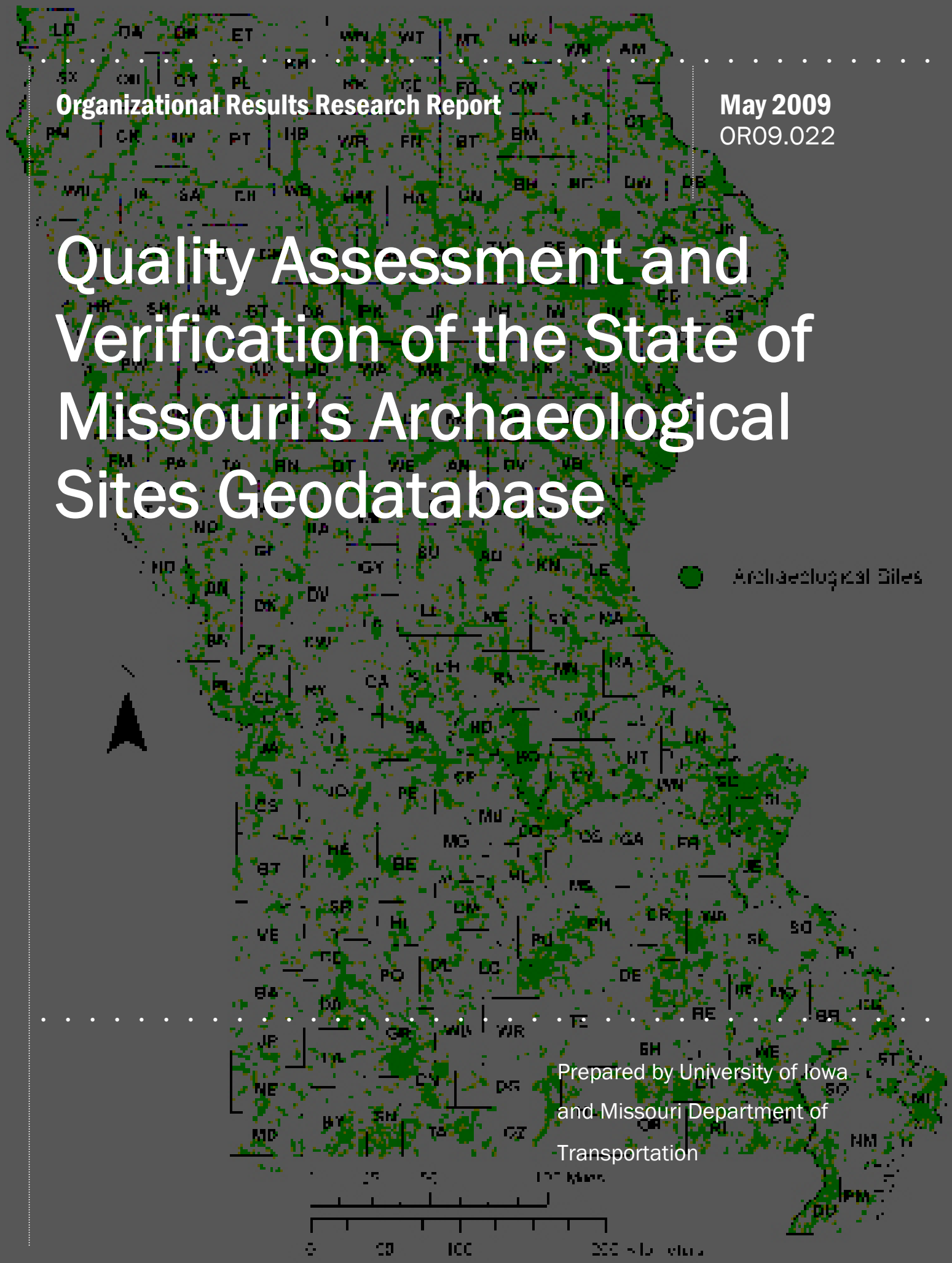
Organizational Results Research Report

May 2009
OR09.022

Quality Assessment and Verification of the State of Missouri's Archaeological Sites Geodatabase



Prepared by University of Iowa and Missouri Department of Transportation

FINAL REPORT

RI07-041

Quality Assessment and Verification of the State of Missouri's Archaeological Sites Geodatabase

Prepared for
Missouri Department of Transportation
Organizational Results

by

Joe Alan Artz, John Hall and Melanie A. Riley
Office of the State Archaeologist
University of Iowa – Iowa City

May 2009

The opinions, findings, and conclusions expressed in this publication are those of the principal investigators and the Missouri Department of Transportation. They are not necessarily those of the U.S. Department of Transportation, Federal Highway Administration. This report does not constitute a standard or regulation.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. OR 09-022	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Quality Assessment and Verification of the State of Missouri's Archaeological Sites Geodatabase		5. Report Date April 2009 6. Performing Organization Code MoDOT
7. Author(s) Joe Alan Artz, John Hall, and Melanie A. Riley		8. Performing Organization Report No. OR 09-022
9. Performing Organization Name and Address Office of the State Archaeologist The University of Iowa Iowa City, IA 52242		10. Work Unit No. RI07-041 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Missouri Department of Transportation Organizational Results P. O. Box 270-Jefferson City, MO 65102		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code MoDOT

15. Supplementary Notes

The investigation was conducted in cooperation with the U. S. Department of Transportation, Federal Highway Administration.

16. Abstract

The Missouri Department of Transportation (MoDOT) determined that a Geographic Information System (GIS) geodatabase of Missouri archaeological sites and surveys required a thorough quality control/quality assessment (QA/QC) by professional archeologists to ensure its accuracy for use in project scoping and development activities. The University of Iowa Office of the State Archaeologist undertook QA/QC for the geodatabase including initial validation and verification of a sample of counties, and recommendations for completing the remaining 99. An initial examination of the geodatabase identified over 12,500 geospatial and logical inconsistencies in data entry. Records for 5,518 sites and 1,781 reports from 14 counties, were examined. On average, 1.9 errors per report and 2.7 errors per site were encountered and corrected. For reports as well as sites, additions (data omitted during data entry) comprise two-thirds of all corrections. Only 6 percent of the reports and 12 percent of the sites required no changes to their database records. Some records could not be validated, often because their source documents are missing or have not been scanned. For the 14 counties, missing documents are less a problem for reports (2 percent missing) than sites (12 percent missing). In validating geodatabase entries against the scanned source documents, OSA averaged 41 minutes per report, and 13 minutes per site. Extrapolated to the remaining counties, an estimated 8,300 person hours is required to complete the geodatabase QA/QC. Fortunately, validation is greatly facilitated by the geodatabase, which is well designed, and the ArchEditor application that gives on-line access to the geodatabase for purposes of data entry and editing.

17. Key Words Archeology, Quality Assessment, Verification, Geodatabase		18. Distribution Statement No restrictions. This document is available to the public through National Technical Information Center, Springfield, Virginia 22161	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 42	22. Price

Form DOT F 1700.7 (06/98)

Table of Contents

Table of Contents v
Figures vi
Tables vi
1. Introduction 1
2. Background 2
3. Geodatabase Structure 3
4. Internal Consistency 4
  Geospatial Consistency 4
    Feature Geometry 4
    County Identifier 4
  Logical Consistency 4
  Summary 6
5. Data Validation 7
  Methods 7
  Reports 8
  Site Forms 9
6. Summary and Recommendations 10
  Geodatabase Schema 11
  ArchEditor Interface 11
    Panning and Zooming 11
    Digitizing Multiple Polygons 12
    Data Validation 12
    Problems Fixed During the Project 12
  Site and Survey Naming Conventions 13
  Duplicated Sites and Report IDs 14
  Time Estimates 14
7. Conclusions 14
8. Acknowledgments 15
9. References Cited 15

Figures

Figure 2.1. Survey (left) and site (right) density in Missouri counties, compiled from Arch9-07, a copy of the Missouri geodatabase that DNR-SHPO provided to OSA for use in the present project. 16
Figure 2.2. Archaeological sites (left) and survey areas (right) of Missouri and Iowa, compiled from Missouri's Arch9-07 geodatabase and Iowa's Iowa Site File GIS. 17
Figure 3.1. Entity-relationship diagram of the Arch9-07 geodatabase. 18
Figure 5.1. Geodatabase entries for 14 counties were validated against scanned documents. The counties are labeled with the number of sites (top number) and surveys (bottom number) in each county. 19
Figure 6.1. Results of geodatabase validation by county for surveys (a) and sites (b). 20
Figure 6.2. Kinds of corrections made to the geodatabase by county for surveys (a) and sites (b). 21

Tables

Table 4.1. Site Polygons with Geometry Errors 23
Table 4.2. Sites That Are Mislocated, Misnumbered, or Not in Missouri 24
Table 5.1. Attribute Values for Selected Geodatabase Tables and Fields. The values lists contain one or more terms for which OSA requested clarification from DNR-SHPO concerning their usage in Missouri archaeology.....25
Table 5.2. Summary of Changes Made to Survey Records During Validation. 27
Table 5.3. Documents Not Validated.....27
Table 5.4. Geodatabase Records Not Validated Because of Missing Report. 27
Table 5.5. Geodatabase Records Not Directly Related to Archaeological Site Recording.....28
Table 5.6. Geodatabase Records With Duplicate Surveys.....29
Table 5.7. Geodatabase Records Requiring DNR Action.....30
Table 5.8. Sites Not Validated. 30
Table 5.10. Summary of Changes Made to Site Records During Validation. 31
Table 6.1. Validation Results for 1,781 Reports, Summarized by County. 33
Table 6.2. Validation Results for 5,581 Sites, Summarized by County.....33
Table 6.3. Corrections Made to Geodatabase Records for 1,517 Reports, Summarized by County.....34
Table 6.4. Corrections Made to Geodatabase Records for 4,167 Sites, Summarized by County. 34
Table 6.5. Total Corrections Made to Reports for Individual Geodatabase Tables by County. 35
Table 6.6. Total Corrections Made to Sites for Individual Geodatabase Tables by County. 35
Table 6.7. Validation Work Effort.....36

Abstract: The Missouri Department of Transportation (MoDOT) determined that a Geographic Information System (GIS) geodatabase of Missouri archaeological sites and surveys required a thorough quality control/quality assessment (QA/QC) by professional archeologists to ensure its accuracy for use in project scoping and development activities. The University of Iowa Office of the State Archaeologist undertook QA/QC for the geodatabase including initial validation and verification of a sample of counties, and recommendations for completing the remaining 99. An initial examination of the geodatabase identified over 12,500 geospatial and logical inconsistencies in data entry. Records for 5,518 sites and 1,781 reports from 14 counties, were examined. On average, 1.9 errors per report and 2.7 errors per site were encountered and corrected. For reports as well as sites, additions (data omitted during data entry) comprise two-thirds of all corrections. Only 6 percent of the reports and 12 percent of the sites required no changes to their database records. Some records could not be validated, often because their source documents are missing or have not been scanned. For the 14 counties, missing documents are less a problem for reports (2 percent missing) than sites (12 percent missing). In validating geodatabase entries against the scanned source documents, OSA averaged 41 minutes per report, and 13 minutes per site. Extrapolated to the remaining counties, an estimated 8,300 person hours is required to complete the geodatabase QA/QC. Fortunately, validation is greatly facilitated by the geodatabase, which is well designed, and the ArchEditor application that gives on-line access to the geodatabase for purposes of data entry and editing.

1. Introduction

In 2007, the Missouri Department of Transportation (MoDOT) determined that a Geographic Information System (GIS) geodatabase of Missouri archaeological sites and surveys, maintained by the state's Department of Natural Resources (DNR), required thorough examination by professional archeologists to ensure its accuracy for use on a daily basis by the agency in project scoping and development activities. The Missouri Highways and Transportation Commission contracted with the University of Iowa Office of the State Archaeologist (OSA) to conduct this examination, the results of which are reported herein. OSA undertook the screening, verification, and correction of a four-county sample of records from the Missouri DNR geodatabase, and subsequently undertook verification of records from ten additional counties.

Verification is intended to give MoDOT a sense of the number and kinds of errors encountered in the geodatabase. In addition, 14 of Missouri's 113 counties were validated and are in good shape for use by the MoDOT, other agencies, and professional archaeologists to plan, preserve, and study archaeological resources.

Additional background on the Missouri archaeological GIS is presented in Section 2. The organization and structure of the GIS is discussed in Section 3. Section 4 presents the results of an initial cross checking of data fields within the GIS. This part of the study identifies the kinds and extent of errors and potential errors in the geodatabase, and documents the need for its thorough cleaning.

Section 5 details the methods and results of the 14 county verification effort. The numbers and kinds of errors are reported. In Section 6, these results are extrapolated to provide a quantitative estimate of the level of effort required to complete verification and correction of the remaining 99 counties. The concluding section also provides recommendations for improving the geodatabase design and its on-line manifestation, ArchEditor.

