutions were needed, but that this audience had similar needs to archaeologists: better standards and data interoperability. Archaeologists’ lack of a standard vocabulary presents strong semantic difficulties, making comparative data difficult to process as a result of different datasets using different terminology. Since this problem is unlikely to be resolved quickly, they suggest instead that data publishers focus on “data portability” (p. 11), publishing data in simple open formats, as a way of making data accessible, rather than waiting on a solution that addresses all issues. This solution is likely to also help with issues

of cost, though it shifts the burden of understanding the data onto the researcher, rather than the data publisher.

Professional archaeologists in teaching roles are aware of the need for better data literacy in access and using data for research. Agbe-Davis et al. (2014) discusses digital data being an under-utilized resource, in a paper explaining how the Digital Archaeological Archive of Comparative Slavery provides material for graduate and undergraduate studies' to "teach the quantitative methods and archaeological theory that are essential for deciphering the archaeological record" (p. 838). The degree of literacy needed to understand the data is a concern voiced by Galle. While the "demand for data has risen" there is a concern that there is a lack of "analytical skills to make sense of the data" (Galle, 2012).

Summary

Understanding effective archaeological dissemination requires referencing literature from archaeology, but also archival practice and information studies. Recent studies and new digital repositories provide opportunities for archival options. Community and professional studies point to the difficulties of managing defined multiple publics. Unsurprisingly, given the emerging nature of the digital record, there is no existing general template for dissemination of digital archaeological data. There are a variety of websites providing different types of access, but the picture is complicated by an existing crisis concerning physical collections and by increasing amounts of digital data and introduction of new data forms. Issues of interoperability and preservation persist. The emergence of national and regional solution for dissemination seem promising, but at present these include a limited number of projects.

Borgman, while questioning the value of open data, points to archaeology as one of the "exemplars which remain the exception rather than the rule" (Borgman, 2015); though in this short article she does not provide examples. At this point there do not appear to be any studies systematically studying how archaeology data are disseminated through definable characteristics. The three approaches taken in this study – open data, scholarly communication, and public outreach – while

overlapping, touch on many of the issues of effective dissemination, and it is hoped that this study will help fill the gap in the literature.

Chapter 3 - Methods

This study examined how Virginia archaeological digital data were made accessible. The research questions are:

- To what extent are websites that present archaeological data from Virginia projects aligning with the requirements of open data?
- Are there characteristics supporting aspects of scholarly publishing?
- Are there characteristics supporting aspects of public outreach?

Conducted through a content analysis of websites featuring archaeological data, the study identified characteristics relating to open data, scholarly publishing and public outreach.

Content analysis

Content analysis allows the investigation of the content of text, and is defined by Neuendorf as “the systematic, objective quantitative analysis of message characteristics” (p.1). While content analysis was developed for study of more traditional “written” documents, it has come to be used for the analysis of websites; to look for the presence, or absence, of specific types of content. McMillan (2000) provides a long list of studies using content analysis for websites as does Car et al., (2012). Neuendorf defines a number of characteristics of content analysis studies, relating such studies to the scientific method (p. 10-13). These include objectivity, reliability, a priori design, validity, and replicability.

Reliability and replicability can be difficult issues when analyzing websites. Connected to these issues, as noted by Weare & Lin (2000), is “unitization,” referencing the often amorphous boundaries of website content. It can be difficult to identify what web pages will be examined, as websites can be voluminous and connected to other content through hyperlinks (p. 10). For this study a navigation structure was constructed to look for archaeological content (see data collection below). Characteristics that were not present within this structure were not recorded. Also, in some cases, the archaeological

data were not found on the institution website, but linked directly from another website. It was assumed that these institutions intended for these data to be accessible, so the data were treated as present on the host site.

Replicability is also affected by coding (Weare & Lin, 2000, p. 17). Since website content is dynamic and fluid, website content can only be said to be present at the time of the study. McMillan (2000) suggests that “fast-paced Web almost demands that data be collected in a short time frame” (p. 92), and the data collection for this study was done within a two-month period by a single collector. Although replicability is difficult for website studies, it is helped by the definition of the tools used for data collection. For this reason a single version of a browser was used for all websites analyzed.

The study examined manifest content (the text as written), since with a quantitative study of this nature, it is inappropriate to assume the intent of the website author(s). This study recorded the presence or absence of characteristics, rather than the interpretation of text, through coding by the data recorder. Looking at websites in this way also fulfilled some of the research goals of objectivity. The characteristics were defined ahead of the study, and includes the idea that “content analysis is nonreactive” (Neuman, 2013, p. 362); the characteristics investigated were not known to the website creators.

Population

Websites were identified from an exhaustive inventory of Virginia archaeological projects. To identify sources of archaeological data, the Council of Virginia Archaeologists (CoVA) survey of collections was used as a starting point for the sampling frame (White & Breen, 2012). The CoVA survey was sent to 135 institutions in September 2010, including the “Archeological Society of Virginia (ASV) chapters, local governments with preservation programs, universities, CRM firms with offices in Virginia, and military installations.” For sampling frame construction the websites for these institutions were initially identified by using the Egloff Atlas of Virginia Archaeology

(<http://www.dhr.virginia.gov/atlas/EAtlas1.html>), which was one of the starting points for the CoVA survey, and is an authoritative listing on the Virginia Department of Historic Resource website. For institutions included in the CoVA survey, but not on the Egloff Atlas, the website address was found through a Google search. Additional archaeological organizations not included in the White & Breen survey were identified based on authoritative sources, supplemented by this author's knowledge of Virginia archaeology as an associate member of CoVA, and with connections to Virginia archaeology for the last twenty-five years. These included National Parks, along with some organizations that have purely digital archaeological content not represented by physical collections (and therefore not considered by White and Breen).

These websites provide an exhaustive sample of archaeology from the State of Virginia. Included for analysis were 150 potential websites. A full list of websites is included in Appendix A.

Data Collection

Analysis came from the gathered data by using deductive measurement. The characteristics to be examined were established a priori (before looking at the websites). They are objective, and replicable, and apply to all websites in this study. The unit of analysis was websites. Since characteristics could occur over multiple pages from the websites, and since websites are fluid constructs, URLs and some screen shots were saved to show where the data elements were found. The manifest content of the websites were first examined to see whether they provided any archaeological web pages. Websites that have archaeological content, but no archaeological data (as defined below), were only coded for the public outreach section, as defined below. Because website characteristics relevant to the study may be spread across webpages, the following structure was used to examine the websites, looking at navigational links for the whole website or archaeological sections:

Home page

- Archaeology
- Exhibits/research/collections/science
- About / Copyright
- Footer section

The characteristics recorded are as follows, sectioned into areas that reflect their primary associations with open data, scholarly publishing and public outreach. The codes for each field are described fully in the coding system (See Appendix B). The coding system was based, in part, on an instrument developed by Deblauwe, in the “Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment” white paper (Kansa & Kansa, 2011, p. 27). Deblauwe reviewed 68 archaeology websites from February 2009 to December 2010. Her criteria examined issues of permanence and licensing, and a separate section focused on evaluating the user experience. Since this study is not concerned with usability, these questions were concatenated into issues of accessibility, and reuse, asking if the data can be accessed without the need for a login, or through navigating web forms. The coding system is also informed by the eight principles of Open Government Data (Open Government Data, n.d.).

Most data-capture fields record nominal data, the presence or absence of a characteristic coded as a 0 or 1 (such as presence of site reports, contact information, data dictionary, or blog). Some fields required more values. The licensing information was recorded in comparison to the commonly defined CC license (with exemptions noted); the type of data reflects the criteria of the five-star open rating discussed earlier, and access is recorded as open, requiring a sign in, or subscription to get a sign in.

For the purpose of this study archaeological data, discussed earlier in chapter 1, was defined by multiple categories:

- Site Reports: webpages/ link(s) to a document showing the site report, which is a standard output of archaeological projects, often published as grey literature. While including analysis and synthesis, site reports can also include artifact data tables and field records.
- Field data: including digital maps, profile drawings, and site photographs.
- Artifact data: including finds lists, artifact inventories and databases.
- Spatial data: including GIS shape files.
- Data visualizations and models: including 3-D models and virtual environments.

For websites that did not include any reference to archaeology, only ID, date reviewed, organization name, URL and type were recorded. These websites were not eligible for further coding. For sites that had clear archaeological content but no archaeological data, public outreach fields only were recorded. Websites with data had all fields coded, shown in Table 1 (below).

Table 1: Data collection fields by section

	Field
General	ID
	Date reviewed
	Organization
	URL
	Institution Type
	Archaeology Online
	Data Online
	No. of archaeological sites
	Basis
	Primary Interface
	Primary data Level
Data Types Present	Site reports
	Field Records
	Collections
	Spatial Data
	Virtual Data
Scholarly publishing	Citation
	Licensing
	Permanent URIs
	Permanence/Archiving
	Contact Information
Open Data	Type of data
	Access
	Accessibility
	Thesaurus/Data dictionary
Public Outreach	Site Summaries
	Blog
	Social Media
	Contextual Information
	Media
	Community Outreach
	Educational Content

All data were collected in an Excel spreadsheet between December 2014 and January 2015. Field level data validation was used to maintain control through the use of data validation functions in Excel. Data normalization was used to verify values at the end of the data capture phase, making sure that all fields were completed with acceptable values. All websites were examined using a Google Chrome browser, version 39.

Metadata were applied to the spreadsheet, and the spreadsheet data will be made available, as a delimited ASCII file, along with this study through TRACE (at The University of Tennessee) and academia.edu, following open data principles.

Data Analysis

After the data were cleaned and normalized, descriptive statistics were generated in Excel. Results were tabulated within each of the conceptual areas. The analysis allowed quantification of how many websites satisfy the requirements of openness and exhibit characteristics in support of scholarly publishing and public outreach. The values in “type of data,” in conjunction with the availability of URIs, were used to enable the calculation of the previously discussed five-star rating openness rating (LATC, 2012), with additional consideration of access and metadata.

Scholarly publishing is informed by the number of websites including citation and licensing information, as well as the interoperability of the data and tools for analysis. The types of data available are important here as well—websites that supply more types of data (site reports, fields records, collections, other) will more strongly support scholarly publishing. Additionally, the number of websites using Uniform Resource Identifiers, allowing linking to resources from other websites, was analyzed as well as the number of websites providing statements on archival policy.

Public Outreach evaluated the presence of contextual information (broader history of the archaeological site, or related themes), as well as supporting media (video, photographs). Along with websites showing only summary data, the presences of community outreach information, and the ability

for users of the website to contribute through comments or tagging of objects, are indicators that the website supports public outreach. Since these characteristics could be seen as incremental, the more characteristics present, the better support for public outreach.

Chapter 4 – Results

As described in the Methods section, 150 potential websites were identified. When these were examined, two institutions from the sample did not have a website. The remaining 148 fell into three clusters (see Table 2/Figure 3): those with no archaeology pages; those with only summary reports, or broad discussion of archaeology; and those with archaeological data.

