# A Model for Transferring Legacy Datasets to Living Documents: A Case Study Using a GIS Geodatabase for Archiving

**DAVID M. PLAZA**
**Eastern New Mexico University**

**Abstract:** Archaeology is faced with the inherit problem of managing legacy datasets, partly due to the high expense of maintaining them. Often these datasets are in a state of disrepair, thus rendering them underutilized, and difficult to properly archive or to integrate into the current archaeological dialogue. Unfortunately, this problem is a common issue and there is not an abundance of current literature on the subject. To address this dilemma, an examination of the condition of the records and artifacts of legacy datasets is needed. In this research, I will use the Anasazi Origins Project (AOP) as a case study and review the condition of the paper documents, such as the site records, of this legacy dataset. The AOP is a perfect example of an investigation that produced an invaluable dataset that was not fully published, analyzed, or properly preserved once fieldwork ended. I will begin by presenting the preservation methods of the AOP paper documents as a first step to archiving. Then, a GIS geodatabase will be used to convert the data from the paper documents of this legacy dataset to an electronic database for archival purposes. The advantage of an electronic database in archiving archaeological research is the ability to easily exchange, store, update, reorganize, and adapt data for various types of analysis in a cost effective manner. Geodatabases add a visual geographic context that is intuitive, offering archivists and archaeologists ease in organizing, inputting, and extracting data. The process described above is a model that produces a living document. This approach offers transparency, context, and provides accessibility for multiple disciplines. Lastly, I will present the AOP case study using this model to give a conceptual framework that is cost effective to transfer legacy datasets to living documents for duplication or adaptation to other similar legacy datasets.

## Introduction

Archivists and archaeologists share many of the same ethical principles and professional objectives. Both are concerned with the organization, preservation, and dissemination of information in an effort to manage cultural resources, such as records. Data from human activities are saved in part as records and should ideally be archive-ready. Data that is archive-ready is information in which its content and metadata that is deposited in an institution is prepared sufficiently for long-term preservation and dissemination without recourse to the original data producer.[1] Records are defined here as data or information that has been stored to a platform that has the capability to be preserved for the purpose of being used again. They have content that gives context and structure as an extension of memory or accountability of past activities.[2] One of the most common types of records in archaeology is the site record. This is used to record the location, environment, artifacts, in-field analysis, interpretations, and other information or

[1] Margaret Headstrom and Jinfang Niu, "Incentives for Data Producers to Create "Archive-Ready" Data: Implications for Archives and Records Management," Society of American Archivists – 2008 Research Forum (2008): 1-2. Accessed June 24, 2012. http://www2.archivists.org.

[2] Richard Pierce-Moses, "A Glossary of Archival and Records Terminology," Society of American Archivist (2012), http://www2.archivists.org, accessed on 8 October 2012.


Published by Society of American Archivists, November 2013

activities associated with an archaeological site and documenting it.[3] Records are important because the body of knowledge accumulated through publications, collections, and datasets are left to future generations for more powerful reasons than the simple transmission of knowledge and the perpetuation of information dissemination. One reason for the preservation and dissemination of information is to conduct research to answer questions about preceding generations to help improve the human condition of future ones. Therefore, data needs to be research-ready in addition to archive-ready.

In order for data to be considered research-ready, it must be in a format that has metadata, accompanying documents, tools to aid researchers that are relevant to the questions they are asking, and organized in a manner that does not require discourse from the institution which received it. One tool that has the capability of storing archive and research-ready data is a geodatabase. This paper will focus in part on the similarities between archaeologists and archivists in their effort to contribute to the dialogue of managing cultural resources, specifically paper documents. In addition, this case study will use the site records from the Anasazi Origins Project (AOP) collection to present a model that can be used to transfer a legacy dataset to a living document using a geodatabase. This article assumes the reader has basic knowledge of GIS.

**Background**

The two AOP field campaigns were conducted from 1964 to 1973 and from1982 to 1986.[4] These field campaigns took place in northwestern New Mexico. The focus of this investigation eventually became the Arroyo Cuervo region, northwest of Albuquerque, New Mexico, USA. A former faculty member of Eastern New Mexico University (ENMU), Cynthia Irwin-Williams, led this project. The purpose of the AOP was to investigate the antecedents of the Ancestral Puebloans, which she called the Oshara Tradition.[5] This project was fundamental to the understanding of the Archaic period in the northern Southwest. The research strategy of the AOP was a multidisciplinary approach that employed related disciplines such as geology, paleobotany, and paleontology. The investigation came at a time when studies on the Archaic northern Southwest of the United States were isolated at best.[6] The greatest contribution of this project was the development of the first regional unifying model for this area and time period.