2. Background

Most states in the United States have a master inventory of recorded archaeological sites. This inventory is often maintained by a designated official such as a state archaeologist or state historic preservation officer. Historically, a state's site file consisted of paper forms on which information about sites was recorded, and paper maps on which the site location was plotted. Beginning in the 1970s and 1980s, many states began to migrate their site forms into computer databases. This migration gained momentum with the advent of personal computers. The next step in the evolution of state site files came in the 1990s and 2000s, with the increasing use by archaeologists of Geographic Information Systems (GIS) software (Artz and Eck 2009; Farley and Gisiger 1996; Kvamme 1999:162-164; Wheatley 1995). This software made it possible to digitize site locations in "real world" coordinates such as degrees of latitude and longitude, or the metric northings and eastings of the Universal Transmercator (UTM) system. The digital representation of the site is linked to tabular data about the site, such as cultural affiliation and investigation methods. The result is a relational database in which the geometry of the sites is incorporated as a kind of information that, along with the tabular data, can be queried and analyzed.

In 2005, an archaeological GIS was created for Missouri by the Missouri Resource Assessment Partnership (MoRAP) at the University of Missouri, in partnership with MoDOT and the DNR. Sites and survey areas (Figure 2.1) were digitized by the University of Missouri-Columbia, and tabular data from site forms and reports were also entered (Diamond 2007a). A 2007 accuracy assessment examined geodatabase records for 40 surveys and 40 sites. These 80 records contained a total of 1160 individual attribute values. Errors were found in 1 in 4 surveys and 4 in 10 sites (Diamond 2007a).

Additional problems came to light when the geodatabase as a whole was examined. For example, at least 7,200 site records (i.e., tabular data) had no associated polygon depicting the site boundaries, and over 1,500 survey polygons lacked complete attribute data (Diamond 2007a). A high incidence of site numbers were found whose county code did not match the county in which their polygon was located.

Errors continued to be found as archaeologists began to use the GIS and as the DNR's State Historic Preservation Office (DNR-SHPO) took over its management. Many of these errors were attributed to the fact that the data entry personnel had little or no familiarity with archaeology and thus misinterpreted information provided in the reports and site forms (Robert Reeder, MoDOT, 2008, personal communication).

The utility of the geodatabase for planning and preservation was perceived as compromised. Applications such as predictive modeling (Diamond 2007b) derived from the database were also of questionable value. In 2007, the MoDOT sent out a Request for Proposals (RFP) to conduct a thorough examination of the geodatabase against site forms and site reports.

OSA responded to the RFP because we had conducted a similar digitizing effort for Iowa's master inventory of archaeological sites, the Iowa Site File (Artz and Eck 2009). With funding from the Intermodal Surface Transportation Enhancement Act (ISTEA) and under contract to the Iowa DOT, OSA digitized Iowa's archaeological site locations and the Iowa SHPO digitized survey areas. Our digitizing included extensive data verification of the kind called for by the MoDOT RFP (Artz and Eck 2009). Another impetus for OSA's responding to the project is that we had long envisioned an archaeological GIS that crosses state boundaries, and this project was seen as a way to work toward that vision (Figure 2.2).

3. Geodatabase Structure

In April 2008, when this project was initiated, the DNR-SHPO geodatabase contained a total of 23,301 site polygons and 8,557 survey polygons, distributed among counties as shown in Figure 2.1. The structure of the geodatabase is outlined in the form of an Entity-Relationship (E-R) diagram (Figure 3.1). The geodatabase consists of 21 entities, including 2 geospatial datasets (SurveyPolygons, SitePolygons) and 19 tabular datasets. Relationships among the entities are shown on the E-R diagram by arrows and accompanying text.

The left and right halves of the diagram show data entities related to archaeological surveys and archaeological sites, respectively. The two halves are joined by the table Site2Survey, at top center in the diagram. Site2Survey is in many respects the most important table in the geodatabase. It tells which sites were investigated during which surveys, and vice versa.

Each table has a primary key (PK) and a foreign key (FK). The primary key is an identification number that is unique to every record in the table. For most entities, the primary and foreign keys are the unique Site_ID and Survey_ID that is assigned to each individual site and survey undertaking. The foreign key is a field containing values that a table shares with its related table, that serve to link the tables together. For example, the primary key in the SurveyData table is a Survey_ID number that is unique to each survey, and Site_ID is a primary key that identifies one and only one site in the SiteData table. In the Site2Survey table these fields are foreign keys whose purpose is to cross-link records between the origin tables (SurveyData and SiteData) and the destination table (Site2Survey). PKs and FKs are stipulated for all tables in the diagram.

The Site_IDs are Smithsonian Institution Trinomial System (SITS) numbers in which the prefix, 23, indicates the state of Missouri, a two-letter code designates the county, and an integer that is issued sequentially as sites are recorded within each county. A single letter suffix is sometimes appended to the site number. The Survey_ID follows a similar format, with the county code and sequential document number separated by a hyphen to prevent confusion of site and survey numbers. In the geodatabase, the prefix 23 is omitted from the Site_ID and Survey_ID.

Most tables in the geodatabase have composite relationships to their respective origin tables. A composite relationship is one in which an entity is dependent on the existence of its related entity. The deletion of a record in the origin table will delete all the related records in the destination table(s). For example, if a record is deleted from the table SurveyData, the deletion will cascade to all the destination tables like Survey Polygon, Methodology, SecondaryAuthor, ProjectType, and SurveyVerified.

On the other hand, records can be added to destination tables and not have a related record in the origin table. For example, site polygons and survey polygons can be added to the geodatabase without having a related record in SurveyData or SiteData.

Table 3.1 lists the field names present in each table of the geodatabase. Some fields accept Boolean (e.g., true-false, yes-no) values, others are populated from lists of accepted values, and others accept free text.

4. Internal Consistency

Errors in data entry can be detected in a database by checking records for logical consistency within and among fields. OSA conducted a series of cross checks looking for instances in which the data for a single site or a single survey contained conflicting information. Two kinds of tests, geospatial and logical, were performed. The results of all tests are saved as tables in dbf format for future reference.

GEOSPATIAL CONSISTENCY

Feature Geometry

The ArcGIS Check Geometry tool, run on the SitePolygons feature class, discovered 55 polygons with invalid geometry. Such errors can occur in several ways in a GIS, but most often are the result of digitizing errors. Table 4.1 lists the sites that have such errors. Errors of this type can often be fixed by manually editing the polygon to remove unwanted vertices and arcs.

County Identifier

The features in SitePolygons were spatially joined to a polygon shapefile of Missouri counties. The join produced a table in which each site was assigned the name of the county within which its digitized polygon falls. For 149 sites, the location of the digitized shape is not consistent with the two-letter county abbreviation that is part of the site's Smithsonian Trinomial System (SITS) number. Table 4.2 provides counts of the number of site polygons where the SITS county code is different from the county where the polygon is located on a map of Missouri. Sources for these errors might include coding errors during GIS creation, or erroneous assignment in the paper records. Checking against the paper records is probably the only way to conclusively determine each case.

Nine of the 149 geospatial inconsistencies are site polygons located in adjacent states. The site polygon for 32CK209 is located in Iowa, and in that state's site file is assigned the SITS number 13LE209. Eight site polygons are located in Arkansas (Table 4.2). Some of these are assigned Missouri county SITS codes, and others are coded "AR," presumably for "Arkansas." Three sites in Montgomery County, Missouri, also have an "AR" SITS code.

LOGICAL CONSISTENCY

The geodatabase was examined for instances in which the value of an attribute in one table should determine the range of values for attributes in other tables. For example, if the only entry in a site's Affiliation table is "Early Archaic," then "pottery" should not occur in its Material_Recovered table. If it does, the two tables are not logically consistent, and one or the other entry is probably erroneous.

Cross-tabulations of geodatabase table and attribute names identified 25 table pairs where logical inconsistencies might exist. Table joins and attribute queries were then devised in ArcGIS to test for the occurrence of such inconsistencies.

In the following, specific attributes are first identified by listing the table name, followed by an exclamation point, then the attribute name. Thus, "Methodology!METHOD_ID" identifies the attribute METHOD-ID in the Methodology table.

The purpose of this analysis was to identify and flag possible errors in the entire geodatabase, but not to correct them. Correction would require case by case evaluation, often requiring primary source checking, which is beyond the scope of this project except for the 14 specified counties.

Duplicate Values in Tables for Survey_ID and Site_ID

Test: **In the feature class SurveyPolygons, the field Survey_ID should have unique values.**
Result: Of the 8,557 polygons, 8,549 have unique values for Survey_ID. Eight polygons have non-unique Survey_IDs, as follows:

- **CR-115:** 2 polygons have this value
- **DV-028:** 2 polygons have this value
- **RI-106**: 2polygons have this value
- **Null:** 2 polygons have null values

Test: **In the table SurveyData, the field Survey_ID should have unique values.**
Result: All 9,414 records have unique values for Survey_ID. There are no duplicate values.

Test: **In the feature class SitePolygons, the Site_ID field should have unique values.**
Result: Of 23,301 records, 192 have a non-unique Site_ID. In these instances, the same site number is assigned to 2 or 3 polygons (Duplicate_SiteIDs_in_SitePolygons.dbf).

Test: **In the table SiteData, Site_ID should have unique values.**
Result: All 29,978 records have unique values. There are no duplicates.

Cross Check SitePolygons against SiteData

Test: **Every polygon in SitePolygon should have a corresponding record in SiteData.**
Result: 862 Site_IDs in SitePolygons have no matching Site_ID in SiteData. This means that a site location is digitized, but no information about that site is in the geodatabase.

Test: **Every record (site form) in SiteData should have a corresponding polygon (site location) in SitePolygons.**
Result: 7,727 SITE_IDs in SiteData do not have polygons in SitePolygons. This means that information for the sites (i.e., a site form) is on file, but the sites' locations are not digitized. Of these, 1,938 have a notation in the "Arch_Notes" field that states that no polygon was plotted because map data were poor or lacking. Another 412 are noted as recorded from a "short form with minimal information."

Cross Check SurveyPolygons against SurveyData

Test: **Every polygon in the feature class SurveyPolygons should have a corresponding record in the table SurveyData**
Result: 32 Survey_IDs in the feature class have no matching record in SurveyData. This means that 32 survey area polygons cannot be matched to data from a report document. Many of these are the result of inconsistent Survey_ID naming conventions, especially in capitalization. For example, ArcGIS fails to recognize a match between SurveyPolygons BU-152a and SurveyData record BU-152A.

Test: **Every record in SurveyData should have a corresponding polygon in SurveyPolygons**
Result: 893 Survey_IDs in the table SurveyData have no matching Survey_ID in SurveyPolygons. This means that 893 report documents cannot be linked to a digitized survey area in the GIS. Some reports may cover an area that is too generalized or poorly described for its boundaries to be mapped. Many of the unmatched records, however, are the result of inconsistent Survey_ID naming conventions. For example, a SurveyPolygon designated WE-64 is digitized, but in the table SurveyData, this identifier occurs in 14 records, designated WE-64A through WE-64N. The geodatabase structure does not allow ArcGIS to recognize the relationship between these 14 records and the survey polygon, if such a relationship exists.

Cross Check Site2Survey against SiteData and SurveyData

Test: **If SurveyData!ARCH_SITE is "Yes" (indicating that the survey encountered one or more sites), then that Survey_ID should have a matching record in the table Site2Survey.**
RESULT: 3 survey documents (KN-041, NO-047, PL-112) have a "1" (meaning sites present) in table SurveyData, but no matching record in table Site2Survey. Any sites recorded by these surveys cannot be linked to the survey document.

Test: **If SiteData!REPORT is yes (indicating a report is on file) that that Site_ID should have a corresponding record in Site2Survey**
Result: All 22,301 site records with Report=1 (meaning report on file) have matching records in Site2Survey, meaning that these site forms can be matched to records in SurveyData.

Internal Checks among Survey Tables

Test: **If Project Type!PRJCT_TYPE_ID is Phase II or Phase III, then Methodology!METHOD_ID should include at least one excavation method.**
Result: Of 871 reports identified as Phase II and Phase III projects, 121 do not list an excavation method in the Methodology section.

Test: **In studying the geodatabase structure, it appeared that reports of Phase I surface survey and/or shovel testing reports were often assigned a Project_Type of "Archaeology." Assuming that "Archaeology" in the database is equivalent to a "Phase I" survey, we tested the proposition that if a survey's Project_Type is "Archaeology" then its METHOD_ID should include at least one surface-survey method.**
Result: Of 8,324 records with Project_Type of "Archaeology," 398 records do not list a surface survey method.

Cross Checks Between Site Tables

Many of these tests involved checks that involved time-dependent attribute values in the Cultural Affiliation, Features, and Material Reported tables. Table 3 lists values for these three tables in terms of whether they represent the historic or prehistoric periods.

Test: **For any site, if SiteData!Collection is "Yes" then there should be one or more records for that site in the Material_Reported table.**
Result: 17 Site_IDs in SiteData are coded "Yes" for Collection, but have no record in the Material_Reported table. (Logic_Collection_Yes_Material.dbf)

Test: **Any Site_ID with reposed collections (i.e., has a record in the table Repository), should have a value of "Yes" for SiteData!COLLECTION.**
Result: 780 Site_IDs in the Repository table are not recorded as having a collection in SiteData!Collection.
Saved as: Logic_Repository_Yes_Collection_Yes.dbf

Test: **Any Site_ID with reposed collections should also have at least one record in the table Material Reported.**
Result: 12 records returned as NOT having material reported.
Saved as: Logic_Repository_Yes_Material.dbf 4/17/08; JH

Test: **Any Site_ID for which Material Report!Material_ID is historic ceramics, glass, or metal should be assigned a historic-period, Euroamerican affiliation (19-25) in CulturalAffiliation!Affiliation.**
Result: 287 sites that yielded historic ceramics, glass, or metal are not assigned a historic affiliation in Affiliation

Test: **Any sites from which prehistoric materials are recovered should be assigned a prehistoric cultural affiliation.**
Result: 1187 sites that yielded prehistoric artifacts are not coded for a prehistoric cultural affiliation. Many of these are sites for which cultural affiliation is "not reported."