Table 2: Clusters by presence of data

	Count	%
No Archaeology pages	109	73.6
No archaeology data	13	8.8
Archaeology data	26	17.6
Total	148	100%

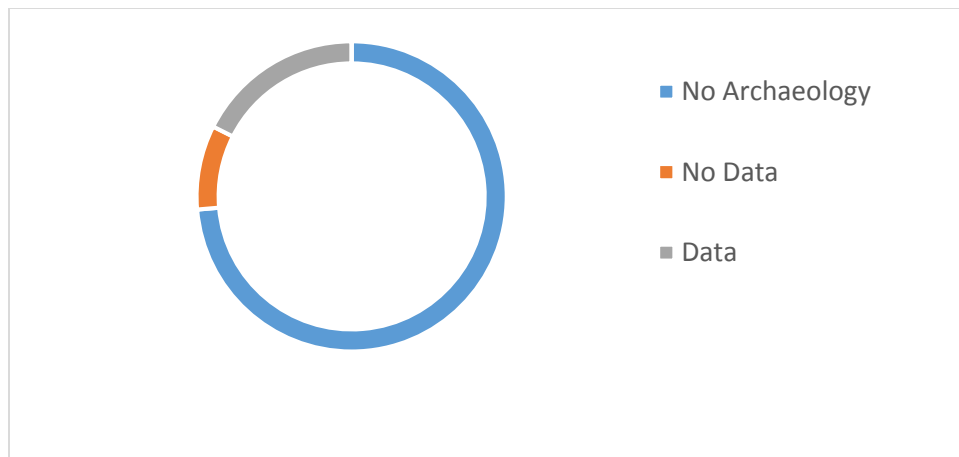


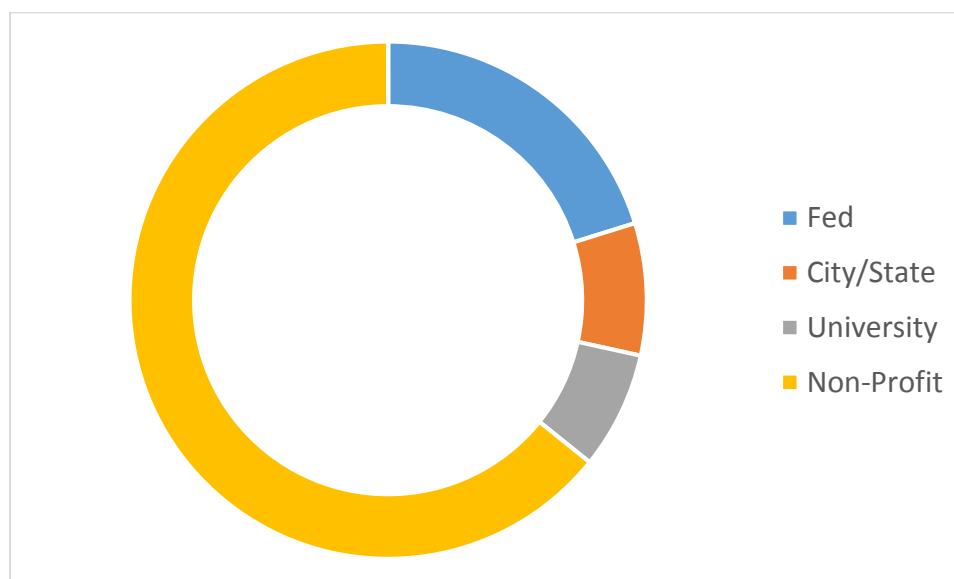
Figure 3: Clusters by presence of data

“No Archaeology” web pages

Of the websites examined, 73.6% (n=109) were identified as having no archaeological webpages. For these websites only basic information was recorded such as the review date, organizational type and the URL.

Table 3: No archaeology pages by organizational type

	Count	%
Federal	22	20.2
City/State	9	8.3
University	8	7.3
Non-Profit	70	64.2
	109	100%

**Figure 4: No archaeology pages by institution type**

Non-profits made up 64.2% of these websites (n=70) (see Table 3/Figure 4). These websites typically had research, collections, and/or exhibit information, but archaeological work related to the institutions was not identified on the websites. The Rosewell Foundation, for example, mentions “preserving, studying, and presenting this historic ruin” (Rosewell Foundation, n.d.) but provides no mention of archaeology, or archaeological reports that have informed the study of this place. Similarly Endview Plantation has a history and architecture section, but archaeology is not mentioned as

contributing to what is known about the plantation (despite the site being identified as having archaeological collections).

Included in this group of “no archaeology web pages” were 10 websites that did make brief mention of archaeology. The Manassas Museum system pages for Mayfield Fort mentions “Prehistoric artifacts uncovered on the site” and “Archeological investigations” of the Fort (City of Manassas, n.d.). The Northumberland County Historical Society mentions “archaeological artifacts dealing with Fort Augusta, from the French and Indian War through the American Revolution, items of significance to the history of Northumberland County” (Northumberland Historical Society, n.d.), and “changing exhibits on topics of historical and archaeological interest.” The Museum of the Middle Appalachians mentions an exhibit titled “Woodland Indian artifacts” (www.museum-mid-app.org/native-american-villages.htm) but provides no further details. The only archaeological reference by the Forensic Science Institute at Radford University was a mention of an archaeological field school.

Archaeology content, but not data

Of the 148 websites, 39 contained archaeological web pages. A third of these (13) had no archaeological data, as defined for this study. These websites were examined only for characteristics of public outreach, along with the 26 websites (17.6%) that included archaeological data.

Public Outreach

The characteristics of public outreach were identified for those websites with and without archaeological data (n=39). One or more of the characteristics of public outreach were found in all of these sites. Appropriately for the archaeology, contextual information was provided by nearly 80% of the websites. Most websites (70%) have taken advantage of the rich visual potential of archaeological excavations and artifacts by including a range of media such as photographs and video. The other two characteristics that appear on the majority of websites are outreach (58%) and summary information

(55%). Community outreach counts are, in part, due to the prevalence of field school opportunities in archaeology, as shared through the website. Twenty-one (55%) websites provide archaeological site summaries; communicating information about the archaeological sites in a format written for a non-archaeological audience (see Table 4/Figure 5).

Table 4: Public Outreach Characteristics

%	Educational	Outreach	Media	Contextual	Social Media	Blog	Summary
Present	15	22	26	31	19	17	21
Present %	38.5%	56.4%	66.7%	79.5%	48.7%	43.6%	53.8%

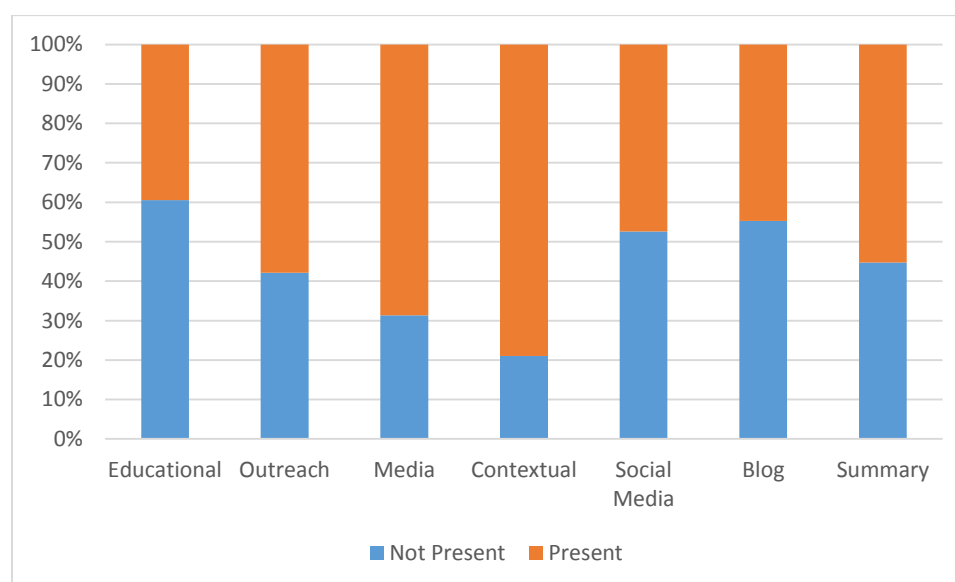


Figure 5: Public Outreach characteristics

Blogging, 43.6%, (n=17) and other social media, 48.7% (n=19) were used by about half of the websites. However, not all of the institutions had active archaeological programs; some reflected content from previously undertaken projects. It may be that social media use and blogs reflected

current or recently completed programs. Almost all the social media was Facebook, though Jamestown Rediscovery had a YouTube feed, and Montpelier included a social media bar with further options. Though not recorded for the survey, it was observed that most blogs had recent content, and appeared to be aimed at a broad, rather than specialist demographic. Explicit educational modules appeared on 38.5% of websites (15). The educational content clustered towards those websites with archaeological data, with only two websites among the non-archaeological data websites having an educational component.

The public outreach characteristics could also be seen as incremental (the more characteristics present the stronger the support for public outreach). On average, websites across clusters evidenced 3.97 out of 7 of public outreach characteristics. By cluster, they average 2.66 for non-data websites and 4.58 for websites with data. Institutional websites that include data are also presenting a greater number of public outreach characteristics (see Table 5).

Table 5: Number of public outreach characteristics.

	Average (out of 7)
Websites with archaeology pages	3.97
Archaeology pages but no data	2.66
Websites with archaeology data	4.58

Looking at the individual websites shows a range of approaches. The George Washington Foundation manages Ferry Farm and Kenmore, and the website had extensive information on the archaeological sites. There were detailed archaeological site summaries, including pictures of artifacts, excavation units and related material, clearly described and written in accessible text. Though they only hit two elements, there was clear and successful intent to support public outreach.

Of the non-data websites, Fairfax County website hit five elements, and their website notes that their collection contains “Upwards of three million artifacts (archaeological discoveries such as spear points, pottery)” (Fairfax County, n.d.) The Fairfax database is not online, but there are a few items shown on the webpage, including a bottle seal. Information is also provided on using the collections for research “Access to either the Historic Object Collection or the Archives Collection for research purposes may be arranged” (Fairfax County a, n.d.). One of their projects, the Colchester Archaeological Research Team, maintains an active blog. Similarly Fort Lee provides information on visiting the collections - “half a million artifacts” (Fort Lee, 2014), as well as an overview publication of the type of material that has been found. The Fairfield Foundation is an example of a site with no archaeological data but a clear intent for public outreach, and their blog was active and full of content.

Archaeological Data

There were 17.6% (n=26) websites that included defined archaeological data. For these websites all characteristics were recorded.

Site reports were the most common form of data represented, appearing on 69.2% of websites (see Table 6), followed by collections data on 57.7% of websites. Field records appeared on only 23% of data websites, and spatial data on only 15%. There was only one website that presented virtual data. Only two websites presented four of the five data types, offering the broadest ranges of data of the websites reviewed; none evidenced all five. One had all but virtual data, and the other all but collections data (see Table 6).

Table 6: Types of Data

	Site Reports	Field Records	Collections	Spatial	Virtual
Present	18	6	15	4	1
Present %	69.2	23.1	57.7	15.4	3.8

Open Data

If we consider open data, just over half of these websites shared data textually as shown in Table 7, usually in the form of site reports through a linked PDF document, but also as simple html pages. Of all websites in the data archaeology cluster (n=26), only 42% (n=11) had re-usable data. Six of these websites included a data dictionary, providing database metadata and detailing how the database fields were used, allowing for greater interoperability. Looking at the total institutions, only 7.3% (n=11) provided easily re-usable data (see Table 7/Figure 6).