The researchers of the AOP analyzed cultural materials in the field and collected some for further study.[7] The AOP collection is in part a result of gathered artifacts and associated documents from the two field campaigns. The site records are part of the associated documents of this project. There are two types of site records: the Museum of New Mexico Site Survey Form and the Eastern New Mexico University Site Survey Form.

---

[3] Thomas W. Neumann, Robert M. Sanford, and Karen G. Harry, "Cultural Resources Archaeology: An Introduction" 2nd Ed. Lanham: AltaMira (2010): 128-130.

[4] Bradley J. Vierra, "Administrative Report on the Status of the Anasazi Origins Project (AOP I) Collections Curated at Eastern New Mexico University" (Manuscript, Eastern New Mexico University, 1996).

[5] Cynthia Irwin-Williams The Oshara Tradition: The Origins of the Anasazi Culture". *Eastern New Mexico University Contributions in Anthropology* 5 (1973): 1-3.

[6] Cynthia Irwin-Williams, *The Archaic of the Southwestern United States: Changing Goals and Research Strategies in the Last Twenty-Five Years 1964 – 1989*, Ed. Bradley J. Vierra (Portales: Eastern New Mexico University Contributions in Anthropology, 1994): 572.

[7] Nicolas Merrill Chapin, "Hunter-Gatherer Technological Organization: The Archaic Period in Northern New Mexico" (PhD diss., University of New Mexico, 2005).

## Problem Statement

Managing collections that are in a state of disrepair or in an obsolete format can be a difficult task. There has been an abundance of literature on the ethics and reasons for preserving collections as well as best practices in management. However, there is a shortage of literature on reintegrating legacy datasets into the current dialogue of their respective disciplines. To address these issues, the site records of the AOP collection will be treated as a legacy dataset in this research. Legacy datasets are defined in this research as any set of data that has been passed from producers to the subsequent generation. It can be a dataset that is old, unused, obsolete, and/or disparate. Legacy datasets that are old, unused, or stored on an obsolete platform can easily be accessible if the platform in which they are saved is updated and are well organized. The process of updating the format in which a legacy dataset is stored can be done with ease if it is well kept, organized properly, and has metadata or associated documentation. Otherwise, legacy datasets can be challenging and can become an ethical and scientific problem. Moreover, the continuous accumulation of data further compounds this dilemma. Fortunately, the AOP collection is organized, however, it is in need of updating its format.

In part, to address the legacy dataset dilemma, an examination of the condition of the records and artifacts of legacy datasets is needed. This research will attempt to contribute to the current dialogue between archaeologists and archivists by offering possible solutions. A methodological direction to preserving and reintegrating legacy datasets by using traditional and digital preservation methods to properly archive and curate the site records of the AOP is used. Then, a GIS geodatabase is used to convert the data from the paper documents of the AOP legacy dataset to an electronic database for preservation and research purposes. The advantage of an electronic database in archiving and preserving archaeological research is the ability to easily exchange, store, update, reorganize, and adapt data for various types of analyses in a cost effective manner. Geodatabases add a visual interface for geographic context that is intuitive, offering archivists and archaeologists ease in organizing, inputting, and extracting data. This process is based on a living documents approach. It is a direction that offers transparency, context, and accessibility for multiple disciplines and establishes continuous relevance for preservation and research of datasets, such as the site records of the AOP legacy dataset.

## Methodology

The living documents approach can be used to modernize the format and organization of legacy datasets to offer methods that are fluid as well as transparent. In addition, this approach has the ability to adapt to the ever-changing cultural and physical environment. Living documents offer the ability to evolve as content is added, modified, or discarded. Living documents are defined here as datasets that meet the following four criteria. The first criterion is that the dataset is accessible for multiple parties. Second, the dataset can be integrated with datasets from other disciplines. Third, the dataset can easily be updated with additional data. Fourth, the dataset will have accompanying documentation, which can include articles that explain or give additional meaning to the dataset.[8] Geodatabases are among the best tools in creating living documents for research and preservation.