Test: **Any sites with historic feature types should list a historic cultural affiliation.**
Result: 31 sites have historic feature types, but do not list a historic cultural affiliation

Test: **Any sites with prehistoric feature types should list a prehistoric cultural affiliation.**
Result: 121 sites have prehistoric feature types, but do not list a prehistoric cultural affiliation. It is possible that features like pits, caches, and hearths may sometimes be present in historic components.

Test: **Any site that has only an Archaic or earlier component should not have pottery.**
Result: Of the 753 sites that are recorded with only Late Archaic or earlier components, 16 have "pottery" as a recovered artifact type. Six of these (CE247, CS12, SL127, SL520, SL859, WN74) have Late Archaic components. Fiber-tempered pottery has been discovered at the Late Archaic Nebo Hill site in Missouri (Reid 1984), but this site is not in the above list.

SUMMARY

Perhaps the most serious inconsistencies identified in this analysis are sites for which the SITS county code disagrees with the site's digitized location. The risk of confusion or errors exists in data tracking and analysis tasks involving these sites.

The next most serious risks are those due to errors in naming conventions. These are especially critical when Survey_IDs and Site_IDs are involved, since they act to link together nearly all the tables in this relational database. The "health" of the geodatabase depends on every polygon having an identical Survey_ID as its matching document(s) in the SurveyData table.

Another situation where an improved naming convention seems called for is for sites and survey areas in adjacent states. OSA recommends that sites located in adjacent states be clearly identified in the geodatabase by prefacing their Site_ID with the appropriate SITS state code, followed by the site's county. In this manner, Arkansas Site_IDs would be prefaced by "3", Iowa sites by "13," and so forth. Missouri sites would continue to be listed in the geodatabase by their SITS codes minus the "23" state code.

Although not discussed in the preceding sections, there are occasional inconsistencies between names and organizations listed in both SurveyData and SiteData. For example, the U.S. Army Corps

of Engineers appears as “Army Corps of Engineers” and “Corps of Engineers.” The Depression Era “Works Progress Administration” also appears as “WPA.”

Inconsistencies among cultural affiliations, materials reported, and feature types are relatively common. OSA would recommend a systematic cross checking to identify sites for which cultural affiliations are “not reported,” but which could probably be assigned a “Prehistoric” or “Historic” affiliation based on recovered artifact or features.

5. Data Validation

METHODS

OSA checked geodatabase records for the 14 counties against the primary-source site forms and reports provided by MoDOT and DNR-SHPO. MoDOT uploaded these documents, scanned into Portable Document Format (pdf) files, onto a secure server, and provided OSA with a userid and password to download them via File Transfer Protocol (FTP). OSA stored the files on OSA’s Local Area Network (LAN), in folders to which only Geospatial Program staff had access rights. All of OSA’s work on this project was conducted in these secure folders.

The accuracy of the GIS-plotted site and survey area polygons were checked against maps provided in the reports and site forms. Tabular data in the geodatabase were checked against textual information in the electronic documents. To ensure that records were not missed, to track project progress, and to compile data on the kinds and frequency of errors, OSA developed a system to track geodatabase modifications. Each worker maintained his or her own Validation spreadsheet, created in Microsoft (MS) Excel. Data from the spreadsheets were dynamically linked to a MS Access database and appended to a master table containing all data validation results. The Validation database is a deliverable for this project, and will be useful to future researchers to independently evaluate the decisions made by OSA during the GIS verification process.

The validation database provides the name of the OSA staffer who made the change; date of change; source of information; the table corrected; and the specific action taken. Changes to site or survey polygons were further documented by noting whether the location, boundary, or labeling of the site was modified. A field called TYPE categorizes the action using the following codes.

- A = a record was added to the table
- C = a record was changed.
- D = a record was deleted.
- All = all the tabular information for the site or survey was entered from scratch
- None = no change was made to the geodatabase en try
- Missing = forms or reports that are missing
- No Map = forms or reports that don’t have a map or it’s too poor to map the site
- Other = none of the above

Some of line entries for a table fit in more than one category. For example, an entry for a site’s Cultural Affiliation may have been changed from “Historic” to a more specific category, such as “Early Industrial, Antebellum”, as well as adding a “Prehistoric” affiliation that was present at the site, but had not been recorded in the geodatabase. For this situation, the code AC was used. AD and CD were other possibilities employed. When a change to a geodatabase recorded required a longer explanation not easily viewed in an ArcGIS table field, that information is entered in a memo field in the validation database.

Validation was done in two phases. In the first phase, from May-November 2008, Boone, Cole, St. Charles, and Jefferson counties were validated. In the second phase, from January to March, 2008, an additional ten counties were digitized (Figure 5.1). The break between the two phases allowed OSA to review our procedures for ways to increase efficiency. The most common difficulty we had encountered involved the geodatabase linkage between sites and survey reports, which is handled in the table Site2Surveys. As designed, this table acts as a cross reference between reports and the sites described in those reports, and conversely, between individual sites and reports that provide additional information on the sites. To validate site data, reports as well as site forms must be checked.

In practice, we found that reports often contain information about sites for which no forms is available, or describe a site revisit for which a corresponding site form was never filed. In the first part of the project, site validation and survey validation proceeded simultaneously. This meant that the site validation team was often working with an incomplete list of report references. Once report validation was completed for a county, the sites had to be rechecked for report links that were not available the first time through. Consequently, on resuming work with the ten new counties, we decided to validate the surveys for a county first, and then begin site validation.

Early in the project, the need to standardize terminology used in the geodatabase tables was addressed. Currently no documentation of attribute definitions exists. To ensure that OSA was using terms in the same way as DNR-SHPO, we conferred with Kerry Nichols, the DNRs Cultural Resource Inventory Coordinator. We went through the attribute lists, identifying attribute values that might have ambiguous meanings. For example, the list of field methodologies included the value “excavation” as well as specific techniques, such as shovel testing, posthole testing, and test units. In this context, with the other specific excavation procedures enumerated, most if not all U.S. archaeologists working in cultural resource management would recognize as applying to large-scale, “block” excavations undertaken for Phase III. Decisions made as a result of that conference call are listed in Table 5.1.

REPORTS

A total of 1,781 reports were checked against scanned documents, distributed among counties as shown in Figure 5.1. Of these, 100 required no edits (Table 5.2) because the geodatabase entries, including the survey polygon, matched information in the report. Eighty two reports had no entry in the geodatabase, and for these, all tabular and geospatial data were entered for the first time. Modifications of existing geodatabase records consisted of additions, in which a previously unentered value was added to a table; edits, in which a previously entered value was modified, and deletions, in which a previous enter value was deleted (Table 5.2).

Adding values that had been overlooked during the initial data entry comprised nearly two thirds of the changes. A significant number of these were the 131 survey polygons that were added to the geodatabase for the first time. Deletion of erroneously entered values comprise about a third of the changes.

Many of the modifications to the Site2Survey table involved the deletion of a field number and addition of the SITS number. The user of field numbers to identify sites is presumed to occur when a report must be submitted before the “official” site numbers are assigned. In most cases, the SITS number could be determined by comparing maps in reports (where sites were labeled with field numbers) to site form maps, labeled with the SITS number.

Eighty three percent of all changes were made to the Methodology and Project Type tables. Most of these involved terms that have special meanings in archaeology, of which data entry personnel in the initial digitizing project were probably not aware. For example, most changes to the Methodology table consisted of the deletion of the value “excavation.” This value had been overused, assigned whenever the survey involved subsurface testing. The kind of testing (e.g., auger, posthole, or shovel

tests) was coded in addition to excavation. In archaeological methodology, however, “excavation” applies to large scale, Phase III excavation. Therefore, most occurrences of excavation in the table had to be deleted.

A total of 256 geodatabase entries for Surveys were not validated by this project (Table 5.3). These are divided into categories as follows:

1. Fifty-seven reports could not be validated because the report could not be located, or because the report lacked an adequate map for digitizing the survey area (Table 5.4)
2. Seventy reports were written for purposes other than archaeological survey or excavation (Table 5.5). These include SHPO letters of Section 106 determinations, literature reviews, archaeological overviews, historic architecture reports, cultural resource overviews, environmental assessments, and mitigation plans.
3. Eleven instances of the duplication of a report were encountered (Table 5.6). These are instances in which the same report was referenced by two Survey-ID numbers, or in which two polygons were digitized for the same survey area.
4. Twenty-six reports are identified as requiring DNR-SHPO action (Table 5.7). These are geodatabase records where the necessary modifications required an operation that was not possible in the ArchEditor interface.

SITE FORMS

Geodatabase records for 5,518 sites were checked against scanned site forms and reports. Of these, 729 (13 percent) could not be validated (Table 5.8) for the following reasons.

1. Eleven are SITS numbers assigned to locations where subsequent observations indicate an archaeological site was not actually present (Table 5.9).
2. Seventeen SITS numbers are numbers that have been assigned in duplicate (Table 5.9). For each of these, the geodatabase has 2-3 polygons, assigned different numbers, at the same location.
3. A total of 108 sites are in the geodatabase but could not be validated because no site forms or report references were found in the scanned documents.
4. A total of 551 sites are not recorded in the geodatabase or referenced in the scanned documents. They are believed to be valid assigned numbers, since they are checked off on Missouri Archaeological Society lists of assigned SITS numbers. These records may exist in the Society’s archives.
5. A variety of other problems precluded validation of 14 sites (Table 5.9).
6. For 18 sites, site forms and reports permitted validation of tabular data, but the reference documents either lacked a map, or the map was of such poor quality that the polygon could not be drawn.

The remaining sites were validated in full (Table 5.10). Of these, 632 (12 percent) were error-free and required no changes. Corrections were made to 4,157 sites (75 percent). Of the nearly 11,500 corrections (Table 5.10), 500 were sites that have never been digitized into the geodatabase, and 257 were sites for which no tabular data was previously present. Additions of previously unentered data comprised two thirds of all corrections, and a large portion of these (2,067 of 7,503) were made to the SiteData table. Most of these consisted of making an entry for Recording Organization, a field that was apparently overlooked during the initial data entry. Edits (changing one value to another, or correcting a free text entry) comprised 20 percent, and deletions 7 percent, of all corrections. Unfortunately, an OSA staffer completed 822 corrections noting only the table and not the type of

correction. Tabulated as A/E/D in Table 5.10, these changes are a combination of additions, edits, and deletions.

The SiteData tabular data required the most corrections (21 percent) followed in descending order by SiteType, SitePolygons, Material, TopographicLocation and Cultural Affiliation. Errors in the SiteType, Material, TopographicLocation, and CulturalAffiliation tables were dominated by errors or oversights made by encoders who were unfamiliar with archaeological and topographic/geomorphological terminology. Cultural affiliation was often listed as “prehistoric” but a more specific affiliation could be applied based on artifacts or features of the site. For example, burial mounds indicate a Woodland affiliation and a Nebo Hill point would indicate a Late Archaic affiliation.

6. Summary and Recommendations

At the beginning of this project, the geodatabase contained records for 23,301 sites and 8,557 reports. An initial examination of the geodatabase identified over 12,500 geospatial and logical inconsistencies in data entry, or 1 error per 2.5 sites/reports. A record-by-record check for a sample of 14 counties revealed a much higher frequency of errors.

Records for 5,518 sites and 1,781 reports were examined. Six percent of the reports and 12 percent of the sites required no changes to their geodatabase entries (Figure 6.1; Tables 6.1-6.2), but for every report that required no corrections, 15 needed changes. For sites, the ratio of changed to not changed records was 7:1.

Eight percent of reports and 13 percent of sites were not validated. The most common reason, especially for sites, was that scanned copies of the primary-source reports and/or site forms could not be located. Missing documents appear to be far less of a problem for reports (2 percent missing) than sites (12 percent missing). The relative frequency of missing reports varies from 0-3.7 percent across the 14 counties (Table 6.1), suggesting no major or systematic patterning. For sites, however, the percentage of missing documentation ranges as high as 30 percent for individual counties, and three counties (Jefferson, Platte, and Osage) account for 70 percent of the total missing site documentation (Table 6.2). The high incidence of missing site forms, and their spatial concentration in a few counties, is probably attributed in part to the historical division of the site form archive between the Missouri SHPO and the Missouri Archaeological Society.

The present project made a total of 14,374 corrections to geodatabase records for 1,517 reports and 4,167 sites (Tables 6.3-6.4). On average, 1.9 errors per report and 2.7 errors per site were encountered and corrected in the geodatabase. For reports as well as sites, additions comprise two-thirds of all corrections. Additions ranged from adding a value to a field in a table, to adding all the tabular data for a previously unentered report or site, to adding a previously undigitized survey or site polygons.