Table 7: Open Data by type

		%	% non-proprietary with Data Dictionary
Text	14	53.8%	0
Structured Text	1	3.8%	0
Non-Proprietary	11	42.3%	54.5%

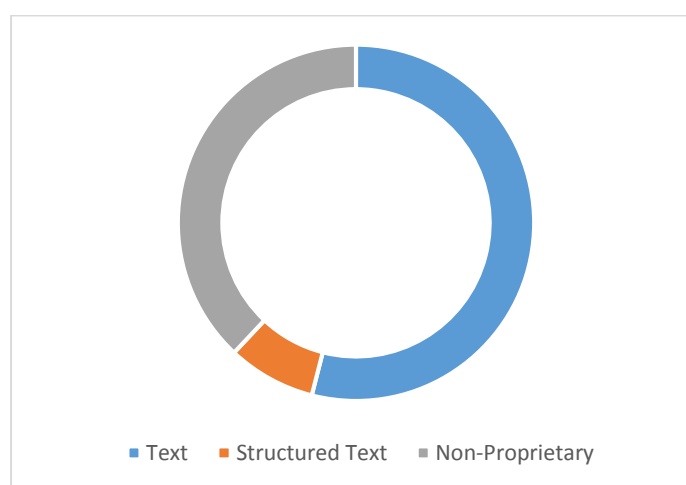


Figure 6: Open data by type

When assessing these websites (n=26) with the five-star open (defined in chapter 2) rating suggested by Berners-Lee (LATC, 2012) 11 websites reach level 3 (see Table 8).

Table 8: Five Star Rating

	Rating		License	URIs
	Count	%	Count	Count
1 Star	14	9.4	1	0
2 Star	1	0.7	0	0
3 Star	11	7.3	4	0
4 Star		0	0	0
5 Star		0	0	0

Using the stricter requirement of requiring an explicit licensing statement with the data for “openness,” then only two websites qualify: Colonial Williamsburg (1 Star), and the DAACS related websites (3 Star). Further limiting the concept of openness, no websites used URIs, nor were there any examples of linked open data (4-star and 5-star requirements).

Accessibility had been included in the survey because of the possibility of finding 3-D data or GIS data that required browser plug-ins. The virtual interface for Colonial Williamsburg did require a plugin, but this was only a small part of the overall archaeology information that the website presented. The only website that required a login/subscription was the Virginia state site repository, V-CRIS.

Scholarly publishing

The websites in the archaeological data cluster (n=26) were assessed based on five characteristics reflective of scholarly publishing. The discussion on open data noted the limited instances of explicit use statements, or licenses at 19.2% (n=5). Citation information was provided on only 26.9% (n=7) of websites. Not all sites (61.5%, n=16) provided easy access to contact information, which was defined as a link or email address to a person connected with the digital or physical collections (rather

than a generic webmaster email). There was no use of URIs (0) and no statements on archiving data were found on any of the websites (see Table 9/Figure 7).

Table 9: Scholarly Publishing

	Citation	License	URI	Archiving	Contact
Present	7	5	0	0	16
Present%	26.9%	19.2%	0%	0%	61.5%

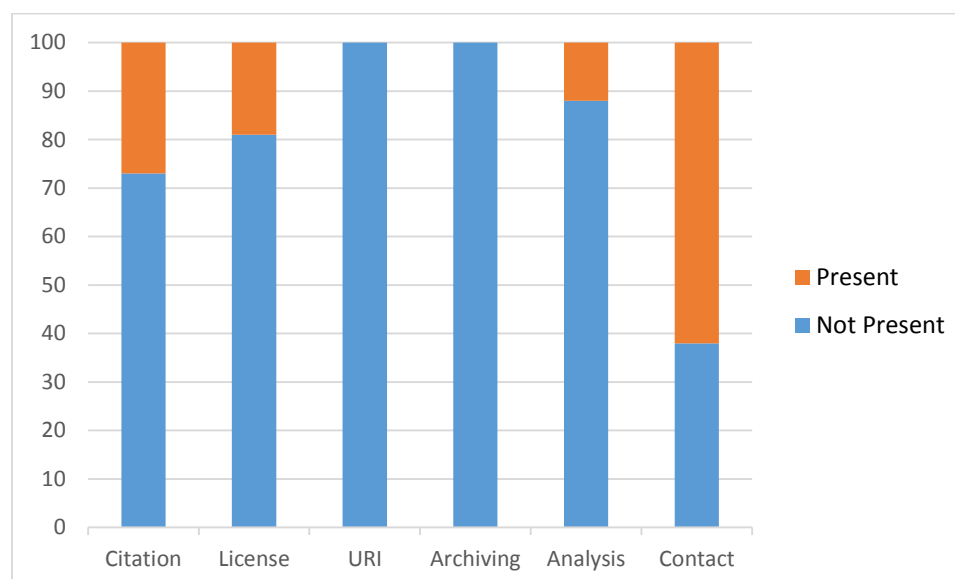


Figure 7: Scholarly publishing characteristics

Archaeological data sub-clusters

One other distinction can be made within the websites with archaeology data cluster. Two websites, DAACS and Comparative Chesapeake, are aggregate sites offering data from a range of institutions. These can be contrasted with 7 institutional websites, such as Alexandria Archaeology Museum and Colonial Williamsburg, which offer a number of archaeological sites within an institution, and 17 individual websites that show data from one, or a few strongly connected, archaeological sites.

Chapter 5 – Discussions

Overview

The analysis of the Virginia websites shows an overall lack of access to archaeological data. Of the 148 websites examined, nearly three quarters had no web pages about archaeology. It is possible that this archaeological information appears in non-digital formats (e.g. brochures), but for many institutions, archaeology is not part of the public discourse showing up on online exhibits, research areas, and descriptive histories. This absence means a lack of access to a potential set of data, and also a lack of visibility for how archaeological scholarship informs our understanding of the past.

Of the websites that did make mention of archaeology, some provided only a public outreach component, providing no archaeological data. Perceptions on the effectiveness of providing such data was not within the scope of this study and hence not considered in this analysis but, as previously noted, other scholars have pointed out both the breadth of the audience for such material, and its potential uses (McDavid 1999; Richards & Robinson 2000; Moyer, 2006; Breen 2011; Kansa & Kansa, 2011; Beale 2012; Beagrie & Houghton, 2013). In the light of increasing demand for access to data generated by public funds (and though the funding for archaeology is complex, much of it comes from public funding), this, again, seems like a missed opportunity.

The websites providing archaeological data were split between those that included only textual data, and those that made data available in a re-usable, structured digital format. The latter serve as a demonstration of how well data sharing can be done. In particular, the Digital Archaeological Archive of Comparative Slavery (DAACS) and A Comparative Archaeological Study of Comparative Chesapeake Culture websites offer support for scholarly publishing, open data and public outreach across a range of archaeological sites. Before discussing the implications of these examples and other study results in more detail, it is necessary to discuss some of the study's limitations.

Limitations

This is a quantitative study looking for the presence or absence of characteristics of open data, scholarly publishing and public outreach. What this study cannot do is speak to the effectiveness of those characteristics. While the degree of data “openness” can partially be quantified, the re-usability of these data, and their usefulness in research is complicated by many factors, some of which are addressed in the literature review. Similarly with regard to scholarly publishing, the presence of supporting characteristics can point to websites being more or less supportive, but they cannot speak to how effective these characteristics are in fostering scholarly communication. Public outreach as measured here has been discussed purely in terms of web-based outreach. The analysis did not look at how effective engagement was through social media – it just noted the presence of blogs and other social media tools. Richardson (2014) has discussed extensively whether social media tools have led to a more reflexive archaeology, concluding that the presence of such tools has not changed the approach of “top-down’ public archaeology” (section: discussion). Similarly the usefulness of contextual, media-rich presentations has not been evaluated; this study has just recorded their presence or absence -- though contextual information and understanding can be seen as integral to the discipline of archaeology.

Web content changes quickly, and new tools and policies affect content and presentation. Accordingly this review is time sensitive, and reflects only the state of affairs between the end of 2014 and early 2015. Even during the period of this study, websites were changing and new content was added.

This study looked at Virginia archaeological projects through institutional websites. While federal requirements and guidelines for grants apply across the United States, differing state environments have led to differences in how data are published and made available. Similarly, the differing levels of public land and the differing cultural environments have led to differences in the types of archaeology conducted and the questions asked. White and Breen’s study (2012) makes a brief

review of other states, noting that Maryland has a centralized repository for collections (though it does not include 100% of archaeological sites). In Arizona, collections are housed in three of four main repositories, while New York and Washington are similar to Virginia in the decentralized nature of archaeological collections (p. 6). Childs and Kagan (2008) identified 221 repositories as part of their national study. As noted in the literature review, the political and financial environments in the rest of the world are different, and have led to different models for data publication, with many countries having a state-supported national archival repository for archaeology data.

The number of archaeological sites presented on the websites was recorded, but not used when reporting results. It was hard to determine how many archaeological sites were covered through content analysis, and the definition of archaeological site wasn't clear on the institutional websites. What is also unknown is whether institutions that presented archaeological data did so for all the archaeological sites under their management. If not, as seems likely, then this study undercounts the amount of archaeological data available, and the problems of access to archaeological data may be greater than it appears from this study.

This content analysis only examines the manifest content - what was visible on the websites. To enhance the reliability of the study, only material that was clearly linked through the website navigation was examined, and other content might have been missed. This seemed to be particularly problematic on University and College websites. For example, The Center for Historic Preservation at Mary Washington has conducted many archaeological investigations over the years but its archaeology lab site (<http://almw.umwblogs.org>), was not findable through this study, since it was not part of the University website. Also the study did not consider the motivations of archaeologists in making (or not making) data available. Indeed it is likely that archaeologists were only a part of the decision about what data went on the institutional website, and they are likely competing for space and visibility. The analysis showed what is, not what was, intended.

There are other forms of archaeological dissemination not covered in this study. Traditional means of scholarly publishing, such as through journals, conference papers, and books has not been included, except where those materials have been made available on the institutional websites (though generally these formats have been slow to include data as part of the dissemination). Furthermore, archaeology has a long and rich history in public outreach. For example many historic houses in Virginia with archaeology programs hold field schools for K12 educators and students; they host open house days focused just on archaeology. Alexandria Archaeology Museum, among others, has a long and distinguished history of public engagement. There is a tremendous amount of work going on that leaves little digital footprint, and concatenating this varied public outreach into a summary “outreach” field clearly underrepresents the many ways in which archaeologists engage with multiple publics.

Finally the study examined the websites of Virginia-based organizations for Virginia archaeological data. Data from Virginia sites made available in other states, for example through museum collections, would have been missed. Reference was made in the literature to the two such institutions (two national repositories).

Scholarly Publishing

Returning to a discussion of the findings, the relatively small amount of accessible, re-usable data obviously has implications for scholarship. Research on Virginia archaeological sites runs the risk of being based on a small range of datasets, chosen for a variety of reasons outside of their research potential. Of the thirty-three DAACS archaeological sites listed in the Chesapeake region, seventeen are from Monticello, home of U.S. president Thomas Jefferson. Adding in another two archaeological sites from Jefferson’s plantation house, Poplar Forest, and 58% of the sites are connected with Jefferson - a peculiar bias in the dataset. Hopefully, over time, more archaeological sites will be added but, at present, research rests on a relatively small amount of data.