The living documents approach also offers ways to create contextual documents. This approach can function as a platform, not only to reinterpret legacy datasets, but also to prevent legacy datasets that have been upgraded into new platforms, such as a geodatabase, from becoming a

[8] Edward W. Tennant, "A Sample Geodatabase Structure for Managing Archaeological Data and Resources with ArcGIS." *Technical Briefs in Historical Archaeology* 2 (2007): 1-2.

legacy dataset in disrepair again. Here, the living document approach is used in an effort to make legacy datasets relevant in research.

A geodatabase is a tool that can be used to organize and structure data. This tool has the capability to store, preserve, and disseminate data for the purpose of management. Furthermore, a geodatabase can also be a great tool for updating legacy datasets into a modern format that is highly useful. An added benefit in storing data to this type of electronic database is when used in ArcGIS, tools can be applied to aid in answering archaeological questions.[9] ArcGIS is a GIS program designed by Environmental Systems Research Institute (ESRI). GIS is a system designed to obtain, store, search, analyze, and display geospatial information. Geodatabases are proprietary electronic central data repository tools designed by ESRI that use Microsoft Access for the management and storage of spatial data. A geodatabase is an object-based vector data model. This model uses x-, y-coordinates to represent features spatially (points, lines, and polygons). Furthermore, it models objects' spatial relationships digitally so that the computer can access and process them. An object is defined as an entity that has attributes and specific actions that can be performed. There are several types of geodatabases. In this research, a file geodatabase is used. The advantage of a file geodatabase is that it can virtually store up to 256 terabytes of various types of data in a single file folder for ease in data migration.[10]

In this research, all the original site records were accessed from the curation facility managed by the Department of Anthropology and Applied Archaeology at ENMU to photocopy onto acid-free paper. The original site records were put in a three ring binder by the original researchers. The binder is kept on file in the archives room of the curation facility. A catalogue number had been previously assigned to the site records. Most of the site records are relatively well preserved. Few records show signs of deterioration. This is possibly due to the nature of fieldwork. Two photocopies of each record were created on acid-free paper. The original, which should not be accessed again, and one acid-free copy of the site records are stored at the curation facility. The acid-free copy stored at the curation facility should only be accessed to make copies to replace the acid-free copy that is stored at the Department of Anthropology and Applied Archaeology curation lab at ENMU. The acid-free copy at the curation lab is for visiting scholars, faculty, and students to copy. Adobe PDFs (Portable Document Format) are then made using this copy to create an electronic version of the site records.

The electronic site records were saved as text-based PDFs. These records were scanned using the minimum museum standard of 300 dpi (digital pixels per inch). The minimum standard was chosen as a balance between quality and quantity. PDFs were saved at 300 dpi to ensure visual quality without occupying much space on digital storage devices. The purpose of creating an electronic document of each site record is to diversify the formats of the records. In addition, the PDF copy is much easier to duplicate and access than the paper counterpart. For instance, a PDF copy of the AOP site records can be copied or emailed to a researcher within a span of a few minutes. In addition, PDFs cannot be altered. Furthermore, space needed to store electronic documents is significantly less than the paper documents. The advantages of electronic copies of paper documents offer additional fluidity and transparency. When digital preservation measures are taken along with traditional means, insurance of long-term management and accessibility can be achieved.

---

[9] David Wheatley and Mark Gillings, *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. (London: Taylor and Francis, 2002): 1-5.

[10] Kang-tsung Chang, *Introduction to Geographic Information Systems*, 6th Ed. (New York: McGraw Hill, 2012): 50.

Once the preservation process of the paper documents is completed, data-preservation using keyboard data entry can begin. The process of creating an electronic database of the original documents is time consuming and labor intensive as with copying and scanning paper documents. The data can be saved in a word processor, spreadsheet program, or database management system.[11] Data of the site records were entered into an Access database, a database management system. Access was chosen because it is robust and accurate, meaning it is widely used and can be easily converted to other database management systems or spreadsheet programs. A database management system allows for easy and accurate reorganization of data. In this study, one table was created for each original site record type, and a table using the Laboratory of Anthropology (LA) format in Access. Creating tables of the original organization of the AOP legacy dataset site records is an additional preservation measure of the data and the provenance in which the dataset was deposited. Since the intent of the original creators in choosing their system of organization is unknown, preservation of the original organization of the AOP legacy dataset site records is necessary. The LA table format is used because it is the standard site form used for fieldwork in the state of New Mexico. By reorganizing the data to this format, it facilitates ease in integration with other datasets. Once the Access database is completed, the conversion process to a geodatabase can begin.