The relative frequency of adds, edits, and deletes varies considerably among the 14 counties (Figure 6.2), especially for sites. This suggests that there is no systematic or recurrent pattern in the kinds of errors that occur in the geodatabase.

Tables 6.5-6.6 track changes to individual data tables across the 14 counties. For report records, Methodology was the most frequently modified table in all 14 counties. The errors appear to be systematic and repetitive, and could probably be fixed for the geodatabase as a whole by a series of global updates to the table.

Corrections of site records, however, are not concentrated in any single table. Although most modifications (21% of the total) were made to the SiteData table, the kinds of corrections being made (Table 6.4) and the tables being corrected (Table 6.6) varied considerably from county to county.

In summary, with the possible exception of the Survey Methodology table, the errors found in the 14 counties are not systematic and predictable. This strongly suggests that the remaining counties can only be validated by examining the geodatabase record-by-record.

The two greatest sources of error in the geodatabase involve missing records. Many site forms are missing from the scanned documents, as are a smaller, but still substantial number of reports. Missing documents precluded validation of 139 geodatabase records, and further identified 557 sites that have been recorded but are not in the geodatabase. A second major source of error is highlighted by the amount of data that needed to be added to the geodatabase. By a margin of 2 to 1 (66% to 33%), more data was added to the geodatabase than was edited or deleted. Omissions appear to have been the greatest failing of the initial data entry project, and are probably the greatest shortcoming of the geodatabase as a planning and research tool.

GEODATABASE SCHEMA

Based on OSA's examination of geodatabase structure, and our intensive work with the system during validation, we recommend no major changes in geodatabase organization, in terms of its tables and the relationships among them. The Site2Survey table, in particular, does a good job of managing the complex many-to-many relationships between sites (each which can be documented in many reports) and reports (each of which can document many sites).

Data quality would be improved by the explicit definition of attribute values, as discussed in Section 5. We would recommend making these definitions part of the ArchEditor interface, so that users entering data will understand the meaning that is imparted to their selections.

Additional attribute values should be added to several tables. In Survey Method ID, a value for 'monitoring' would capture projects in which the archaeologists were monitoring construction rather than archaeologically excavating. In the site tables, we recommend adding more detailed attribute states such as chipped stone tools; ground/pecked stone tools; fire-cracked rock, and charcoal.

ARCHEDITOR INTERFACE

In addition to quality checking of records in the geodatabase, OSA worked extensively with the DNR's ArchEditor web application. This application went on line early in the project and greatly facilitated our work. As initially proposed, OSA was to work with a copy of the geodatabase, making all our revisions to that database, and subsequently returning the corrected version to DNR-SHPO. Given the complex relationships of tables and geospatial feature classes (Figure 3.1), it would have been a tedious, time consuming task to incorporate our revisions with the master geodatabase at DNR-SHPO.

Using ArchEditor, OSA interacted directly with the master geodatabase, adding and modifying data like any other remote user. At a regularly scheduled time each Friday afternoon, DNR-SHPO would take down the server and synchronize user modifications with the master tables. In the process of using ArchEditor, we thoroughly tested the interface. Problems we encountered and our suggestions for further improvement are detailed below.

Panning and Zooming

In ArchEditor, all editing must be done at the scale of 1:24,000 or closer. This is limiting in several ways. If a particular survey is excessively large, for instance a highway survey that extends across several counties, only a portion of the survey can be created or edited at one time. The interface does not allow panning the display while in editing mode, which would be necessary to create the entire polygon. Instead, a number of small polygons must be created and 'unioned' together. The same

problem arises when multiple polygons are associated with a single survey. A multipolygon survey, such as, for example the 13 culverts surveyed by report JE-263, is usually spread out across several townships and cannot be viewed unless zoomed out beyond 1:24,000. Editing this multi-part polygon requires the user to zoom and edit to each individual part.

We recommend a 'pan to' function as well as a 'zoom to' function, as the 'pan to' function allows the user to remain at the same scale. A 'pan to' function would also come in handy since there is no way to set a map scale specifically. The only way to set scale is to continually zoom in and zoom out until the correct scale is found. Once the appropriate scale is found, the "pan to" function would allow it to be maintained as the user "jumps" from polygon to polygon in the ArchEditor interface.

Digitizing Multiple Polygons

Some large-scale surveys are comprised of multiple, nonadjacent polygons, for example, a survey of culvert replacements along a stretch of highway. In a GIS (including the geodatabase), these can be represented as a "multipart polygon," in which the polygon is actually a collection of two or more polygons, grouped together into a single geospatial unit, and having a single set of records in the tabular data. To digitize or edit such polygons, the "polygon union" function is used. After the first polygon is created, the tabular data for the survey are entered. Then the polygon is selected, and the "quick zoom' function is used to move the display to where the next polygon will be placed. The "union polygon" feature is activated, and the second polygon is digitized. This links the two polygons into a multipart shape that shares the tabular data entered for the first polygon. This process is repeated until all the polygons have been digitized and unioned. In these examples, the "Polygon Union" feature allows the user to create an areally-extensive, multipart polygon. The drawback to the method is not being able to keep a polygon in editing mode while the view is panned, as discussed above.

Data Validation

A "data dictionary" defining the terms used in the geodatabase would be a useful resource to develop. As discussed above, many of the errors in the initial data entry occurred because encoders were not familiar with archaeological terminology. As demonstrated by our discussions with DNR-SHPO on this matter, archaeologists use a lexicon based on concepts that appear to be widely shared, but are often not explicitly defined. To avoid confusion and ambiguity that is often present in our jargon, a well-thought-out data dictionary is an indispensable component of any statewide archaeological database.

Additional fields within Survey Method ID would also help narrow down the type of survey for future users. One instance would be creating the field 'monitoring' for when an archaeologist is only monitoring an excavation instead of doing controlled excavation. While checking site forms, several instances were noted in which a subdivision of the Materials variable would created a more informative record. For example, the value "lithic tool" could be subdivided into chipped stone tools and ground/pecked stone tools. We also recommend subdividing the attribute value "lithics" to include flaking debris, fire-cracked rock, and other distinguishable kinds of stone material found on sites.

Problems Fixed During the Project

Early in OSA's use of ArchEditor, we noticed that the Survey Methods table was displaying only a single value, "shovel testing," for any record that was selected. Checking our copy of the geodatabase, we found that this was true even for records where additional methods were present. Even if multiple methods existed for a survey, the ArchEditor interface displayed only "shovel test." We reported the problem to Kerry Nichols and Stuart Harlan at the Missouri DNR, and it was soon fixed.

Another problem resulted from a geodatabase validation rule that was established after our work began. The format of previously entered Survey IDs is the two-letter, county SITS code, followed by a dash, then a three digit survey number. Zeroes were used to pad one and two-digit numbers (e.g. BR-001, BO-037). At one point, relatively early in the project, our attempts to enter previously unentered surveys resulted in the following ArchEditor error message.

- ERROR: Bad id format.
  The ID you have entered does not conform to standards. For sites, the format must be two characters (representing the county abbreviation) followed by 1-3 trinomial numbers separated by '/' characters. (Examples: 'BO1' or 'BO1/2' or 'BO1/2/3') For surveys, the format must be two characters (representing the county abbreviation) followed by a '-' character, followed by 1 trinomial number. (Example: BO-1) Make sure there are no leading zero's on the trinomial number for either sites or surveys.

DNR-SHPO, contacted for guidance, requested that we follow the new survey numbering convention, which omits the padded 0's.

We received the same error message when trying to enter a Survey_ID with a letter suffix, e.g., OS-38a. The geodatabase validation rules presently do not allow nonnumeric suffixes, which is a reasonable change for new data entry, but prevents adding older "legacy" data from ArchEditor.

One update to ArchEditor that proved very helpful was the addition of being able to see tabular data from a survey when there was no polygon associated with it. When validation with ArchEditor first began, tabular data could not be accessed if a polygon for that survey or site did not exist. From our previous work with our copy of the geodatabase, we knew that many sites and surveys, especially early ones, lacked sufficient geospatial data to digitize a polygon to represent them in the GIS. Tabular data, however, had been entered. Nichols, alerted to the problem in ArchEditor, modified the interface so that tabular data was viewable whether or not a polygon was associated with the site or survey.

If a site polygon is entered with no tabular information or link to a survey, you cannot go back later and open the feature's table for editing. A 'site id does not exist' error comes up instead. A work around is to delete the shape, redraw it, then enter the information when the attribute entry form pops up. This did not happen for every site that had no data in the attribute table, but the frequency of occurrence seemed high if the site polygon was not related to a survey. This phenomenon was not restricted to any one county.

SITE AND SURVEY NAMING CONVENTIONS

A number of reports that OSA validated contained only temporary site numbers. The temporary site numbers were those designated in the field by surveyors prior to receiving the official SITS number. For example, the Environmental Research Center of Missouri, Inc., or ERC, designated a temporary site in Boone County as 23BO-ERC-3, where 23 stood for Missouri in the Smithsonian Trinomial System, BO stood for Boone County, ERC stood for Environmental Research Center, and '3' stood for the third site found by ERC in Boone County.

When a query for the sites associated with a certain survey return only temporary site numbers, it indicated that either SITS numbers had never been assigned or, if assigned, the change had not been made in the geodatabase. Sometimes the SITS number would appear in ArchEditor's map, and other times the temporary site numbers would appear. When there were permanent site numbers in the ArchEditor map display, it was simple enough to compare the temporary site labels on the survey report map with the SITS numbers displayed by ArchEditor, and update the tabular data to show the permanent numbers. When temporary site numbers appeared in both the ArchEditor map and in the survey report, it was not possible to determine the permanent site number. For these sites, DNR-SHPO was contacted, and Kerry Nichols would provide the SITS numbers from his records, or assign a SITS number if one had never been issued. Once it became obvious that a number of reports contained only temporary site numbers, Mr. Nichols was asked for a spreadsheet correlating temporary and

permanent site numbers. We received this document on August 22nd, 2008, and it was subsequently used to replace temporary numbers with SITS numbers in the geodatabase.

DUPLICATED SITES AND REPORT IDS

We recommend that DNR-SHPO develop a set of rules for dealing with duplicate Survey and Site numbers. Duplication occurs when the same site or survey area is mistakenly assigned more than one number. Numerous examples were encountered in this project (Tables 5.6, 5.10). Duplicate numbers create confusion in the archaeological literature, and can lead to incomplete and erroneous query results. Ideally, when multiple numbers for a single entity are found, one of the numbers should be given precedence and the others withdrawn from use. In the Iowa Site File, OSA keeps a list of these so-called “retired site number” in a table that also lists the existing site numbers to which their records now pertain.

TIME ESTIMATES

In this project, a total of 2,395 person hours were spent validating the geodatabase records for 1,781 reports and 5,518 sites (Table 6.7). On average, 41 minutes were spent on each report, and 13 minutes on each site (1.5 and 4.7 per hour, respectively). No allowance is made for missing documents in the calculations in Table 6.7. If a site form was missing from the scanned documents, information about the site had to be searched for in reports. Time was also expended in correspondence with DNR-SHPO, tracking down missing documents.

To calculate the time required to complete geodatabase validation, we take as a basis the number of survey (8,557) and site polygons (23,301) in the geodatabase copy we received at the beginning of the project. We estimate that about 8,300 person hours would suffice to validate the remaining 6,800-some reports and 17,800-some sites (Table 6.7).

The per-record validation time could be reduced by an initial phase that would have two objectives. First, missing site forms and reports would be identified, located, and if found, scanned. Second, internal inconsistencies, including but not limited to those identified in Section 4 of this report, would be identified and corrected in the database. This would save the per-record validators from making repetitive modifications to tables such as Survey Methodology where certain kinds of data entry errors appear to have been frequently made.

7. Conclusions

Statewide archaeological data repositories throughout the United States evolve over time, and the current stage of that history, their transformation to digital formats, is never straightforward. The primary source data were collected over decades, in multiple formats, and under standards that, if not lacking, were inconsistently enforced. Digital datasets require a rigor that is not well adapted to the task of transforming these legacy data.

Missouri’s archaeological GIS is a well-designed and useful tool that has great potential for planning and research. It is built on software created by ESRI, the GIS industry leader, and makes use of ESRI’s geodatabase format that dynamically links map data and tabular data into a single relational database. Missouri is one of the few states that makes its geodatabase available on the Internet, and may be the only state that allows users to digitize newly recorded sites and survey areas, using the browser-based ArchEditor application. In OSAs opinion, the ArchEditor application places Missouri on the leading edge of archaeological data management in the nation.

Although the container is excellent, the contents are incomplete and error-ridden. The present project has attempted to partially remedy the problem by extensive quality control and quality assessment of geodatabase records for 14 counties. We have also estimated the work effort required to complete validation of the remaining 99 counties. While time-consuming and complex, validation will be a much less complex and lengthy task that it would be without the sound foundation offered by the geodatabase, and the excellent tools offered by ArchEditor.

8. Acknowledgments

The lead author thanks those who worked on the project at OSA. John Hall, Melanie Riley, John Willman, and Bryan Kendall did all the actual data verification work. Kayla Resnick assisted with map production for the report. The project would not have been possible without the assistance of Kerry Nichols at DNR-SHPO, whose already heavy workload was further burdened by our many requests and questions. Bob Reeder and Rebecca Geyer at MoDOT were also appreciated sources of knowledge and support. Finally, I'm grateful to Iowa State Archaeologist John Doershuk, who initially opened the doors for OSA's involvement in cultural resource work for the Missouri DOT. His vision, in keeping with the mission of our home university, has encouraged OSA to extend our research and service endeavors beyond our home state.