The state V-CRIS system does provide a large dataset of archaeological sites across the commonwealth, though most of these data are at the archaeological site level. However, access to the database is not free. This fee-based structure makes the dataset useful for project planning, as cultural resource companies often have user accounts because of their need to submit data. Access may be more difficult for researchers and educators, since universities need to have accounts, though free short-term accounts can be made available for research, as it was for this study. Access is extremely limited for non-professionals, though considerations of protecting archaeological sites are part of the reason for this restriction.

The lack of use of URIs presents a concern for the long-term accessibility of the resources. Research on the persistence of web addresses shows that most website links have a limited life span. For example, Prithviraj and Sampath (2014) investigated the availability of web addresses for papers published in LIS conference proceedings. They estimated the life span of URLs at less than five years, and over 50% of URLs examined were unavailable at the time of their study (p. 45). The lack of access here suggests long-term concerns about the persistence of web-based content. This concern is mirrored by the lack of information on permanent archiving on the websites. This omission may not reflect a lack of institutional policy, but without an explicitly stated policy posted to respective websites, data users must contend with uncertainty about how long resources will remain available. The availability of the national archive (tDAR) could fill this gap – and some of the Comparative Chesapeake archaeological sites are also included on the tDAR website - but there is limited Virginia content in tDAR at present. Some of the websites examined seem no longer active (no recent content, content dated from several years ago), and there must be some uncertainty as to how long content on mothballed websites remains available. Issues concerning the preservation of digital archaeological data are being discussed in the archaeological community through the previously mentioned “Caring for digital data in archaeology” book by the Archaeology Data Service & Center (ADS) for Digital Antiquity (2013), and through

information provided on the ADS website and presentations by tDAR (<http://www.tdar.org/why-tdar/preservation/>). Further research is needed to understand the extent to which digital preservation is or should be a concern among Virginia archaeologists, but it seems very useful to return to a quote from White and Breen (2012): “There is much to be done within the professional community to ensure that the tangible remains of our archaeological research are preserved for the future” (p. 11). Tangible clearly includes digital assets and data.

Few of the websites provide sufficient bibliographic information for citing, or specific recommendations for how to cite webpage content. Though the general rules of citing websites of various style conventions apply, these can be problematic without persistent URIs, and authorship is often unclear. DAACS’s clarity on citations is helpful and a model for other websites (<http://www.daacs.org/about-the-database/citing-query-results/>). Despite the easy availability of Creative Commons licenses, only a few websites make explicit statements on how the data can be re-used. As stated by Kansa and Kansa (2011), data interoperability needs to be matched with “legal” interoperability (p. 12). Explicit licensing statements on use and re-use provide clarity, rather than the assumption that data are free to use. Costa (2014), in a wider exploration of this point, suggests “Using anything else [creative commons open license], even for content or data that is available for downloading, is not unlike putting it on display without allowing bystanders any actual interaction.” Making explicit statements on licensing, archiving and citation seems achievable, and would help considerably in facilitating use and attribution in other research endeavors and scholarly publications.

In a 2014 survey of members of the Society for Historical Archaeology, over 50% of respondents rated the use and importance of digital archaeology collections as very, or somewhat important (Society for Historical Archaeology, 2015). Similarly a national survey of repositories in 2010 asked “How important is the collection/curation and access to digital data to your organization?”, with nearly 50% answering highly important (Watts, 2010, p. 7). Kintigh and Altschul (2010) note in their exploration of

the need for a trusted archival digital repository for archaeology, “As consumers, archaeologists genuinely desire effective, on-line access to the results of others’ work” (2010, p. 270). There are reasons for the lack of sharing as reported in this study’s results. The discussion by Harley et al. (2010) focuses on the difficulty of academic scholars in getting recognition for creating digital resources and asserts that in judging scholarship, it is “difficult for review committees to assess the audience and impact of work published in non-traditional formats” (p. 41). Only 4 of the 26 “open data” institutions, were universities or colleges. Perhaps, as Richards (2014) suggests, we need more carrots and sticks, or, as Kansa (2012) proposes, for data sharing to become a more explicit part of archaeology ethics. It might be concluded that the desire for using digital collections is not yet matched by the desire to create or disseminate them, and clearer guidelines on what constitutes a scholarly interoperable dataset might be helpful.

An example of the access policy posted on the website for the collections at Alexandria Archaeology Museum, an institution known for its public archaeology program, is useful for comparing online access to physically visiting the collections:

“Scholarly and Professional Access: Students and scholars requesting access to the collection shall submit a proposal of their planned research project, including collections and data to be examined, objectives, publication information, proposed timetable and schedule of hours to be spent at Alexandria Archaeology.”

“Public Access: Members of the public not engaged in approved research will not be permitted to handle artifacts in the collection without the permission of the Collections Manager.”

(Alexandria 2013, section: Collections Policy)

Such restrictions to physical materials are understandable, particularly in the light of resources need to provide access, and the potential consequences from the physical handling of materials. But the policy also shows the potential of digital access for providing resources to both scholars and a broader public,

and reducing wear and tear on materials. Moreover, the lack of specific contact information found in this study suggests a potential problem for researchers wishing to follow up with questions, and requesting access to physical collections.

Open Data

As positioned in this study's conceptual framework, closely tied to scholarly publishing is open data – the availability of web-based data for re-use. Twelve websites were recorded as offering structured open data. One of the few comparable studies by Kansa and Kansa (2011), found 25% of websites have data “easy to reuse” (p.7), compared to the 7.3% of the total websites with reusable data in this study. The Kansa and Kansa study reflected websites from Iraq and other ancient cultural heritage sites, and both studies suggest that there is much more to be done to develop and facilitate data sharing.

Three issues are particularly important in considering open data. First, not all websites published the respective database's metadata, and because of this omission it is hard to know how fields are being used, such as what data types are acceptable, and what data values are allowed or restricted. This information is essential in understanding the data. As already noted by Galle (2014), the database metadata information is one of the most used sections of the DAACS website. The development of website thesauri, relating the specialist archaeological language, could serve a broader public, while the development of ontologies would serve scholarship through codifying the institutional knowledge of the collections and their contexts. At present there are few external guidelines for data standards, and as previously discussed, data interoperability remains an issue. Publishing the database metadata could help in the development of standards. In addition, the lack of metadata and metadata standards, make use of archaeological data more difficult for disciplines outside of archaeology. Greater use of accepted metadata formats would support wider data interoperability.

Second, there is the issue of data provenance. The Comparative Archaeological Study of Comparative Chesapeake Culture website provides metadata for the artifact database, but information on data provenance (how it got from the original data to the aggregate database) is not shared. Data provenance is defined as the “process of tracing and recording origins of data and its movements between databases” (Bruneman et al, 2000, p. 1). All archaeological data goes through processing from the field to the database – such as which data are collected, how they are aggregated, what fields are captured, and the granularity of artifact description. Field methods are sometimes explicit on the websites to help understand how different field methodologies will affect findings (for example <http://www.daacs.org/sites/building-c/#before>). But lab manuals – showing how data are captured, and manipulated, are typically not shown. Making these processes transparent and providing information on data provenance would make the data more useful for comparative studies.

While outside of the scope of this study, a third issue to be addressed is: what data should be made available? With an existing backlog of physical collections, and an understanding by archaeologists that not all archaeological sites are equal (e.g. some archaeological data are from surveys, while others represent full-scale excavations). Should there be an expectation of data access for all type of archaeological sites? While this is a discussion well beyond the scope of this study and includes explorations around prioritizing collections (as discussed by While and Breen (2012), and even the specter of deaccessioning, there are conversations on this topic that are currently on-going (e.g. Federal Register, 2014).

Public Outreach

Public outreach characteristics were generally well supported, with all websites meeting one characteristic and most sites meeting several. There are few, if any, technical barriers to satisfying requirements for public outreach, with social media practices having become well established and standardized (use of Facebook and Twitter for example). Blogs, and other social media were not on all

websites, but they make more sense for ongoing projects, since they provide both an avenue for dissemination and the chance for public engagement in the archaeology. With nearly 80% of the websites with archaeology pages presenting archaeology with contextual information, images of artifacts unsupported by spatial or thematic contextual information seem to be a thing of the past.

The lack of explicit digital educational material (38.5% of sites, $n=15$) may be caused by several factors. The creation of such material may require resources and expertise not always available in-house. In Virginia, providing materials that can be used in schools also requires that the modules tie into the State of Virginia's Standards of Learning (SOLs). Nevertheless, in a broad sense, these educational components can provide disciplinary domain level knowledge – an understanding of the website data requires some understanding of the archaeological process. The Virginia Department of Historic Resources website includes a “What Do Archaeologists Do?” educational component that lays out how archaeologists work. Jamestown Rediscovery provides a virtual lab and field interactive exercise that asks students to answer research questions. Though these educational modules provide little data, they do provide some of the tools needed to understand archaeology and archaeological data; contributing to some of the sense-making referred to by Gurstein (2011, section: A model for effective data use). Public archaeology has a strong history in archaeology, so it is perhaps surprising that not more archaeology educational content was found on archaeology websites in this study.

For public outreach, all but one website provided at least one supporting characteristic. These characteristics could be seen as incremental (the more characteristics present, the better). The survey results support the contention that websites are attempting to reach audiences beyond other archaeologists (see Chapter 4: Table 3). Interestingly, institutional websites that include data are also presenting a greater number of public outreach characteristics, suggesting that this is not an either-or proposition. But there appears to be no consistent template, or approach, for how institutions are presenting archaeological information. The literature on different publics has been discussed earlier, but

what are the data needs of these different groups? The data needs of different publics would be an interesting question for further research. The role and use of digital data for non-archaeologists may be debated, and archaeologists certainly communicate to the public in many other ways – field schools, open houses, exhibits, public talks, books and other media - but perhaps public outreach could make better use of digital archaeological data.

Findings by institutional type

It is likely that different factors are at play regarding how institutions make archaeological data available. At the federal , as seen through an examination of the National Park Service aggregate collections website, individual parks make decisions about what content, if any, to share (National Park Service (n.d. a). Of the 20 National Parks in Virginia with collections, only six have made their archaeological collections public (National Park Service, n.d.).

For other agencies, for example Loudoun County, the focus of the websites is not on presenting research, or collections, and if archaeological content is present, it is buried within the website. The Forest Service website for the George Washington and Jefferson National Forests – an organization that employs an archaeological staff - provides no results, or menu links, when searching for archaeology or archeology. The Virginia state repository, V-CRIS, offers a huge number of archaeological sites, though the information about them is limited and varied, as they include legacy information collected over many years. V-CRIS is also unusual in offering a GIS interface that allows queries by drawing shapes on a map – and the ability to download GIS shape files. The decision to monetize access to archaeological site data has been offset by their sharing some of this information through Open Context, and other parts of the Virginia Department of Historic Resources website provide a broad range of contextual material, with a range of educational content.

Archaeological data were found on only a small number of university and college sites. The total figure of 8 can be compared to the 60 four-year institutions listed on the State College of Higher

Education for Virginia website (<http://www.schev.edu/students/collegelistalpha.asp>) – though not all of these have archaeology/anthropology programs.

Aggregated websites

The sample for this study included three federated websites, providing discovery and access across multiple archaeological institutional sites. While the National Park Service web catalog included only limited data from a few parks, it does serve public outreach in showing that the National Parks have collections – when for most people they are places, not museum collections – and the content can be accessible in an open format. However the scholarly value is limited, particularly for other archaeologists, due to the lack of related contextual information. The other two websites, DAACS and Comparative Archaeological Study of Colonial Chesapeake Culture however, provide strong support for scholarship and open data. These websites provide thematic context for these data, and the ability to query across multiple archaeological sites solving, at least in part, some complicated issues of interoperability. The majority of the interoperable data came from these two sites and their associated institutions.