To create a geodatabase from an Access database, ArcGIS 10 and key attributes from the site records are needed. The ArcGIS component used in this research is ArcMap. The key attributes that are needed to create a geodatabase are a site identifier, cultural affiliation, and location. The site identifier is the site number or site name. This identifier is used to store or relate attributes. Cultural relevance of a site is used for organization of the geodatabase, such as time period or cultural group. Location is needed to geographically reference a site. The more accurate the location, the better the geodatabase will serve. For this case study, the geodatabase will need a projected coordinate system for the site records. A projected coordinate system displays the Earth's spherical surface onto a two-dimensional plane.[12]

In ArcMap there is a catalog option that allows viewing capacity as a window. In the catalog window, connect to the folder in which the Access database of the relevant data is stored. Next, under the folder in which the Access database is saved, create a file geodatabase. This option is listed in a drop-down menu by right-clicking the folder in which the Access database is saved. Then, create a feature dataset in the geodatabase. A feature dataset stores feature classes that share the same projected coordinate system and area of extent. A feature class stores spatial features of the same geometric type.[13] A series of windows will appear to name the feature dataset and assign a projected coordinate system. Since the UTMs recorded in the AOP site records are in NAD (North American Datum) 1927, the feature dataset needs to be in the same projection. Import the tables from the Access database to the feature dataset as feature classes so that they will be projected in NAD 1927. Since the standard is NAD 1983, partly because of its accuracy, the feature datasets will need to be projected into NAD 1983. The project tool in the ArcToolbox can transform the projected coordinates of the feature datasets from NAD 1927 to NAD 1983. Once this is completed, move each feature class from the geodatabase to the table of contents. The feature classes will show on the left hand side under the table of contents. All sites that have UTMs should appear in the data view window, represented by points.

[11] Robert V. Kemper, *The Potentials and Problems of Computers*, ed. Sydel Silverman and Nancy J. Parezo. (New York: Wenner-Gren Foundation for Anthropological Research Inc., 1992): 108.
[12] Chang, *Introduction to Geographic Information Systems*, 31.
[13] Chang, *Introduction to Geographic Information Systems*, 54.

A topographic base map will be needed to show the location of sites in relation to geographic features. The United States Geological Survey National Geospatial Program offers free topographic base maps on their National Map Viewer and Download Platform.[14] There are several types of data available for download, such as Elevation and Orthoimagery. In the National Map Viewer, there are several options to select and download an area. An email address will be needed for the downloaded data to be sent to. Once the data has been retrieved, save to the same folder as the Access database.

In this research, the type of Elevation data downloaded is a National Elevation Dataset (1 arc second) pre-packaged ArcGrid format in a GeoTIFF format. Due to the size of the Arroyo Cuervo Region, several sections of this area were downloaded. Import the downloaded GeoTIFFs as raster datasets to the geodatabase. Move each raster dataset from the geodatabase to the table of contents under the site record feature classes. The map should appear under the site records in the data view window. Next, the hillshade tool in ArcToolbox is used to help visualize the topography information of the raster dataset by creating a shaded relief appearance. This last step completes the geodatabase of the AOP legacy dataset with a visualization of the sites on a map. Using the identify tool, click on a point that represents a site, and all the attributes of the site from the Access database will show in a window (Figure 1).