9. References Cited

Artz, Joe Alan, and Colleen R. Eck
2009 I-Sites: An Online Database and Geographic Information System for Archaeology in Iowa, USA. *Plains Anthropologist*, in press.

Diamond, David D.
2007a GIS Capture of Archaeological Surveys and Sites in Missouri: Final Report. Missouri Resource Assessment Partnership, University of Missouri, Columbia.
2007b Predictive Modeling of Archaeological Site Potential: Final Report. Missouri Resource Assessment Partnership, University of Missouri, Columbia.

Farley, J. A., and A. Gisiger
1996 Managing the Infrastructure: The Use of a Corporate Metadata for Archaeology. In New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research, edited by H. D. G. Maschner, 275–300. Occasional Paper 23. Center for Archaeological Investigations. Carbondale: Southern Illinois University.

Kvamme, K. L.
1999 Recent Directions and Developments in Geographical Information Systems. Journal of Archaeological Research 7:153–201.

Reid, Kenneth C.
1984 Fire and Ice: New Evidence for the Production and Preservation of Late Archaic Fiber-Tempered Pottery in the Middle-Latitude Lowlands. *American Antiquity* 49:55-76.

Wheatley, D.
1995 The Impact of Information Technology on the Practice of Archaeological Management. In Managing Archaeology, edited by M. A. Cooper, A. Firth, J. Carman, and D. Wheatley, 163–174. London: Routledge.

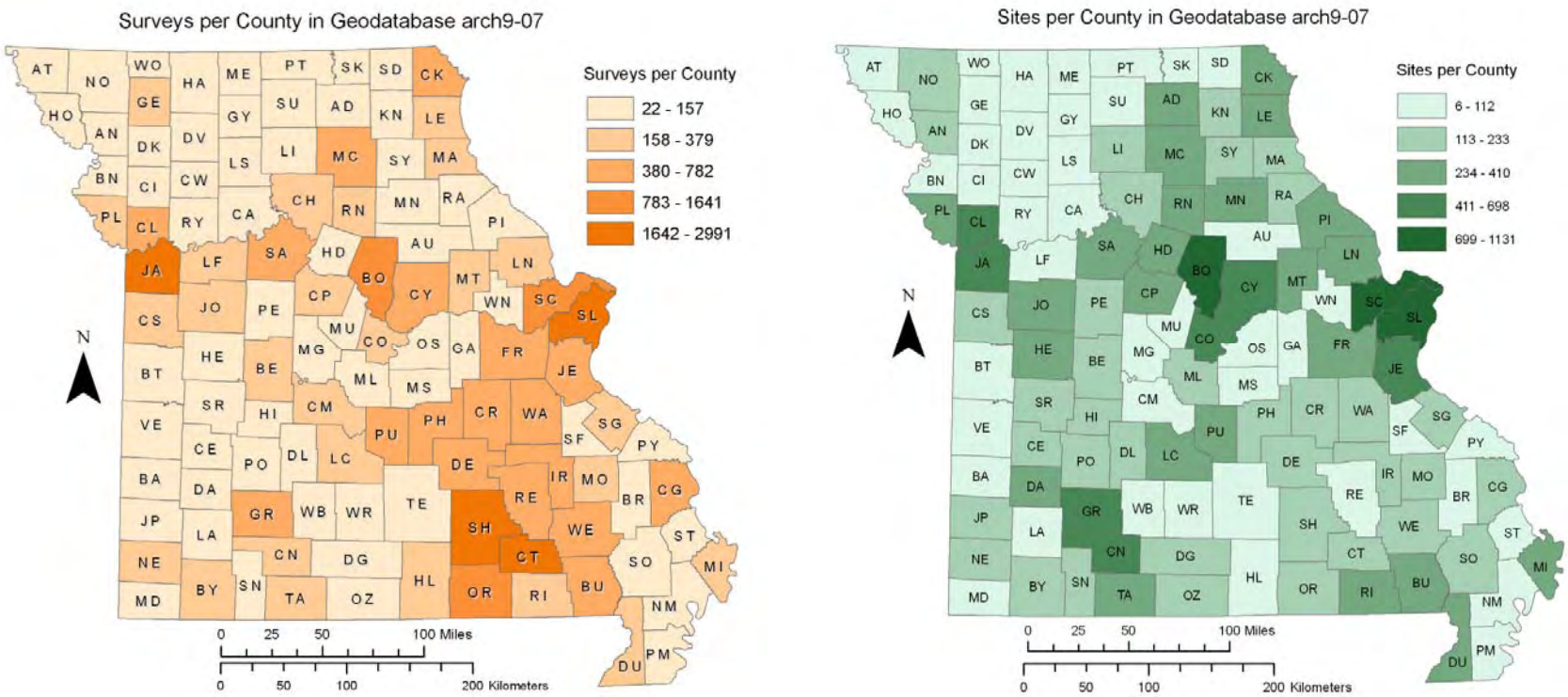


Figure 2.1. Survey (left) and site (right) density in Missouri counties, compiled from Arch9-07, a copy of the Missouri geodatabase that DNR-SHPO provided to OSA for use in the present project.

.

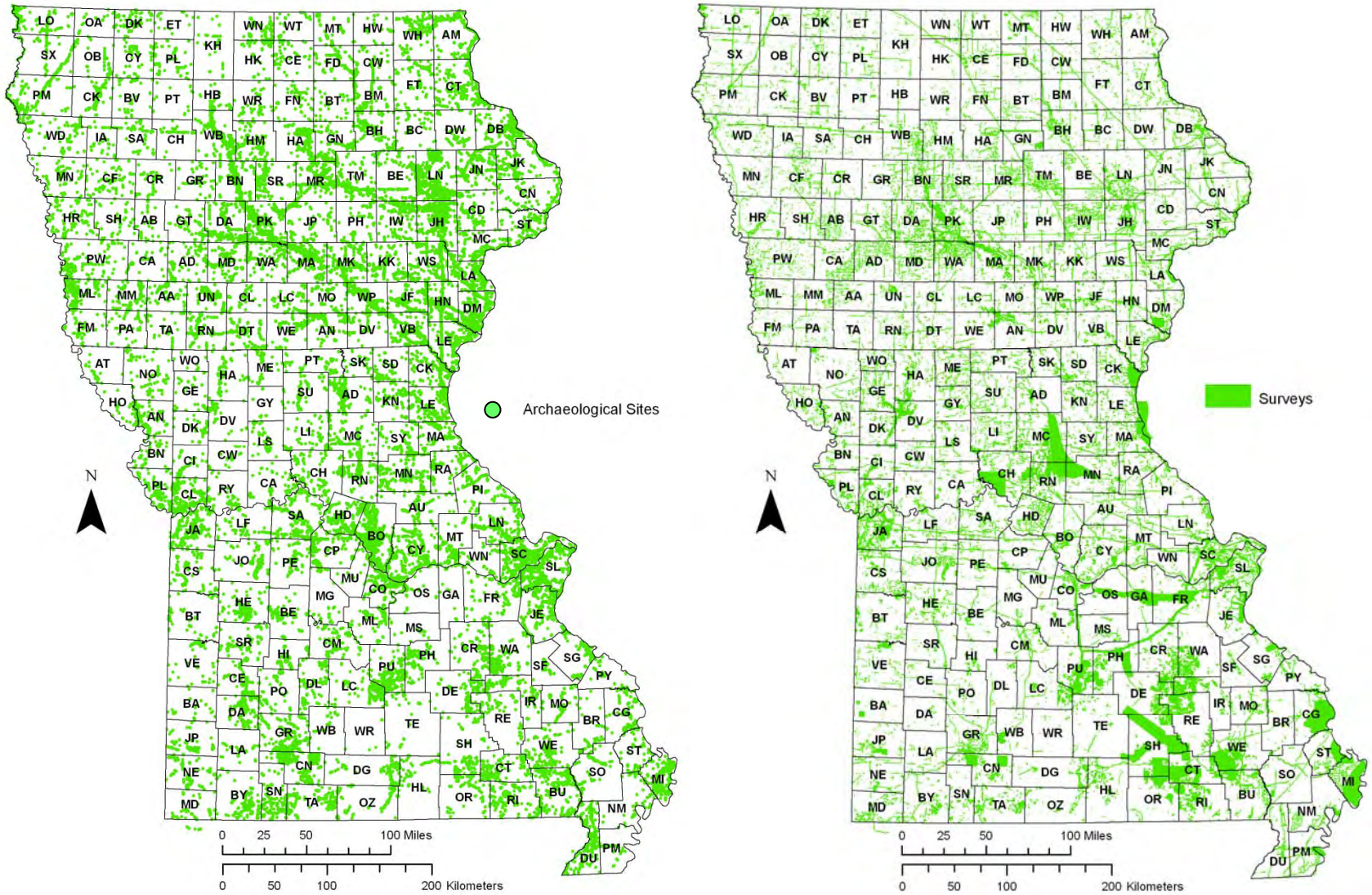


Figure 2.2. Archaeological sites (left) and survey areas (right) of Missouri and Iowa, compiled from Missouri's Arch9-07 geodatabase and Iowa's Iowa Site File GIS.

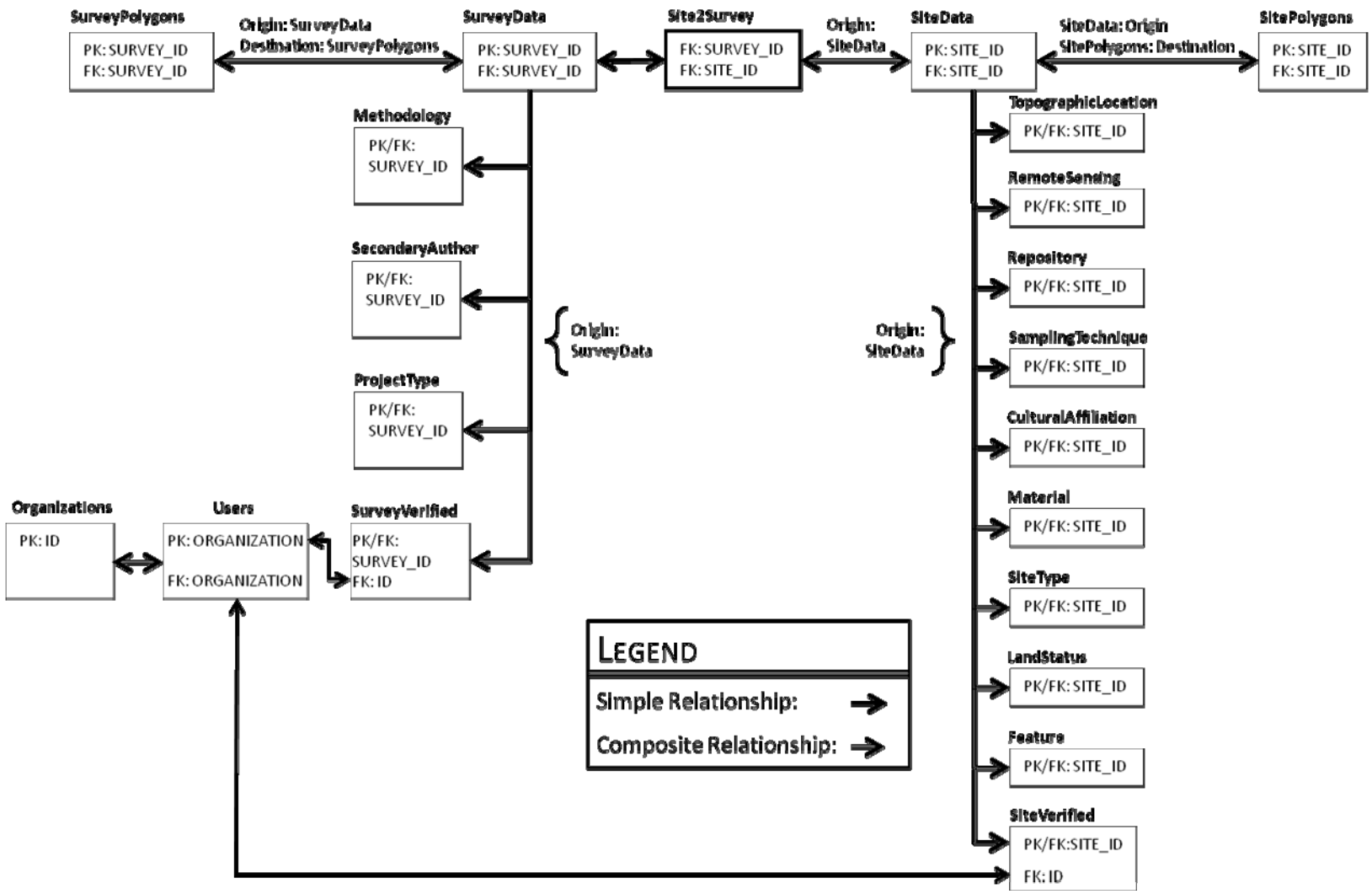


Figure 3.1. Entity-relationship diagram of the Arch9-07 geodatabase.