DAACS contains archaeological sites not just from Virginia but also from the Caribbean, Maryland, South Carolina and Tennessee. The history of DAACS was discussed by Galle in an SHA blog (1012). She notes the website content was created by “physically reanalyzing the assemblages” (para 2), according to protocols and data structures developed by DAACS staff and the steering committee. Galle notes that “Many DAACS users go straight for the archive’s meta-data” (para 4), suggesting the website is being used as a de facto standard for cataloging. Galle also notes “the critical aspect of the DAACS program—providing the standardized data that are essential to any comparative archaeological study” (para 2) DAACS is in many ways an exemplary model for archaeological data publishing. Its strict adherence to cataloging standards overcomes some of the problems of cross-site analysis. Its thematic context (comparative slavery) means the website is inherently contextual, and supported by a range of

papers, manuscripts and an extensive bibliography. The regional and thematic nature of DAACS means that not all archaeological sites are suitable for inclusion and, because of the re-cataloging requirement, it may require more resources than many institutions are able to provide. DAACS also includes information on archaeological sites associated with other institutions in this study. Some of these institutional websites were not included in the data section of this analysis, because the website did not include a link to DAACS (James River Institute and the Fairfield Foundation). Adding links from the host institution websites would make these resources more findable, and it is unknown why these links were not provided.

The Comparative Archaeological Study of Colonial Chesapeake Culture achieves similar goals with a regional and temporal focus. The comparative artifact catalogs are not as extensive as DAACS, and have been created without re-cataloging (Alston, 2005). The project's archaeological sites involve some digitization from paper records and "compiling the context and artifact information from each site into a structured database with authorized fields and terms" (section: Databases and Using digital Technologies). This involved an "extensive modification of some of the original cataloguing fields" (Comparative Archaeological Study of Colonial Chesapeake Culture, 2009, section: Database glossary). Data can be downloaded, and along with V-CRIS, it was one of two websites allowing sharing of spatial data in an open format. The website provides a broad range of re-usable, open data, though it is more limited in public outreach characteristics.

Colonial Williamsburg has an interactive eWilliamsburg component that allows visitors to examine buildings through a map interface, walk through the streets and buildings, and see the documents that support the research for these buildings. While the virtual data are not easily accessible, the site shows the potential for providing data on 3-D models of buildings. Colonial Williamsburg also has another area of the website dedicated to archaeology, with 25 archaeological site reports, available as a mix of PDFs and web pages.

When looking at individual archaeological sites, the Mount Vernon Midden and University of Tennessee website provide strong support for all aspects of this study. The Mount Vernon website includes collections data linked through the DAACS website, and also contains a database of searchable objects. Information is available on archaeological site and excavation units, thematic and temporal information to place objects in a broader context, and references to books, articles, conference papers and reports.

Conclusions

Archaeologists are creating ever increasing amounts of digital content in a variety of formats. While this study has touched lightly on aspects of preservation, the lack of disseminated data suggests that only a portion of these data will survive into the future, and that a digital curation crisis has been added to the physical curation crisis. A principle of preservation technique is LOCKSS – lots of copies keep stuff safe – and increasing dissemination would allow for datasets to be minimally, informally archived and kept in multiple locations. A better solution would be more institutions putting digital archaeological data in trusted archival repositories. It should also be said that archaeology is rare in that there can be a conflict between access and preservation. One of the reasons for the lack of archaeological site information on National Park websites is the very real fear of looting. This is also one of the considerations in restricting access to the state V-CRIS system. Providing information on archaeological sites can endanger them, though there are ways to redact certain pieces of sensitive information.

Why aren't more data made available? The answer perhaps lies in a number of factors worthy of further study. They include the documented lack of standards in data description that makes it hard to combine data from different excavations; archaeologists' attitudes to data release— though it is likely that archaeologists are only part of the decision about how to share digital data on institutional websites—; and funding issues that can focus on excavation rather than curation. New requirements for

data management plans may require greater attention to data curation and access in the future, helping to solve the latter problem. Certainly, if funds are available, the presence of tDAR and Open Context provide new options for data preservation. Finally there is a lack of lightweight models for data publication, particularly for interoperable data. However, while there are technical issues in developing further metadata standards and in implementing trusted repositories, at a basic level putting up PDFs or spreadsheet data are immediately achievable goals. The proviso here is that if only part of the archaeological data is published -- artifact data without accompanying field notes, methods, image and site data -- then important components of the contextual record has been lost.

Which data should be made available, or prioritized for preservation and access, considering the data backlog? This study suggests the current situation falls far short of full access, but deciding on the collection criteria - or priorities -- for future data dissemination should be a conscious decision based on vetted and established selection and appraisal criteria. Clearly there is much more that could be done.

While basic access is relatively simple, access will become more complicated with the increasing introduction of new data formats: GIS and spatial data; 3-D models; and virtual environments. The Virtual Curation Lab at Virginia Commonwealth is working on 3-D scans of artifacts and currently no data are provided, though there is a reference to downloadable scans being made available in the future, hinting at the possibility of new, shared archaeological data types.

There was no support for Linked Open Data in the study, which leaves archaeological data unconnected to other datasets. Beyond immediate data interoperability is the availability of “machine-readable representations of data or [even] structures that facilitate linking and indexing by search engines” (Kansa & Kansa 2011, p. 2). Spectrum, the standard for museum collections in the United Kingdom, has recently announced support for the CIDOC-CRM standard, which is an emerging model of presenting cultural data with “rich and contextual semantics necessary for meaningful data integration,

cross collection exploration, research and collaboration” (CIDOC, 2015). The technologies, software and philosophies of sharing data are evolving, while the backlog of unshared data grows.

Archaeologists are concerned with separating themselves from the practice of pot-hunting. One of the more easily understood differences between archaeologists, and “pot-hunters” who dig archaeological sites to collect artifacts, is data. Archaeologists undertake scientific excavations, and making data transparent could be considered part of archaeological advocacy. Making explicit the data upon which reports and summations are written supports the notion of archaeology as a scientific process, as well as serving the needs of other professionals who can verify and build upon those data. There are issues of data literacy, standards, and supporting metadata in using and making sense of data for multiple publics, but just making them accessible shows the basis of archaeological conclusions, allowing for a more reflexive process. The data should be present on the website even if most visitors don’t look at the data tables.

The extent to which various publics can benefit from data is yet to be seen, but much of these data are derived from public funds in a political climate calling for public access. Further studies would be needed to understand the effectiveness of this dissemination to scholarship, education, and in building support for archaeological projects and cultural resource preservation. Are archaeologists changing views about culture, history and preservation through public engagement? What is the role of data in these conversations? Archaeologists’ control of information raises ethical questions and considerations of authority, trust, provenance, and the role of gatekeepers. While this study is discipline specific, the questions raised concern the digital literacy needed to take advantage of data, regardless of discipline. What information needs to be provided with datasets to make them useful? If, as Nellie Kroes states, “Data is the new gold” (European Commission, 2011) then we might more broadly ask how well data are able to provide value? Who is in a position to mine this gold? Making sense of data, creating value from data – by government, businesses, other publics and archaeologists – will require a degree of

digital literacy; meaningful access requires a number of factors discussed earlier. Is open data democratic, in sharing the results with all, or ironically, is it elitist, providing more information to those with the tools to take advantage of it? It will require those serving up the data to think about the audiences they are trying to reach, and how best to present these data. This analysis of a particular type of data –archaeological – is a small step in trying to explore some of those considerations.

There are good examples of data sharing identified in this study. The Digital Archaeological Archive of Comparative Slavery and the Comparative Archaeological Study of Colonial Chesapeake Culture met most of the characteristics for scholarly access, open data and public outreach. These are exemplary examples of federated collections that allow for regional and thematic analysis. These websites are providing resources on archaeology sites that were excavated forty years ago, showing the potential long-term value of these datasets. Alexandria Archaeology Museum and Colonial Williamsburg provide information on multiple archaeological sites, much of it accessible for multiple audiences, and Colonial Williamsburg's work on virtual environments suggests how archaeological reports and data can be woven into the interpretive presentation fabric. On a single archaeological site level, the Mount Vernon Midden website is an explicit attempt to serve the non-archaeology publics at a sophisticated level, showing the archaeological data in a rich, thematic contextual setting.

This quantitative content analysis of websites provides, at a moment in time, a benchmark to available data from archaeological projects in Virginia and an insight into what data are unavailable. The instrument used could be re-used in analyzing websites from another state to get a broader picture of archaeology in North America. Hopefully the study may help in defining some of the website characteristics demanded of openness, scholarship and public outreach that will help archaeologists decide which types of data dissemination may be the most suitable for their needs.

References

- ACCORD. (2014). About accord. Retrieved from <http://accordproject.wordpress.com/about/>
- Advisory Council on Historic Preservation. (2009). Section 106 archaeology guidance. Retrieved from: <http://www.achp.gov/docs/ACHP%20ARCHAEOLOGY%20GUIDANCE.pdf>
- Agbe-Davies, A. S., Galle, J. E., Hauser, M. W., & Neiman, F. D. (2014). Teaching with digital archaeological data: A research archive in the university classroom. *Journal of Archaeological Method and Theory*, 21(4), 837-861. doi:10.1007/s10816-013-9178-3
- Alexandria (2013, January 8). Alexandria Archaeology Museum. Retrieved from: <http://alexandriava.gov/historic/archaeology/?id=33502>
- Alston, C. L. (2005). Digital technology in comparative studies. Retrieved from <http://www.chesapeakearchaeology.org/Interpretations/SHAAIstonDigitalTechnology.cfm>
- Anderson, D., Yerka, S., & Kansa, E. (2013). Developing the cyberinfrastructure for a national archaeological site database project. Retrieved from: <http://ux.opencontext.org/blog/wp-content/uploads/2012/09/DINAA-NASD-Technical-Proposal-2011.pdf>
- Andrew W. Mellon Foundation (2008). President's report 2008. *The Andrew W. Mellon Foundation*. Retrieved from: http://www.mellon.org/news_publications/annual-reports-essays/presidents-reports/2008
- Archaeology Data Service & Center for Digital Antiquity (2013). *Caring for digital data in archaeology: A guide to good practice*. Oxford, UK: Oxbow Books.
- Association for Library Collections and Technical Services. (2009). Definitions of digital preservation. Retrieved from: <http://www.ala.org/alcts/resources/preserv/2009def>
- Beagrie, N., & Houghton, J. (2013). The value and impact of the Archaeology Data Service, a study and methods for enhancing sustainability. *Jisc*. Retrieved from: http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf
- Beale, N. (2012). How community archaeology can make use of open data to achieve further its objectives. *World Archaeology*, 44(4), 612–633. doi:10.1080/00438243.2012.743252
- Beck, A., Bevan, A., & Costa, S. (2012). What does the Holy Grail look like? Defining open data in archaeology and the related issues. [PowerPoint slides]. Retrieved from: <http://www.slideshare.net/stefanocosta/what-does-the-holy-grail-look-like-defining-open-data-in-archaeology-and-the-related-issues>
- Beck, A., & Stott, D. (2013). Unleashing the potential of collaboration, archaeological detecting in the 21st century. [PowerPoint slides]. Retrieved from: <http://www.slideshare.net/DARTProject/arb-presentation-dublin>
- Berners-Lee, T. (2010, February). The year open data went worldwide. TED talks. Retrieved from: http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide

- Bland, J. (2011, October 1). Costs and benefits of open data – Professor Trefethen. The Royal Society In Verba. Retrieved from: <http://blogs.royalsociety.org/in-verba/2011/10/04/costs-and-benefits-of-open-data-professor-anne-trefethen/>
- Borgman, C. L. (2010). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Borgman, C. L. (2015). If data sharing is the answer, what is the question? ERCIM News. Retrieved from: <http://ercim-news.ercim.eu/en100/special/if-data-sharing-is-the-answer-what-is-the-question>
- Breen, E. (2011, May 16). Popular Archaeology. Mount Vernon’s archaeological collections online project. Retrieved from: <http://popular-archaeology.com/issue/june-2011/article/mount-vernons-archaeological-collections-online-project>
- Canadian Archaeological Association (2014). *Position Paper: excavation of archaeological material in the popular media*. Retrieved from: <http://canadianarchaeology.com/caa/news-announcements/position-paper-excavation-archaeological-material-popular-media>
- Car, T., Raspor, S., Komsic, J., & Markovic, S. (2012). Content analysis of hotel websites: case study of Croatia. *Studia UBB Negootia*, LVII(4), 27–37.
- Chenhall, R. G. (1971). The archaeological data bank: a progress report. *Computers and the Humanities*, 5(3), 159-169.
- Comparative Archaeological Study of Colonial Chesapeake Culture (2009). Database glossary. Retrieved from: <http://www.chesapeakearchaeology.org/AboutTechnical/DatabaseStructure.cfm>
- Childs, S. T., & Kagan, S. (2008). *A decade of study into repository fees for archeological curation*. Washington, DC. doi:10.6067/XCV8NS0T7S
- CIDOC (2015). CIDOC Conceptual reference model (CRM) extension for SPECTRUM. Retrieved from: <http://www.cidoc-crm.org/docs/CRM-SPECTRUM-PressRelease-2015Feb.pdf>
- City of Manassas (n.d.) Mayfield Fort. Retrieved from: <http://www.manassascity.org/index.aspx?NID=220>
- Cole, R. (2012). Some observations on the practice of “open data” as opposed to its promise. *The Journal Of Community Informatics*, 8(2). Retrieved from: <http://ci-journal.net/index.php/ciej/article/view/920/917>
- Costa, S. (2014). Open licenses for archaeological data matter: the case of AustArch. [Blog post]. Retrieved from: <http://archaeology.okfn.org/>
- Culture Embossed (2012). About Culture Embossed. Retrieved from: <http://cova-inc.org/wineseals/project-description.html>

- DAACS (2014). DAACS receives grant from Mellon Foundation for innovative collaborative project known as The DAACS Research Consortium. Retrieved from: <http://www.daacs.org/2013/10/04/daacs-receives-grant-from-mellon-foundation-for-innovative-collaborative-project-known-as-the-daacs-research-consortium/>
- Denning, K. (2004). The storm of progress' and archaeology for an online public. *Internet Archaeology*, 15, 322-349. Retrieved from: http://intarch.ac.uk/journal/issue15/denning_toc.html
- Doerr, M., & Oldman, D. (2013). The costs of cultural heritage data services: the CIDOC CRM or aggregator formats? [Blog post] Retrieved from: <http://www.oldman.me.uk/blog/costsofculturalheritage/>
- Dunn, S., & Hedges, M. (2012). *Crowd-sourcing scoping study engaging the crowd with humanities research centre for e-research*, Department of digital humanities King 's College London. Retrieved from: <http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/12/Crowdsourcing-connected-communities.pdf>
- European Commission (2011). European press release database. Retrieved from: http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm?locale=en
- Fairfax County (n.d.). Cultural resource management and protection. Retrieved from: <http://www.fairfaxcounty.gov/parks/resource-management/resources-crp.htm>
- Fort Lee (2014). Cultural resource program area. Retrieved from: <http://www.lee.army.mil/dpw/emd/cultural.resources.aspx>
- Falkingham, G. (2005). A whiter shade of grey: a new approach to archaeological grey literature using the XML version of the TEI guidelines. *Internet Archaeology*, 17. Retrieved from: <http://intarch.ac.uk/journal/issue17/5/toc.html>
- Faniel, I., Kansa, E., Kansa, S., Barrera-Gomez, J., & Yakel, E. (2013). The challenges of digging data: A study of context in archaeological data reuse. *JCDL Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 295-304. New York, NY: ACM. doi:10.1145/2467696.2467712.
- Fantoni, S. F., Stein, R., & Bowman, G. (2012). Exploring the relationship between visitor motivation and engagement in online museum audiences. *Museums and the Web 2012: Proceedings*. Toronto: Archives & Museum Informatics. Consulted October 15, 2013. Retrieved from: http://www.museumsandtheweb.com/mw2012/papers/exploring_the_relationship_between_visitor_mot.html
- Freeman, M. (2011, April 12). Should you object? Mount Vernon midden archaeology. [Blog post]. Retrieved from: <http://storiespast.blogspot.com/2011/04/should-you-object-mount-vernon-midden.html>
- Freeman, M. (2013). NPS web catalog study. Unpublished. Retrieved from: https://www.academia.edu/5790012/NPS_Web_Catalog_Study

- Freeman, M. (2015). Digital Archaeology; A new specialization. Retrieved from: <https://www.academia.edu/>
- Galle, G. (2012, May 29). Will today's graduate training in Historical Archaeology predict the future of digital research archives? [Blog post]. Retrieved from: <http://www.sha.org/blog/index.php/2012/05/will-todays-graduate-training-in-historical-archaeology-predict-the-future-of-digital-research-archives/>
- Gurstein, M. (2011). Open data: empowering the empowered or effective data use for everyone?. *First Monday*, 16(2). doi:10.5210/fm.v16i2.3316
- Hamma, K. (2004). The role of museums in online teaching, learning, and research. *First Monday*. 9(5). Retrieved from: <http://firstmonday.org/ojs/index.php/fm/rt/prINTERfriendly/1146/1066>
- Hank, C. (2013). Communications in blogademia: An Assessment of scholar blogs' attributes and functions. *New Review of Information Networking*, 18(2), 51–69. doi:10.1080/13614576.2013.802179
- Harley, D., Accord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. UC Berkeley. Retrieved from: <http://escholarship.org/uc/item/15x7385g>
- Heritage Bytes (2012). Developing the cyberinfrastructure for a national archaeological site database. [Blog post] Retrieved from: <http://ux.opencontext.org/blog/wp-content/uploads/2012/09/DINAA-NASD-Technical-Proposal-2011.pdf>
- Jamestown (2014). Public archaeology enhanced by today's technology. [Video] Retrieved from: <http://www.youtube.com/watch?v=0sPa2nq09SE&list=UUVtdK0zt6-tj4FIWOEHfJiw>
- Jeffrey, S. (2012). A new digital dark age? Collaborative web tools, social media and long-term preservation. *World Archaeology*, 44(4), 553–570. doi:10.1080/00438243.2012.737579
- JOAD. (n.d.). Why share archaeology data (image). Retrieved from: <http://openarchaeologydata.metajnl.com/about.html>
- Jones, S., MacSween, A., Jeffrey, S., Morris, R. , & Heyworth, M. (2003). 'From the ground up the publication of archaeological projects: a user needs survey', *Internet Archaeology* 14. Retrieved from: http://intarch.ac.uk/journal/issue14/puns_index.html.
- Kansa, E. (2005). A community approach to data integration: authorship and building meaningful links across diverse archaeological data sets. *Geosphere* 1(2), 97-109. Retrieved from: <http://geosphere.gsapubs.org/content/1/2/97.full.pdf+html>
- Kansa, E. (2012). Openness and archaeology's information ecosystem. *World Archaeology*, 44(4), 498–520. doi:10.1080/00438243.2012.737575

- Kansa, E., & Kansa, S. (2010). Publishing data in Open Context: methods and perspectives. *CSA Newsletter, XXIII*(2). Retrieved from: <http://csanet.org/newsletter/fall10/nlf1001.html>
- Kansa, S (2012). Digital Index of North American Archaeology (DINAA). *Heritage Bytes: updates and news about Open Context*. [Blog post] Retrieved from: <http://ux.opencontext.org/blog/archaeology-site-data> (accessed 1 November 2013).
- Kansa, S., & Kansa, E. (2011). Enhancing humanities research productivity in a collaborative data sharing environment. *White Paper NEH Division of Preservation and Access*. Retrieved from: http://ux.opencontext.org/blog/wp-content/uploads/2011/06/white_paper_PK_50072.pdf
- Kansa, S. W., Kansa, E. C., & Arbuckle, B. (2013). Publishing and pushing : mixing models for communicating research data in archaeology. *International Journal for Digital Curation. 9*, 1–15.
- Kintigh, K. (2006). The promise and challenge of archaeological data integration. *American Antiquity, 71*(3), 567–578. doi:10.2307/40035365
- Kintigh, K. W., & Altschul, J. H. (2010). Sustaining the digital archaeological record. *Heritage Management. 3*(2):264- 274.
- Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. (2014). Grand challenges for archaeology. *Proceedings of the National Academy of Sciences of the United States of America, 111*(3), 879–80. doi:10.1073/pnas.1324000111
- LATC (2012). Five star open data. Retrieved from: <http://5stardata.info/>
- Limp, F., Payne, A., Simons, K., Winters, S., & Cothren, J. (2011). Developing a 3-D digital heritage ecosystem: from object to representation and the role of a virtual museum in the 21st century. *Internet Archaeology, (30)*. <http://dx.doi.org/10.11141/ia.30.1>
- Marquardt, W., Montet-White, A., & Scholtz, S. (1982) Resolving the crisis in archaeological collections Curation. *American Antiquity 47*(2), 409-418
- McDavid, C. (1999). From real space to cyberspace: contemporary conversations about the archaeology of slavery and tenancy. *Internet Archaeology, (35)*. <http://dx.doi.org/10.11141/ia.6.4>
- McManamon, F., Kintigh, K., & Brin, A. (2010). Digital Antiquity and the digital archaeological record (tDAR): Broadening access and ensuring long-term preservation for digital archaeological data. *CSA Newsletter, XXIII*(2). Retrieved from: <http://csanet.org/newsletter/fall10/nlf1002.html>
- McMillan, S. J. (2000). The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. *Journalism & Mass Communication Quarterly, 77*(1), 80-98.
- McMillan, L., Hatch, D., & Heath, B. (in press). Dating methods and techniques at the John Hallowes site (44WM6): A 17th-Century Example. *Northeast Historical Archaeology*.