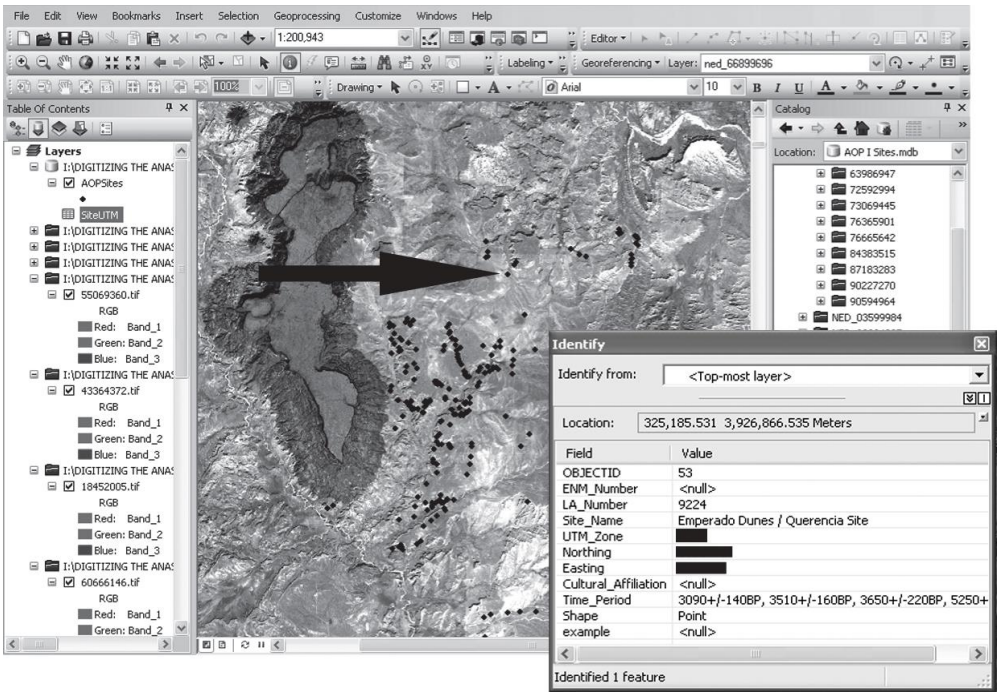


Figure 1. Example of the AOP geodatabase.

The end result is a geodatabase of the AOP site records, an electronic database that has a graphical interface that resembles a map. This electronic database has the capability for location and attribute information of a site to be projected with relational spatial information of geographic features on a platform that is intuitive. Moreover, this tool, as mentioned before, facilitates a living documents approach.

[14] United States Geological Survey National Geospatial Program, accessed on November 29, 2012, http://www.Nationalmap.gov.

The role of GIS in archaeology and archiving has increased rapidly within the last decade. Most disciplines that are concerned with the interpretation, storage, dissemination, and organization of spatial information are becoming savvier with the use of GIS. As mentioned before, the purpose of this paper is to present a model to store legacy datasets onto a platform in which it can be analyzed through a case study using the AOP legacy dataset. These methods demonstrate the use of a tool in which data can be managed in an archive and research-ready format. Using this case study, the proposed model can be applied to similar legacy datasets.

**Findings and Results**

Applying the proposed model to update legacy datasets offers opportunities to integrate with other databases. Centralized online digital repositories like the New Mexico Cultural Resource Information System (NMCRIS) offer additional storage options. The Archaeological Research Management Section maintains this state regulated online computer information system. NMCRIS can be used here to represent other state-run information systems. This online computer information system integrates geographic, management, and research-related data. Moreover, it offers security, access, and preservation for archaeological datasets.[15] The benefits to using this model to integrate legacy datasets to electronic databases that are accessible via the internet are the ease in access to relevant data that can be cost effective and time efficient, thus adding a larger dataset to research without duplication of effort.

Another online digital repository is tDAR. This online database is an international repository and archive managed by Digital Antiquity. The benefits of tDAR are the accessibility and preservation of archaeological datasets from across the globe that are continuously enhanced. The added bonus of this online repository and archive is that it is an excellent starting point for gathering data for research without ever leaving the home institution.[16] In other words, researchers from different countries can access valuable data of archaeological projects from other countries in an efficient manner that can be cost effective with ease.[17] Additionally, this can reduce the need to travel. By no means is this a replacement for fieldwork. Instead, it is a means to reduce the time and cost to extract data from previous and current archaeological projects.

The application of this model or any method to transfer a legacy dataset to a living document using an electronic database is truly limited to the creativity of the custodians of collections and researchers. The intention here is not to offer a static model, but a fluid model that is adaptive to other kinds of datasets. At best, this case study should offer a conceptual framework for reintegrating legacy datasets into the current archaeological dialogue.

**Conclusion**

This case study is an effort to update the AOP legacy dataset. In addition, the purpose is to contribute to the current dialogue of managing archaeological legacy datasets by presenting a

---

[15] Archaeological Records Management Section, "New Mexico Cultural Resource Information System: Guidelines for Submitting Archaeological Records." Accessed on June 9th, 2012. http://www.nmhistoricpreservation.org.

[16] Francis P. McMamanon and Keith W. Kintigh "Digital Antiquity: Transforming Archaeological Data into Knowledge." *The SAA Archaeological Record* 10 (2010): 38.