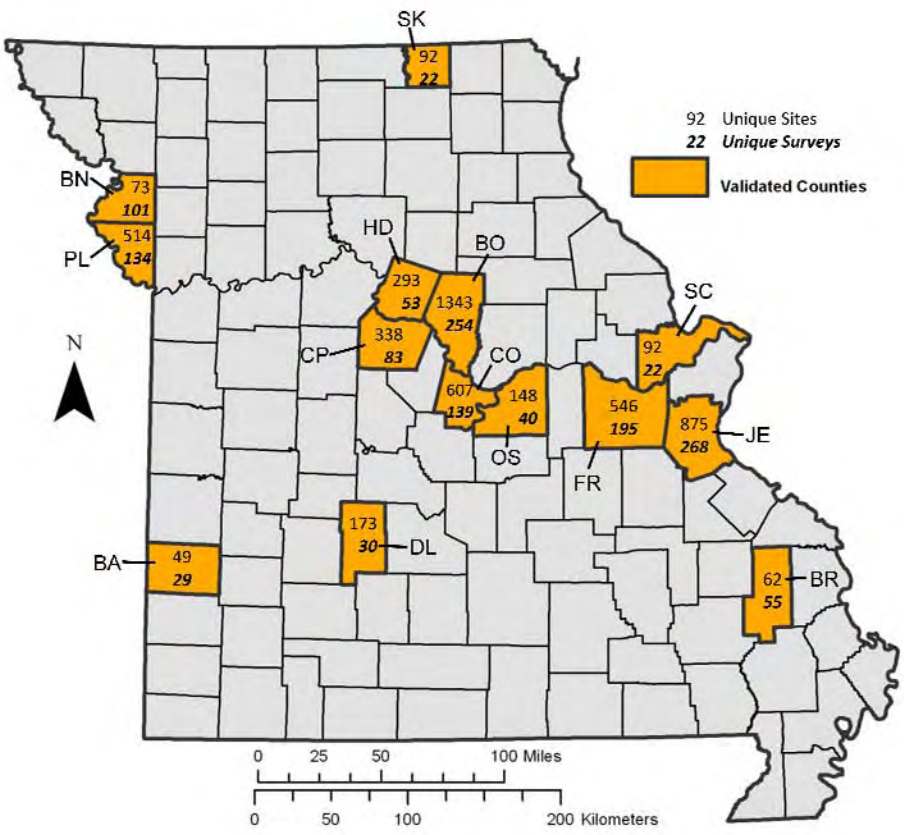


Figure 5.1. Geodatabase entries for 14 counties were validated against scanned documents. The counties are labeled with the number of sites (top number) and surveys (bottom number) in each county.

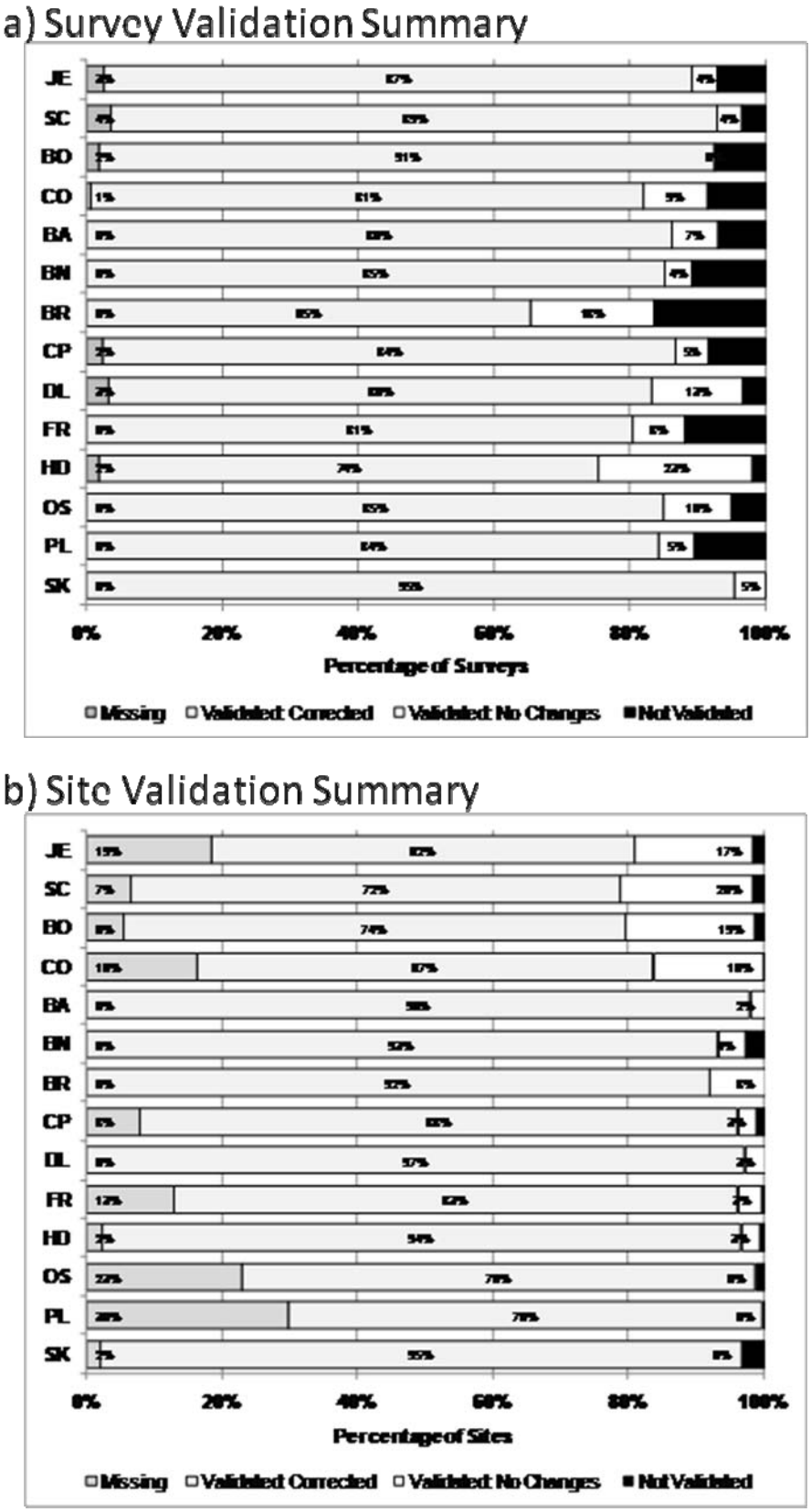


Figure 6.1. Results of geodatabase validation by county for surveys (a) and sites (b).

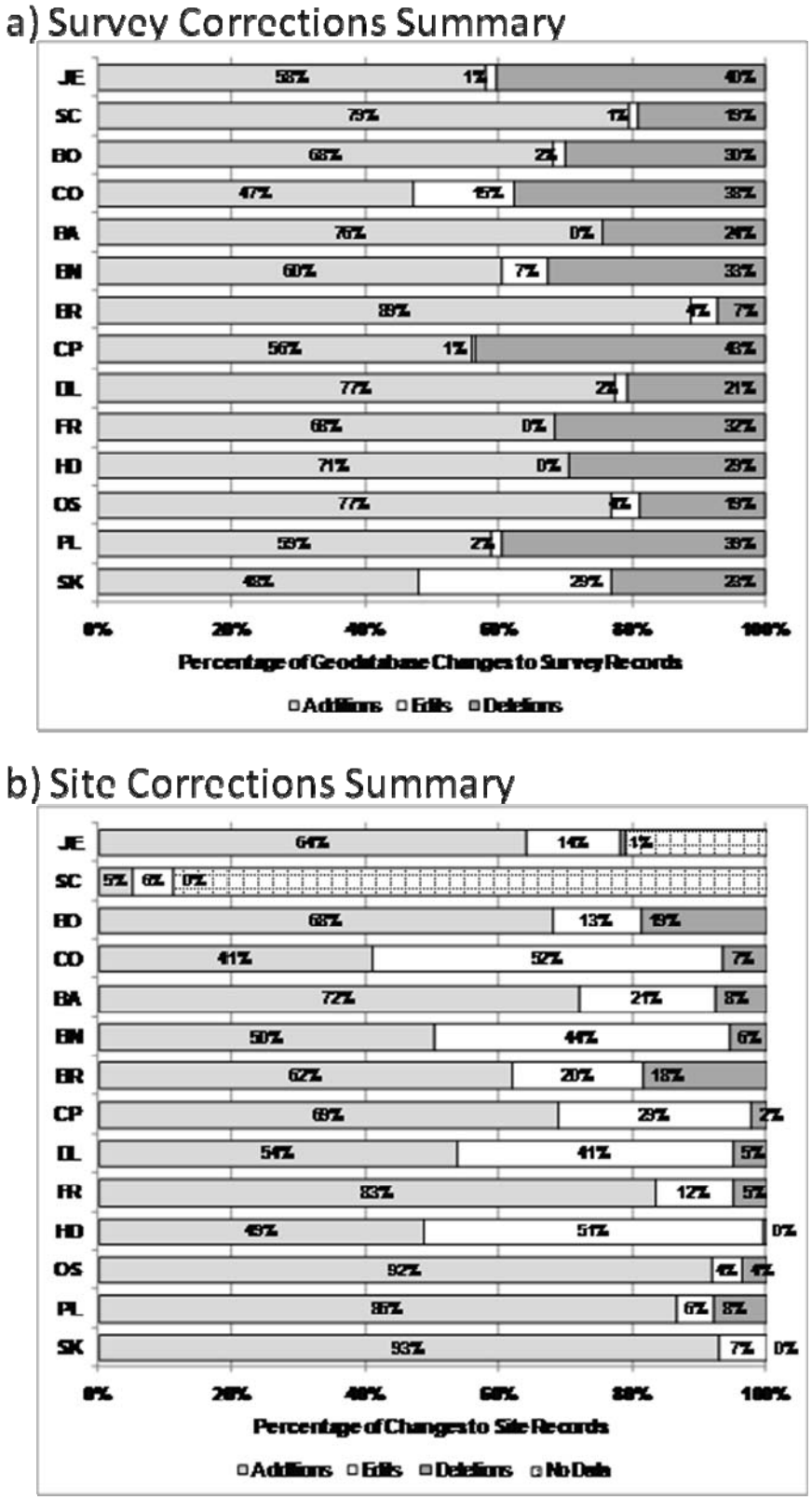


Figure 6.2. Kinds of corrections made to the geodatabase by county for surveys (a) and sites (b).

Table 3.1. Geodatabase Attribute Fields

Table	Field Name	Type	Table	Field Name	Type
SurveyData	ARCH_SITE	drop down	Site Data, continued	HUMAN_REMAINS	drop down
	AUTHOR	free text		NRHP	drop down
	CNDCTD_BY	free text		ARCH_NOTES	free text
	CNDCTD_FOR	free text			
	TITLE	free text	Cultural Affiliations	AFFILIATION	check box
	ARCH_DATE	free text			
	ARCH_NOTES	free text	Features	Feature	check box
Methodology	METHOD_ID	check box	Land Status	LAND_USE	check box
Project Type	PRJCT_TYPE_ID	check box	Material Reported	MATERIAL	check box
Other Authors		free text			
			Remote Sensing	TECHNIQUE	check box
SurveyPolygons	SHAPE_Length	self calculated			
	SHAPE_Area	self calculated	Repository	Repository	check box
Site2Survey	ObjectID	self calculated	Sampling Technique	TECHNIQUE	check box
SiteData	COLLECTION	drop down	Site Type	TYPE	check box
	DESCRIPTION	drop down			
	ILLUSTRATION	drop down	Topo_Location	TOPO_LOCATION	check box
	ARCH_DATE	free text			
	RECORDER	free text	SitePolygons	SHAPE_Length	self calculated
	RECORDING_ORG	drop down		SHAPE_Area	self calculated
	REPORT	drop down		NEW_EDIT	list
	REVISIT	drop down		NEW_EDIT	list

Note: Primary key fields (e.g., Site_ID, SURVEY_ID are common to all tables and not included above

Table 4.1. Site Polygons with Geometry Errors

County	Numbers	Count
CA	1172-1174	3
CK	10	1
LE	20	1
MD	1158	1
MT	1420-1421	2
NM	1618	1
PI		3
PI	1388	1
PL	1465-1471	7
PL	1476-1483	8
PL	1485	1
PL	1487-1488	2
PL	1490-1494	5
PL	1496-1505	10
PL	1507-1512	6
PO	1457	1
RN	1426	1
SH	1467	1
TOTAL		**55**

Table 4.2. Sites That Are Mislocated, Misnumbered, or Not in Missouri

Polygon Location	Site ID	Polygon Location	Site ID
County Conflict		**County Conflict (Continued)**	
ANDREW	HO50	POLK	DA460
	NO1132		DA461
BOONE	MU179	PULASKI	LC1399-1404
CALLAWAY	BO61		LC1406-1434
	BO124		TA231
	CO55-56	REYNOLDS	DE1214-1215
CARTER	SH334		DE1224
CLAY	CO1483-1486	SALINE	CA4
	PL35		CA5
CLINTON	CL427	SHANNON	PU22
	CL1088	ST CHARLES	LN29
COLE	ML245	ST GENEVIEVE	LC1366-1383
COOPER	BO461-462		LC1385-1395
	SA1572	STODDARD	BR1
CRAWFORD	DE1241-1243		BR21
DADE	CE321		BU208
FRANKLIN	CR320		BU1454-1458
	SL651	STONE	TA294
GREENE	CN76		TA312
	CN81	TANEY	FR150
HENRY	BE508	WARREN	MT50-MT54
HOLT	AN80		SC985
	AN100	WASHINGTON	CR237
HOWARD	BO788		CR255
	CP67		CR1533
JACKSON	CL266	WAYNE	BU1467
LACLEDE	PU1851		CT155
	PU1857		
MONITEAU	BO850	**State Conflict**	
	CO232	Arkansas	AR6
MONTGOMERY	AR12		AR16
	AR15		AR19
	AR18		AR20
MORGAN	CM195		DU310
	CM196		PV05-06
PHELPS	PU1838		RI146
PLATTE	BN387	Iowa (13LE209)	CK209

Table 5.1. Attribute Values for Selected Geodatabase Tables and Fields. The values lists contain one or more terms for which OSA requested clarification from DNR-SHPO concerning their usage in Missouri archaeology.