- Meho, L. I., & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, 54(6), 570–587. doi:10.1002/asi.10244
- Montpelier (2015). Topics. [Blog post] Retrieved from: <http://www.montpelier.org/blogs/tags/Archaeology>
- Morrison, W., Thomas, R. & Gosden, C. (2014). Laying bare the landscape: commercial archaeology and the potential of digital spatial data. *Internet Archaeology*, (36). doi:10.11141/ia.36.9
- Moyer, Teresa S. (2006). Technical Brief 19: archeological collections and the public: using resources for the public benefit. DOI Departmental Consulting Archeologist/NPS Archeology Program, National Park Service, Washington, DC
- Museum of the Middle Appalachians (n.d.). Woodland Indians museum of the middle Appalachians. Retrieved from: <http://www.museum-mid-app.org/native-american-villages.htm>
- Mullins, Paul (2013, October 4). Archaeological personalities and the profession's future. [Blog post]. Retrieved from: <http://www.sha.org/blog/index.php/2013/10/archaeological-personalities-and-the-professions-future/#more-3264>
- National Endowment for the Humanities (2011). Data management plans for NEH Office of digital humanities proposals and awards. Retrieved from: http://www.neh.gov/files/grants/data_management_plans_2014.pdf
- National Park Service (n.d.). Part 79--Curation of federally-owned and administered archeological collections. Retrieved from: <http://www.nps.gov/archeology/tools/36cfr79.htm>
- National Park Service (n.d. a). Parks. Retrieved from: <http://museum.nps.gov/ParkPList.aspx>
- National Park Service (n.d. b). FAQ. Retrieved from: <http://museum.nps.gov/ParkFaq.aspx>
- National Science Foundation (n.d.). Archaeology - senior award Info. Retrieved from: <http://www.nsf.gov/sbe/bcs/arch/senior.jsp>
- National Science Foundation. (2013). GPG Chapter II. Retrieved from; http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#IIC2j
- Nebbia, M. (2014). Review of Gabii goes digital. *Internet Archaeology* (36). <http://dx.doi.org/10.11141/ia.36.3>
- Neuendorf, K. A. (2002). The content analysis guide. Sage Press.
- Neuman, W. L. (2013). *Social Research Methods: Pearson New International Edition*. Pearson Education Limited.

- Northumberland Historical Society (n.d.) Explore Our Collections. Retrieved from:
www.northumberlandcountyhistoricalsociety.org/page.asp?tid=152
- Open Knowledge. (n.d.). Open definition. Retrieved from: <http://opendefinition.org/od/>
- Open Government Data. (n.d.). The annotated principles of open government data. Retrieved from;
<http://opengovdata.org/>
- Peacock, D., & Brownbill, J. (2007). "Audiences, visitors, users: reconceptualising users of museum online content and services." In J. Trant and D. Bearman (eds.). *Museums and the Web 2007: Proceedings*. Toronto: Archives & Museum Informatics. Retrieved from:
<http://www.archimuse.com/mw2007/papers/peacock/peacock.html>
- Pilaar Birch, S. (2013). Using social media for research dissemination: The digital research video project. *Internet Archaeology*, (35). Council for British Archaeology. doi:10.11141/ia.35.4
- Prithviraj, K. R., & Sampath Kumar, B. T. (2014). Corrosion of URLs: Implications for electronic publishing. *IFLA Journal*, 40(1), 35–47. doi:10.1177/0340035214526529
- Punch, K (2005). *Introduction to Social Research Quantitative and Qualitative Approaches*. Sage.
- Richards, J. D. (2004). Online archives. *Internet Archaeology*. Retrieved from:
<http://intarch.ac.uk/journal/issue15/7/jr1.html>
- Richards, J. (2014). Towards a history of archaeological computer. Proceedings from CAA 2014 Concepts Methods Tools. Paris, France.
- Richards, J.D. & Robinson, D. (eds) 2000 *Digital Archives from Excavation and Fieldwork: Guide to Good Practice*, Second Edition. AHDS Guides to Good Practice. Retrieved from:
<http://ads.ahds.ac.uk/project/goodguides/excavation>.
- Richardson, Lorna. (2014). Understanding archaeological authority in a digital context. *Internet Archaeology* 38. <http://dx.doi.org/10.11141/ia.38.1>
- Rosewell Foundation (n.d.) History of Rosewell. Retrieved from: <http://www.rosewell.co/history-of-rosewell.html>
- Sarraf, S. (1999). A survey of museums on the web: who uses museum websites? *Curator: The Museum Journal* 42.3. Blackwell Publishing Ltd
- Scott, S. (2013). Pioneers, publishers and the dissemination of archaeological knowledge: A study of publishing in British archaeology 1816-1851. *Internet Archaeology*, (35). Council for British Archaeology. doi:10.11141/ia.35.1 Retrieved from:
http://intarch.ac.uk/journal/issue35/scott_index.html
- Shanks M., & Witmore C. (2012). 'Archaeology 2.0? Review of *Archaeology 2.0: New Approaches to Communication and Collaboration* [Web Book]', Internet

- Archaeology **32**. <http://dx.doi.org/10.11141/ia.32.7> Retrieved from: <http://intarch.ac.uk/journal/issue17/5/toc.html>
- Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A Survey of Data Provenance in e-Science. *SIGMOD Record*, 34(3), 31–36. doi:10.1145/1084805.1084812
- Society for American Archaeology. (n.d.). Principles of archaeological ethics. Retrieved March 14, 2014, from <http://www.saa.org/AbouttheSociety/PrinciplesofArchaeologicalEthics/tabid/203/Default.aspx>
- Society for Historical Archaeology. (2003). SHA - ethics statement. Retrieved from: <http://www.sha.org/index.php/view/page/ethics>
- Society for Historical Archaeology. (2013). Awards and Awardees. Retrieved from: http://www.sha.org/index.php/view/page/awards_awardees
- Society for Historical Archaeology. (in press). SHA 2014 survey. SHA Newsletter
- Sullivan, L., & Childs, S. (2003). *Curating archaeological collections: from the field to the repository* (Vol. 6). Rowman, Altamira.
- tDAR (2014). Teaching. Retrieved from: <http://www.tdar.org/why-tdar/access/teaching/>
- Tschan, R. (2002). A comparison of Jenkinson and Schellenberg on appraisal. *The American Archivist*, 65(2), 176–195.
- Warner, M. (2014). Turning Inwards: Collections-driven research and the vitality of the discipline. Paper presented at the annual meeting of the Society for Historical Archaeologists, Quebec, 2014.
- Watts, J. (2011). Policies, preservation, and access to digital resources : The Digital Antiquity 2010 National Repositories Survey. doi:10.6067/XCV8CC1022
- Weare, C., & Lin, W. Y. (2000). Content analysis of the world wide web opportunities and challenges. *Social Science Computer Review*, 18(3), 272-292.
- White, E. C., & Breen, E. (2012). *A survey of archaeological repositories in Virginia*. Retrieved from: <http://www.cova-inc.org/resources/COVAcollectionsSurvey.pdf>
- White House (2013). Open Government Initiative. Retrieved from: <http://www.whitehouse.gov/open>
- White House (2013a). Increasing access to the results of federally funded scientific research. Retrieved from: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- William and Mary (2015). Archaeology at the Brafferton. [Blog post] Retrieved from: <http://www.wm.edu/sites/braffertonproject/archaeology/index.php>

Winters, J. (2006). Editorial. *Internet Archaeology* (19). doi:10.11141/ia.19.8

Winters, J. (2013). Editorial. *Internet Archaeology*, (35). doi:10.11141/ia.35.12

Appendices

Appendix A – institutional websites holding archaeology collections

Repository Name	URL
African American Heritage Sites	www.visithampton.com
Alexandria Archaeology Museum	www.alexandriaarchaeology.org
Alexandria Black History Museum and African-American Heritage Park	oha.alexandriava.gov/bhrc/
Alexandria Canal Lift Lock and Pool No. 1	oha.alexandriava.gov/oha-main/oha-alexandria-canal.html
Amazement Square	www.amazementsquare.org
Anthony A. Burke	
Appomattox Court House National Historical Park	www.nps.gov/apco/
Arlington House, The Robert E. Lee Memorial	www.nps.gov/arho
Assateague Island National Seashore	www.nps.gov/asis/
Avoca Museum	www.avocamuseum.org
Bassett History Center	www.bassetthistoricalcenter.com
Bedford City/County Museum	www.bedfordvamuseum.org
Belvoir Plantation Historic Site	www.belvoir.army.mil/history.asp?id=Antebellum
Black History Museum and Cultural Center of Virginia	www.blackhistorymuseum.org
Blue Ridge NPS	www.nps.gov/blri
Booker T. Washington National Monument	www.nps.gov/bowa/index.htm
Browning & Associates, Ltd.	
Brunswick County Museum	www.tourbrunswick.org/brunswick_county_museum.htm
Camp Roosevelt (Forest Service)	www.fs.fed.us/r8/gwj/lee/
Cape Charles Museum and Welcome Center	www.smallmuseum.org/capechas.html
Carlyle House Historic Park	www.nvrpa.org/park/carlyle_house_historic_park
Carroll County Historical Society Museum	www.historicalsociety.chillsnet.org
Cedar Creek Battlefield	www.cedarcreekbattlefield.org
Center for Historic Preservation	www.umw.edu/chp/
Chesapeake and Ohio Canal National Historical Park	www.nps.gov/choh
Chesterfield Museum Complex	www.chesterfieldhistory.com/
City Point Early History Museum (Weston Manor)	www.historichopewell.org
College of William and Mary	powhatan.wm.edu/
College of William and Mary wmcar	www.wm.edu/sites/wmcar/
Colonial National Historical Park (Jamestown)	www.nps.gov/colo/
Colonial Williamsburg Foundation	research.history.org/Archaeological_Research.cfm
Comparative Archaeological Study of Colonial Chesapeake Culture	www.chesapeakearchaeology.org
Culture Embossed	cova-inc.org/wineseals/index.html

Dahlgren Naval Surface Warfare Center	www.navsea.navy.mil/nswc/dahlgren
Daniel Harrison House	www.heritagecenter.com/Fort%20Harrison/fortharrison.html
Digital Archaeological Archive of Comparative Slavery	www.daacs.org/
Eastern Shore of Virginia Historical Society (ESVHS)	www.shorehistory.org/
Elizabeth Furnace	www.fs.fed.us/r8/gwj/lee/
Endview Plantation	www.endview.org
Fairfax County Park Authority	www.fairfaxcounty.gov/parks/resource-management/resources-crp.htm
Fairfield Foundation	www.fairfieldfoundation.org/
Fayette Area Historical Initiative African American Museum	fahimuseum.org
Flowerdew Hundred	www.flowerdewunderground.com/44pg65.html
Fort Lee	www.lee.army.mil/dpw/emd/cultural.resources.aspx
Fort Monroe	www.nps.gov/fomr/index.htm
Fredricksburg and Spotsylvania National Military Park	www.nps.gov/frsp/index.htm
Gadsby's Tavern Museum - Alexandria	oha.alexandriava.gov/gadsby/
George Mason	soan.gmu.edu/
George Washington's Birthplace National Monument	www.nps.gov/gewa/
George Washington's Ferry Farm	www.kenmore.org/archaeology.html
George Washington's Office Museum	www.winchesterhistory.org/george_washington.htm
Glencoe Museum	glencoemuseum.org/
Gloucester Museum of History	www.gloucesterva.info/MuseumofHistory/tabid/1033/Default.aspx
Graves Mountain Lodge	www.gravesmountain.com
Great Falls Park - NPS	www.nps.gov/grfa
Gunston Hall Plantation	www.gunstonhall.org/index.php/archaeology-grounds/archaeology
Halifax County Historical Society	www.halifaxcountyhistoricalsociety.org/
Haller-Gibboney Rock House Museum	museums.wytheville.org/museums.php
Harpers Ferry	museum.nps.gov
Henricus Historical Park	www.henricus.org
Historic Christ Church Foundation	www.christchurch1735.org/archaeology.html
Historic Crab Orchard Museum and Pioneer Park	www.craborchardmuseum.com
Historic Hopewell Foundation Inc (HHFI)	www.historichopewell.org/
Historic Pole Green Church Foundation	www.historicpolegreen.org
Historic Services Department, City of Newport News	www.nngov.com/880/Museums-Historic-Sites
Historic Smithfield Plantation	www.smithfieldplantation.org
History Museum of Western Virginia	www.vahistorymuseum.org/