[17] Joshua Watts "Policies, Preservation, and Access to Digital Resources: The Digital Antiquity 2010 National Repositories Survey." *Reports in Digital Archaeology* 2. Accessed June 10, 2011. http://www.digitalantiquity.org/.

possible viable model for organizing, modernizing, preserving, managing, and disseminating these types of datasets. Currently, there are initiatives to modernize the data format of other components of this legacy dataset. Although there are parts of this dataset in need of transfer to a current platform for data preservation, there is still further research potential. This paper discusses preservation issues with archaeological legacy datasets, in particular the site record paper documents. The AOP legacy dataset can be used to answer preservation, archival, and archaeological research questions.

Future research that can build from this paper will be the integration of the other components of the AOP legacy dataset, such as the Excel spreadsheets of the cultural materials of each site with the geodatabase used in this research. Currently, Dr. John Montgomery, chair of the Department of Anthropology and Applied Archaeology at ENMU, is directing graduate and undergraduate students in curating, preserving, and archiving the AOP collection in an effort to manage long-term access and dissemination. These activities are recorded to an Excel spreadsheet and can be integrated into the AOP geodatabase from this research. Additional efforts for the preservation and dissemination of this legacy dataset will be its integration of the site records from the AOP collection and associated documents into NMCRIS and tDAR. Other potential research is in exploring viable tools such as Google Earth to create living documents and creating dataset-specific online geodatabases for legacy datasets. The limits to reintegrating legacy datasets into the current academic dialogue are defined only by the boundaries of our imaginations.

**Resources**


Archaeological Records Management Section, "New Mexico Cultural Resource Information System: Guidelines for Submitting Archaeological Records." Accessed on June 9, 2012. http://www.nmhistoricpreservation.org.

Chang, Kang-tsung, *Introduction to Geographic Information Systems*, 6th Ed. New York: McGraw Hill, 2012.

Chapin, Nicolas Merrill. "Hunter-Gatherer Technological Organization: The Archaic Period in Northern New Mexico." PhD diss., University of New Mexico, 2005.

Hedstrom, Margaret and Jinfang Niu, "Incentives for Data Producers to Create "Archive-Ready" Data: Implications for Archives and Records Management." *Society of American Archivists - 2008 Research Forum* (2008): 1-9. Accessed on June 24, 2012. http://www2.archivists.org/.

Irwin-Williams, Cynthia. "The Oshara Tradition: The Origins of the Anasazi Culture". *Eastern New Mexico University Contributions in Anthropology* 5 (1973): 1-28.

Irwin-Williams, Cynthia. "The Archaic of the Southwestern United States: Changing Goals and Research Strategies in the Last Twenty-Five Years 1964 – 1989." Edited by Bradley J. Vierra. Portales: Eastern New Mexico University Contributions in Anthropology, 1994.

Kemper, Robert V. *The Potentials and Problems of Computers*. Edited by Sydel Silverman and Nancy J. Parezo. New York: Wenner-Gren Foundation for Anthropological Research, 1992.

McManamon, Francis P. and Keith W. Kintigh "Digital Antiquity: Transforming Archaeological Data into Knowledge." *The SAA Archaeological Record* 10 (2010): 37-40.

Neumann, Thomas W., Robert M. Sanford, and Karen G. Harry. *Cultural Resources Archaeology: An Introduction* 2nd Ed. Lanham: AltaMira, 2010.

Richard Pierce-Moses, "A Glossary of Archival and Records Terminology," Society of American Archivist (2012), http://www2.archivists.org, accessed on 8 October 2012.

Tennant, Edward W., "A Sample Geodatabase Structure for Managing Archaeological Data and Resources with ArcGIS. *Technical Briefs in Historical Archaeology* 2 (2007): 12-23.

United States Geological Survey National Geospatial Program, accessed on November 29, 2012, http://www.Nationalmap.gov.

Vierra, Bradley J. "Administrative Report on the Status of the Anasazi Origins Project (AOP I) Collections Curated at Eastern New Mexico University." Manuscript, Eastern New Mexico University, 1996.

Watts, Joshua. "Policies, Preservation, and Access to Digital Resources: The Digital Antiquity 2010 National Repositories Survey." *Reports in Digital Archaeology* Number 2 (2011): 1-36. Accessed June 10, 2012. http://www.digitalantiquity.org/.

Wheatley, David and Mark Gillings, *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. London: Taylor and Francis, 2002.