Table Name	Field Name	Values	Agreed upon Definition
Cultural Affiliation	**Cultural Affiliation**	Pre-Clovis	
		Paleo-Indian	
		Dalton	
		Archaic	
		Early Archaic	
		Middle Archaic	
		Late Archaic	
		Woodland	
		Early Woodland	
		Middle Woodland	
		Late Woodland	
		Terminal Late Woodland (Emergent Miss.)	
		Mississippian	
		Early Mississippian	
		Middle Mississippian	
		Late Mississippian	
		Proto-Historic	
		Historic Native American	
		Historic	Historic, uncertain specific period
		Colonial (1700-1803)	
		Territorial (1804-1820)	
		Antebellum (1821-1861)	
		Civil War (1861-1865)	
		Early Industrial (1866-1899)	
		Urban / Industrial (1900-1960)	
		Other	
		Prehistoric	Prehistoric, uncertain specific period
		Not Reported	
Material	**Material**	lithics	Includes flaking debris, fire-cracked rock, and any other non-tool stone artifact
		lithic tool	chipped, pecked, and groundstone tools.
		ceramics (prehistoric)	
		ceramics (historic)	
		human remains	
		floral	
		faunal	
		historic	
		glass	
		metal	
		building materials	historic (brick, cement) and prehistoric (daub)
		other	
		Not Reported	
SamplingTechnique	**SamplingTechnique**	pedestrian survey	
		shovel test	<50 cm diameter
		soil cores / probes	solid core extraction – Giddings, Oakfield
		auger tests	as defined in methodology table.
		trenching	
		test units	>= 50 cm square and <= 2 m square
		other	
		Not Reported	
SiteType	**SiteType**	lithic scatter	stone artifacts (inc. FCR) only

Table 5.1. Attribute Values for Selected Geodatabase Tables and Fields. The values lists contain one or more terms for which OSA requested clarification from DNR-SHPO concerning their usage in Missouri archaeology.

Table Name	Field Name	Values	Agreed upon Definition
		mound / cairn	
		cemetery / mortuary	
		habitation (prehistoric)	
		habitation (historic)	
		extraction camp	
		cave / rockshelter	
		quarry	
		historic scatter	
		rock art	
		commercial / industrial	
		public	
		military	
		governmental	
		trail / trace	
		other	
		Not Reported	
Methodology	**Methodology**	auger probe	mechanical or manual; flight, posthole, or bucket auger types
		excavation	any excavation larger than 2 x 2 m
		interviews	
		literature search	
		mechanical stripping	
		metal detector	
		pedestrian	any kind of surface walk-over examining exposed ground, including cutbank inspection.
		photographic analysis	any photo documentation in report
		plowing	plowing to improve surface visibility
		raking	
		surface collection	controlled, gridded surface collection.
		test pit	50 x 50 cm to 2 x 2 m
		transects	surveyors walking on transects or digging tests on transects
		trenching	
		screening	
		soil core	solid core extraction using an “Oakfield” or Giddings
		shovel testing	less than 50 x 50 cm in size
		GPS	
		Not Reported	
ProjectType	**ProjectType**	Archaeology	Phase I
		Architecture	
		Event Location	field schools? “pure research”
		Phase II	
		Phase III	
		Not Reported	

Table 5.2. Summary of Changes Made to Survey Records During Validation.

Table Name	None[a]	Adds	Edits	Deletions	Total[a]	Pct.
All Tables		82	0	0	82	2.8%
SurveyPolygons		131	39	14	184	6.3%
Site2Survey		93	3	58	154	5.3%
SurveyData		10	44	1	55	1.9%
Methodology		1342	4	673	2019	69.0%
Project Type		270	1	145	416	14.2%
SecondaryAuthor		8	0	1	9	0.3%
SitePolygons		0	8	0	8	0.3%
TOTAL[a]		1936	99	892	2927	
pct		66.1%	3.4%	30.5%		
No Changes	100				100	

Change" records.

Table 5.3. Documents Not Validated

Action Code	Definition	Number of Instances
DNR Action	Validation Requires DNR Action	26
Not Applicable	Document is not a survey or excavation report	70
Duplicate	The report or survey polygon duplicates another document in the geodatabase	11
Missing	Report is missing	31
No Map	Report lacks a map, or map is too poor to determine survey limits	26
Total		156

Table 5.4. Geodatabase Records Not Validated Because of Missing Report.

BO-045	DL-023	JE-252	SC-140
BO-056	HD-041	SC-037	SC-150
BO-163	JE-010	SC-060	SC-151
BO-189	JE-199	SC-077	SC-161
BO-193	JE-209	SC-106	SC-162
CO-100	JE-214	SC-118	SC-298
CP-052	JE-215	SC-120	SC-299
CP-068	JE-249	SC-122	

Table 5.5. Geodatabase Records Not Directly Related to Archaeological Site Recording

SURVEY_ID	Action
BA-027	pdf is a SHPO letter ONLY. Does not contain report. Letter pertains to CRM report # BA-028.
BN-001	pdf is a thesis, not a survey; no change made
BN-002	pdf is a thesis, not a survey; no change made
BN-075	pdf is documenting an historic bridge, not really a survey. Did not add survey polygon
BO-001	no action; pdf is not a survey report
BO-028	no change; pdf is a 'research design', not a survey report
BO-031	pdf is a thesis, not a survey report
BO-035	pdf is a record/lit review of previously recorded sites; no new surveys
BO-041	pdf is a publication/study, not a survey report; no change made
BO-046	pdf is a Phase III research design, not an actual survey report. No changes made
BO-050	pdf is a lit search/overview of previously recorded sites; no changes made
BO-074	pdf is a watershed management report, not a survey report; no changes made
BO-089	pdf is an 'Environmental Assessment', rather than an archaeological survey, see Surv. Polygn memo
BO-183	no action; pdf is a historical/architectural report
BO-228	no action; pdf is a historical/architectural report
BR-001	BR-001 is a dissertation on the Mississippian Culture in SE Missouri. Not a specific survey and shows no usable maps.
BR-002	BR-002 is a study of Indian Villages in SE Missouri. Not a specific survey and has no maps. No changes made.
BR-003	BR-003 is a lit review of the Archaeological/Historical resources in the lower Mississippi alluvial valley, not a new survey.
BR-004	BR-004 is a report on the St. Francis River Basin in Arkansas & Missouri. Not a survey. No changes made.
BR-005	BR-005 is a predictive model for cultural resources in the St Francis River Basin. A research design, not a new survey.
BR-007	report is a Cultural Resources overview of the Mark Twain National Forest, not a new survey.
BR-027	report consists of photos & locations of existing cell towers. No methodologies included in report. No changes made.
BR-043	report is a "Determination of Effect" for burn units by the UDSA Forest Service
CO-001	no action; pdf is not a survey report
CO-002	no action; pdf is not a survey report
CO-044	no action; Pdf. is a lit review
CP-001	pdf is an archaeological resource review of Blackwater-Lamine River Basin, no map included. No changes made
CP-002	pdf is an environmental impact review; no changes made
CP-041	pdf is a document concerning a sunken steamboat. Not a survey. No changes made.
CP-059	pdf is a review of all the mitigative measures used to protect a National Register property. No changes made.
CP-066	pdf is a historic engineering record. No changes made
CP-082	pdf is an historical/architectural review that doesn't have a map that shows survey area. Unable to add survey data
DL-021	DL-021 is an architectural survey that appears to be a continuation of DL-020. no maps included in report. No change made.
FR-011	report is a lit search, no changes made
FR-012	report is a lit search, no changes made
FR-064	pdf is a "Wetlands Mitigation and Monitoring Plan", not an archaeological survey. No changes made.
FR-068	pdf is a compilation of maps of previously recorded cultural resources. No change.
FR-070	pdf is a records & lit search.
FR-083	pdf is a research design proposal for a Phase III, not a survey. No change made.
JE-003	pdf is a 'museum brief', not a survey report; no changes made
JE-019	pdf is a final environmental impact statement w/ only basic archaeological survey info
JE-028	pdf is a 'guide book' for MAS Spring meeting, not a survey; no changes made
JE-036	pdf is a lit search of historic, prehistoric and architectural, cultural resources; no changes
JE-093	pdf is of a 'feasibility study', not an actual survey; no changes made
JE-094	pdf is a roadway proposal plan, not a survey; no changes made
JE-097	pdf is a 'Data Recovery Plan', not a survey; no changes made
JE-107	pdf is a lit search of archaeological & architectural sites within proposed hwy relocation
JE-127	pdf is a lit search of Archaeological and Historical records; no changes
JE-131	pdf is a 'Tractable Alternatives Analysis', not a survey; no changes made
JE-158	pdf is a proposal doc, not a survey report; no changes made
OS-014	report is a lit search and architectural survey. No changes made.

Table 5.5. Geodatabase Records Not Directly Related to Archaeological Site Recording

SURVEY_ID	Action
PL-005	pdf is an environmental evaluation, not a survey. Has no polygon. No change.
PL-007	pdf is paper on Kansas City Hopewell Lithics, not a survey. No change.
PL-017	pdf is an Environmental Impact statement. Not a survey. No changes made
PL-028	pdf is a bridge inspection report, not a survey. No change made.
PL-033	pdf is a paper "A Few Archaeological Observations", not a new survey. No change made.
PL-040	pdf is a short paper on a mound site, not a survey. No change made.
PL-057	pdf is 'Site Type Variablilty ' paper, not a survey. No change made.
PL-075	pdf is a review of prehistoric Indian sites in Platte Co. Not a new survey. No change made.
SC-009	pdf is a historical/architectural lit search; no survey polygon associated with report
SC-049	full survey report not included in pdf; see survey polygon
SC-062	pdf is an 'Environmental Assessment', not a survey; no changes made
SC-186	pdf is an 'overview' of known historical, archaeological and architectural resources; no change made
SC-205	pdf is a 'Research Design', not a new survey; no changes made
SC-209	pdf is a preliminary report for a habitat enhancement project, not a survey; no changes made
SC-235	pdf is a lit review, not a survey; no changes made
SC-285	pdf is an 'Environmental Assessment', not a survey; no changes made
SC-321	pdf is SHPO letter only. Unclear where polygon & tabular data in ArchEditor came from.
SC-322	pdf is a review of previous surveys, not a new survey; no changes made
SC-373	pdf is a data recovery plan, not a new survey; no changes made

Table 5.6. Geodatabase Records With Duplicate Surveys

SURVEY_ID	Action
BN-007	appears BN-007 is duplicate or excerpt from BN-004; no changes made
BN-096	Draft (duplicate) of BN-097; did not enter
BN-099	Duplicate data of BN-103 and BN-097; did not enter
BN-100	Draft (duplicate) of BN-097 and BN-096; did not enter
BN-105	Draft (duplicate) of BN-107; did not enter
FR-102	same as FR-080. No changes made.
FR-105	report is a duplicate of FR-082. no changes made.
FR-153	FR-153 is the same survey report as FR-091. No changes made
	survey report FR-166 is a duplicate of survey report FR-163. no change made.
JE-010	pdf is a duplicate of survey report JE-107; no changes made
SC-266	pdf SC-266 is actually a duplicate of SC-366

Table 5.7. Geodatabase Records Requiring DNR Action

SURVEY_ID	Action
BA-027	could not remove tabular data
BN-071	survey polygon is in Davis Co, report states survey in Davis Co, polygon labeled as Buchanan Co. survey; no change
BN-108	Polygon is same as BN-089; asked Kerry to copy it
BO-096a	Kerry reassigned polygon for BO-096a to BO-204 (8/28/08); see polygon memo for B0-096a
BO-205	survey report BO-205 does not match information in the map and table, under the number BO-205. Unclear if survey report BO-205 is mislabeled. No changes made. (added 8/28/08- pdf appears to be a mislabeled copy of survey report BO-097a)
BR-37	report number written on front of report is BR-036 and incorrect.
CO-013	per Kerry Nichols email 8/12/08, report needs to be rescanned
CO-108	per Kerry Nichols email 8/12/08, report needs to be rescanned
CO-125	per Kerry Nichols email 8/12/08, report needs to be rescanned
CO-128	per Kerry Nichols email 8/12/08, report needs to be rescanned
CO-133	per Kerry Nichols email 8/12/08, report needs to be rescanned
CP-023A	CP-023 had an addendum attached called CP-023A. The polygon in ArchEditor was labeled CP-023a (small 'a'), and didn't have any tabular data associated with it. However, when CP-023A (capitol 'A') was entered only tabular data would pop up. Emailed Kerry to fix labels and correct problem.
FR-035	attempted to add site FR334; ArchEditor returned that the site id does not exist in Site Data Table. Unable to add
FR-039	report is a lit search. No new survey. It is unclear how sites were chosen for site2survey. Did not change.
FR-049	attempted to add site FR334; ArchEditor returned that the site id does not exist in Site Data Table. Unable to add
FR-060	report had "FR-102" written at the top and is the exact same as pdf FR-102. No info for FR-060 in Arch Ed. Did not add.
FR-138	report has a temp site number 23FR-ERC-2. Unable to add perm# FR1552. not in site data table.
FR-146	multiple site numbers have been assigned to the same sites (also FR-079). Emailed Kerry
FR-160	pdf FR-160 is incorrectly named; survey actually in Shelby County and doesn't match survey polygon/data FR-160 in ArchEditor
FR-165	survey appears to be in St. Louis County rather than Franklin
FR-189	unable to discern what Range the survey is in. Sent note to Kerry to explain and have them digitized
HD-011	it appears that part of report HD-012 is attached to the end of this (HD-011) report.
JE-002	scans are poor quality; unable to discern what was done in new survey of bone bed
JE-107	both JE-107 and JE-107a are contained in pdf JE-107; two pdf, JE-107 and JE107a, were created from this pdf
JE-263	13 culverts to be digitized in ArcMap by Kerry per email.
OS-038a	unable to add polygon to ArchEditor. Program will not allow the 'a' after the report number. (also, township should be '42' not '24')

Table 5.8. Sites Not Validated.