Ingleside Plantation Vineyards	www.inglesidevineyards.com/
Isle of Wight County Museum	www.historicisleofwight.com/museum.html
James Madison University	www.jmu.edu/archaeology/
James River Institute, JRIA	www.jriarchaeology.com/
Jamestown Rediscovery Preservation Virginia	apva.org/rediscovery/page.php?page_id=1
Jamestown Settlement	www.historyisfun.org
Ker Place	www.shorehistory.org/
Longwood College (library)	library.longwood.edu/research/
Loudon Archaeological Foundation	loudounarchaeology.org/
Loudoun County	www.loudoun.gov/
Lynchburg Museum System	www.lyncburgmuseum.org
MacCallum More Museum & Gardens	www.mmmg.org/#!arthur-robertson/c1n5d
Madison County Historical Society/Madison Arcade	madisonvahistoricalsociety.org/
Madison County Library	www.madisoncountyvalibrary.org/
Maggie L. Walker National Historic Site	museum.nps.gov
Manassas	museum.nps.gov
Manassas Museum System	www.manassascity.org/index.aspx?NID=211
Mariners' Museum - Monitor collection	www.marinersmuseum.org/catalogs/
Mary Ball Washington Museum & Library	mbwm.org
Mattaponi Indian Museum	www.uppermattaponi.org/
Menokin Foundation	www.menokin.org/
Monacan Ancestral Museum	www.monacannation.com/museum.shtml
Monacan Village at Natural Bridge	www.naturalbridgeva.com
Monticello	www.monticello.org/archaeology/index.html
Montpelier Foundation	www.montpelier.org/research-and-collections/archaeology
Mount Vernon Archaeology	www.mountvernon.org/research-collections/archaeology/
Mount Vernon Midden	mountvernonmidden.org/
Museum of Culpeper History	www.culpepermuseum.com
Museum of the Middle Appalachians	www.museum-mid-app.org
Naval Shipyard Museum, Portsmouth	www.portsmouthnavalshipyardmuseum.com/ naval_shipyard/index.html
NCR Regional Archaeology Program	www.nps.gov/rap/
New London	www.newlondonmuseum.org/
New River Gorge National River	www.nps.gov/neri
Newtown History Center	newtownhistorycenter.org/stone-house-restoration-project/ archeology/
Northumberland County Historical Society	www.northumberlandcountyhistoricalsociety.org/ page.asp?tid=152
Occoneechee State Park Visitors Center	www.dcr.virginia.gov/state_parks/occ.shtml
Old Jail Museum	www.chesterfieldtourism.com/museum.shtml
Pamplin Historical Park and National Museum	pamplinpark.pastperfect-online.com/32996cgi/ mweb.exe?request=ks

Pamunkey Indian Museum	www.pamunkey.net/museum.html
Petersburg National Battlefield	www.nps.gov/pete
Preservation Virginia	preservationvirginia.org/
Prince William County Archaeology	www.pwcgov.org/government/dept/planning/pages/archaeology.aspx
Prince William Forest Park - but info from NPS regional	www.nps.gov/rap/
Pulaski County Courthouse	www.courts.state.va.us/courts/circuit/Pulaski/home.html
Quantico	www.usmcmuseum.org/
Radford (special collections and archives)	www.radford.edu
Radford University Forensic Science Institute	www.radford.edu/content/csat/home/forensic-science.html
Red Hill	www.redhill.org
Reuel B. Pritchett Museum	www.bridgewater.edu/StudentServices/AlexanderMackMemorialLibrary/ReuelBPritchettMuseum
Richmond County Museum	www.co.richmond.va.us/visitors/richmond-county-museum
Richmond National Battlefield Park	www.nps.gov/rich
Rockbridge Historical Society (Campbell House)	www.rockhist.org
Rosewell Foundation	www.rosewell.org
Salem Museum & Historical Society	www.salemmuseum.org
Science Museum of Western Virginia	www.smwv.org
Shenandoah National Park	www.nps.gov/shen
Shenandoah Valley Discovery Museum	www.discoverymuseum.net
South Boston-Halifax County Museum of Fine Arts and History	www.sbhcmuseum.org
Southwest Virginia Museum Historical State Park	www.dcr.virginia.gov/state-parks/southwest-virginia-museum.shtml#general_information
Staunton River Battlefield State Park	www.stauntonriverbattlefield.org/arch.html
Stratford Hall Plantation - DAACS	www.stratfordhall.org/collections-research/archaeological/sbc.edu/course-overviews/archaeology
Sweet Briar College	www.tredegar.org/
The American Civil War Center of Historic Tredegar	
The Museums of Wytheville	museums.wytheville.org/
Thomas Jefferson's Poplar Forest	www.poplarforest.org/archaeology-at-poplar-forest/
University of Tennessee	web.utk.edu/~bheath2/
USDA Forest Service-George Washington and Jefferson National Forests	www.fs.usda.gov/main/gwj/home
Valentine Richmond History Center	richmondhistorycenter.com/index.asp
Virginia Aquarium & Marine Science Center	www.virginiaaquarium.com/
Virginia Archaeological Resource Center of the ASV (also Kittiewan Chapter)	www.kittiewanplantation.org/
Virginia Commonwealth University	vcuarchaeology3d.wordpress.com/
Virginia Department of Historic Resources inc. VCRIST	www.dhr.virginia.gov/arch_DHR/archaeo_index.htm

Virginia Institute for Marine Science (VIMS)	www.vims.edu/
Virginia Military institute	www.vmi.edu/
Virginia Museum of Natural History	www.vmnh.net/collections
Virginia National Guard Artifact Curation Facility	vko.va.ngb.army.mil/VirginiaGuard/environmental/cultural%20resources.htm
Virginia State Parks	www.virginia.gov/visit/state-parks
Washington and Lee University	www.wlu.edu/sociology-and-anthropology/archaeology-at-wandl
Westmoreland County Museum & Library, Inc.	www.westmoreland-county.org/index.php?p=visiting&c=countyMuseum
Wilderness Road Regional Museum	www.newriverheritage.org/members-wildernessroad.htm
Wilton House Museum	www.wiltonhousemuseum.org
Wolf Creek Indian Village and Museum	www.indianvillage.org
Yorktown Victory Center	www.historyisfun.org

Appendix B – Website Coding System

Data Management

- A. Website ID #: Unique 3-digit number assigned at sampling. Enter in digits.
- B. Sampling Date: Date website reviewed. Enter as mm-dd-yy

Institutional Information

- C. Organization Name: from Home page. Enter as text
- D. Website URL: from Home page. Enter as text
- E. Organization Type: from home page name, or from domain (.gov .va .org .com). 1=Federal
2=City/State 3=University 4=Non-profit 5=Private. Enter as digits
- F. Is there any archaeological section? **If 0 or 1 no further coding will be done on this record.**
0=No 1=Mention of Archaeology Only 2=Yes Enter as digits
- G. Is there any archaeological data? **If “0” go to skip open data and scholarly data sections.**
0=No 1=Yes Enter as digits
- H. Number of archaeology sites represented: Enter as digits; 0 means unknown
- I. Basis of dissemination: Are the sites presented on national, regional or thematic basis
0=Unknown 1=National 2=Regional 3=Thematic 4=Site. Enter as digits
- J. Primary Interface: from Home page or from query/search results page: 1=Text 2=Spatial 3=3-
D. Enter as digits
- K. Primary Data Level: from the Home page or from query/search results page 1=Summary
2=Site report 3=Field records 4=Spatial data 5=Artifact collections. Enter as digits

Data Content

- L. Site reports: Are site reports available? 0=No 1=Yes. Enter as digits
- M. Field Records: Are field records available? 0=No 1=Yes. Enter as digits

N. Artifact records: Are artifact records available? 0=No 1=Yes. Enter as digits

O. Spatial data: Are spatial data available? 0=No 1=Yes. Enter as digits

P. Virtual data: Are virtual data available? 0=No 1=Yes. Enter as digits

Scholarly Publishing

Q. Citation: does the site provide information on how to cite content? 0=No 1=Yes. Enter as digits

R. CC license: What is the CC license? Where not expressed as a CC license what is the CC equivalent? 0=None 1=CC By 2=BY-SA 3=BY-ND 4=BY-NC 5=BY-NC-SA 6=BY-NC-ND: Enter as digits

S. Permanent URIs: Is data content supported by permanent URIs? 0=No 1=Yes. Enter as digits

T. Permanence/Archiving: Does the website provide information on how content is archived? 0=No 1=Yes. Enter as digits

U. Contact Information: is a phone number/email provided for more information about the collections? 0=No 1=Yes. Enter as digits

Open Data

V. Type of open data: based on 5 star open data rating 1=Text 2=Structured 3=Non-proprietary structured. Enter as digits

W. Access: is access to the site open/requires a login/requires a subscription. 0=Free 1=Login Required 2=Subscription. Enter as digits

X. Accessibility: are any website plugins required to access the data 0=No 1=Yes. Enter as digits

Y. Data dictionary/Thesaurus: is a data dictionary provided? 0=No 1=Yes. Enter as digits

Public outreach

Z. Summary reports: Are site summaries available? 0=No 1=Yes. Enter as digits

- AA. Blog: does the website provide a link to a blog? 0=No 1=Yes Enter as digits
- BB. Social Media: does the website provide a link to a related social media account (Facebook, Twitter etc.)? 0=No 1=Yes Enter as digits
- CC. Contextual Information: does the website include broader contextual information (historical sources, material culture studies, geographic land studies etc. 0=No 1=Text 2=Images 3=Video. Enter as digits
- DD. Supporting Media: does the website include supporting media? 0=No 1=Yes. Enter as digits
- EE. Outreach: does the site link to information about field schools/open houses/volunteer opportunities? 0=No 1=Yes enter as digits
- FF. Education: does the site provide any educational/K12 material? 0=No, 1=Yes

Vita

Mark Freeman was born in Walthamstow, England in 1962. He attended college at the University of Lancaster (1980-1983) where he completed a B.A. Hons. in Economics and Accounting. Following this he successfully undertook a Post Graduate Certificate in Education at the University of Warwick. Mark started his teaching career at Greenshaw High School in Sutton, England and, after two years, he moved to Charleston High School in Nevis, West Indies through the Voluntary Services Overseas (V.S.O) program, teaching Economics and Accounting.

Mark moved to the United States in 1988. He worked for First Colony Life Insurance for five years, before moving to a position at Re:discovery Software as Client Service Manager. He still has an affiliation with Re:discovery Software, but in the last twelve years has worked through his own company, Stories Past, on nearly 50 website and installations for museums and archaeologists. Clients have included Monticello, Mount Vernon, Jamestown, and recently as the P.I. for the web catalog for the National Park Service. Mark entered the Masters of Science in Information Sciences program in 2012.