Action Code	Definition	Number of Instances
Not a Site	Notes indicates the record is not of an archaeological site	11
Duplicate	More than one site number is assigned to the location	17
Missing	Polygon and/or tabular data are present in geodatabase, but there is no form or report to validate against	108
	No polygon, site form, or report referring to this number was found	551
No Map	Report lacks a map, or map is too poor to determine survey limits	18
Other	Other kinds of problems	14
Total		719

Table 5.9. List of Unvalidated Sites.

Type	Site_ID	Action
Not a Site	SK53	not a site
	SK54	not a site
	SK87	may not be an archaeological site
	HD1	No entry in geodatabase; no form or report; a later report suggests it may not be a site
	FR132	Not a site
	FR147	no entry in geodatabase; no form or report; a later report suggests it may not be a site
	JE334	non archaeological
	JE459	not clear that this is archaeological
	JE518	not clear that this is archaeological
	BO1090	No known archaeological material
	BO1093	No known archaeological material
Duplicate	CO1588	Site does not exist -number was incorrectly assigned to a site already on record (CO523)
	SC2051	Likely SC45
	JE154	Same site as JE2; same changes made as JE2
	JE321	Same as JE329
	JE458	same as JE456
	JE526	same as JE203
	JE527	same as JE190
	JE528	same as JE190
	JE530	same as JE216
	JE602	Site is duplicate of JE-599
	BO1276	Site number should be BO193 - encompassed by this site
	BO931	Site is the same as BO469; should not be a BO931
	BO939	No Report with field visit of site - same site as BO192?
	BO308	no edits, appears to be the same site as BO308B
	BO469	same site as BO931, one needs retired
	OS44	polygon is shared with OS97
	OS132	same as site OS146
Other	FR237	site number is no longer used, and has not been assigned to a site.
	SC21	This site is incorrectly identified in Survey SC163 The actual location according to all previous reports place it 1.5 km SW of the location tested. Site SC24 is identified at the location tested by Survey 163
	SC29	Incorrectly named LN29
	SC33	Incorrectly identified as SC32 in Survey 186
	JE333	Same as JE-360
	JE412	Site is duplicate of JE129
	BO612	Map is poor and polygon shape does not correspond to shape in the map. Info provided in site record form is quite different from info that I entered in the Table. Is there a missing site update?

Table 5.10. Summary of Changes Made to Site Records During Validation.

Table Name	None[a]	Adds	Edits	Deletions	A/E/D	Total[a]	Pct.
All Tables		755	2	-	-	757	6.6%
SitePolygons		543	873	9	-	1,425	12.4%
Site2Survey		5	1	1	-	7	0.1%
SiteData		2,067	127	4	247	2,445	21.4%
CulturalAffiliation		756	149	72	125	1,102	9.6%
Feature		183	6	2	183	374	3.3%
LandStatus		391	125	75	26	617	5.4%
Material		905	84	85	85	1,159	10.1%
RemoteSensing		15	1	13	-	29	0.3%
Repository		180	71	12	15	278	2.4%
SamplingTechnique		456	82	77	19	634	5.5%
SiteType		597	579	200	101	1,477	12.9%
TopographicLocation		650	187	285	21	1,143	10.0%
TOTAL[a]		7,503	2,287	835	822	11,447	
pct	--	65.5%	20.0%	7.3%	7.2%		
No Changes	632						

[a] Totals and percentages are for changed records (shaded area), and exclude the "No Change" records

Table 6.1. Validation Results for 1,781 Reports, Summarized by County.

Type	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	TOTAL	Pct
DNR Action	2		1	5	1	1				7		1			18	1.0%
Duplicate	1	1				5				4					11	0.6%
No Map	4	1	6	4		1				4			6		26	1.5%
Not Survey	11	11	11	3	1	3	8	6	1	6		1	8		70	3.9%
Other	1		1			1	1	1		2	1				8	0.4%
Missing	7	14	5	1	0	0	0	2	1	0	1	0	0	0	31	1.7%
No change	10	14	0	13	2	4	10	4	4	15	12	4	7	1	100	5.6%
Corrected	232	337	230	113	25	86	36	70	24	157	39	34	113	21	1517	85.2%
Total	268	378	254	139	29	101	55	83	30	195	53	40	134	22	1781	100.0%
Percent	15%	21%	14%	8%	2%	6%	3%	5%	2%	11%	3%	2%	8%	1%	100%	
% Validated	90%	93%	91%	91%	93%	89%	84%	89%	93%	88%	96%	95%	90%	100%	91%	
% Missing	2.6%	3.7%	2.0%	0.7%	0.0%	0.0%	0.0%	2.4%	3.3%	0.0%	1.9%	0.0%	0.0%	0.0%	1.7%	

Table 6.2. Validation Results for 5,581 Sites, Summarized by County.

Type	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	TOTAL	Pct
Duplicate	8	1	5	1								2			17	0.3%
No Map	1	3	7			2		4			1				18	0.3%
Not a Site	3		2							2	1			3	11	0.2%
Other	4	3	6							1					14	0.3%
Not Validated	16	7	20	1	0	2	0	4	0	3	2	2		3	60	1.1%
Missing	163	27	75	99				27		71	7	34	154	2	659	11.9%
No Changes	151	79	254	98	1	3	5	9	5	18	8	0	1	0	632	11.5%
Corrected	545	291	994	409	49	68	57	298	168	454	276	112	359	87	4167	75.5%
Total	875	404	1343	607	50	73	62	338	173	546	293	148	514	92	5518	100.0%
Percent	15.9%	7.3%	24.3%	11.0%	0.9%	1.3%	1.1%	6.1%	3.1%	9.9%	5.3%	2.7%	9.3%	1.7%	100.0%	
% Validated	80%	92%	93%	84%	100%	97%	100%	91%	100%	86%	97%	76%	70%	95%	87%	
% Missing	18.6%	6.7%	5.6%	16.3%	0.0%	0.0%	0.0%	8.0%	0.0%	13.0%	2.4%	23.0%	30.0%	2.2%	11.9%	

Table 6.3. Corrections Made to Geodatabase Records for 1,517 Reports, Summarized by County.

Type	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	TOTAL	Pct
Additions	245	437	294	106	31	124	87	85	41	212	48	53	148	25	1936	66.1%
Edits	6	8	9	34	0	14	4	1	1	0	0	3	4	15	99	3.4%
Deletions	170	105	129	85	10	67	7	66	11	98	20	13	99	12	892	30.5%
Total Corrections	421	550	432	225	41	205	98	152	53	310	68	69	251	52	2927	100.0%
Percent	14.4%	18.8%	14.8%	7.7%	1.4%	7.0%	3.3%	5.2%	1.8%	10.6%	2.3%	2.4%	8.6%	1.8%	100.0%	
Reports Corrected	232	337	230	113	25	86	36	70	24	157	39	34	113	21	1517	
Corrections per Report	1.8	1.6	1.9	2.0	1.6	2.4	2.7	2.2	2.2	2.0	1.7	2.0	2.2	2.5	1.9	

Table 6.4. Corrections Made to Geodatabase Records for 4,167 Sites, Summarized by County.

Type	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	TOTAL	Pct
Additions	703	34	1628	450	105	100	101	733	244	1123	257	617	1252	158	7505	65.6%
Edits	154	42	314	576	30	88	32	307	188	160	268	30	84	12	2285	20.0%
Deletions	8	0	451	72	11	11	30	25	23	65	2	25	112	0	835	7.3%
No Data	231	591	--	--	--	--	--	--	--	--	--	--	--	--	822	7.2%
Total Corrections	1096	667	2393	1098	146	199	163	1065	455	1348	527	672	1448	170	11447	
Percent	9.6%	5.8%	20.9%	9.6%	1.3%	1.7%	1.4%	9.3%	4.0%	11.8%	4.6%	5.9%	12.6%	1.5%	100.0%	
Sites Corrected	545	291	994	409	49	68	57	298	168	454	276	112	359	87	4167	
Corrections per Site	2.0	2.3	2.4	2.7	3.0	2.9	2.9	3.6	2.7	3.0	1.9	6.0	4.0	2.0	2.7	

Table 6.5. Total Corrections Made to Reports for Individual Geodatabase Tables by County.

Table	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	Total	Pct
All	11	7	8	0	1	5	12	3	4	10	6	2	10	3	82	2.8%
Methodology	328	388	284	149	32	132	39	118	27	236	42	42	178	24	2019	69.0%
Project Type	39	77	96	19	5	35	24	20	9	31	7	15	36	3	416	14.2%
SecondaryAuthor	1	1	3	3	0	0	0		1						9	0.3%
Site2Survey	17	55	18	14	1	8	4	4	4	12	4	5	8		154	5.3%
SurveyData	6	8	8	5	0	1	3	2	1	1	1	1	4	14	55	1.9%
SitePolygons	0	0	0	0	0	7	0	0	0	0	0	0	0	1	0	0.0%
SurveyPolygons	19	14	15	35	2	17	16	5	7	20	8	4	15	7	184	6.3%
TOTAL	421	550	432	225	41	205	98	152	53	310	68	69	251	52	2927	100.0%
	14%	19%	15%	8%	1%	7%	3%	5%	2%	11%	2%	2%	9%	2%	100%	

Table 6.6. Total Corrections Made to Sites for Individual Geodatabase Tables by County.

Table	JE	SC	BO	CO	BA	BN	BR	CP	DL	FR	HD	OS	PL	SK	Total	Pct
All	246	22	65	44		4				186	119			71	757	6.6%
CulturalAffiliation	96	85	48	111	25	25	14	139	30	167	21	85	254	2	1102	9.6%
Feature	38	169	57	10	10	10	2	12	10	20	5	4	26	1	374	3.3%
LandStatus	46	7	119	34	6	8	16	85	28	105	11	71	77	4	617	5.4%
Material	76	50	196	105	16	29	18	146	62	123	63	82	187	6	1159	10.1%
RemoteSensing				27						2					29	0.3%
Repository	23	3	4	17		21		49	3	60	30	48	20		278	2.4%
SamplingTechnique	25	12	150	128	9	7	6	61	18	94	20	20	81	3	634	5.5%
Site2Survey		2		5											7	0.1%
SiteData	320	205	623	154	31	51	37	210	132	174	42	108	343	15	2445	21.4%
SitePolygons	85	45	204	240	29	18	32	131	71	188	117	87	113	65	1425	12.4%
SiteType	101	61	440	160	11	18	26	115	59	129	84	82	188	3	1477	12.9%
TopographicLocation	40	6	487	63	9	8	12	117	42	100	15	85	159		1143	10.0%
TOTAL	1096	667	2393	1098	146	199	163	1065	455	1348	527	672	1448	170	11447	
	10%	6%	21%	10%	1%	2%	1%	9%	4%	12%	5%	6%	13%	1%	100%	

Table 6.7. Validation Work Effort.

	Reports	Sites
Number validated in this project (14 counties)	1781	5518
Person hours spent on validation	1210	1185
Number validate per person hour	1.5	4.7
Numbers remaining to be validated (99 counties)	6776	17783
Estimated person hours to validate	4517	3784
Estimated person weeks (40 hours per week)	113	95

Missouri Department of Transportation
Organizational Results
P. O. Box 270
Jefferson City, MO 65102

573.526.4335
1 888 ASK MODOT
innovation@modot.mo.gov