InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Overview

Case Study 14:
Archaeological Records in a Geographical Information System:
Research in the American Southwest

Peter Gagné, Université Laval

August 2006

The Creator Context / Activity

Creator: Center for Desert Archaeology (CDA) in Tucson, Arizona.

Creator type: Scientific focus / Private sphere (corporate). A small, private non-profit institute dedicated to research. In some ways, the creator may be closer to a university research project than to a private corporation.

Juridical context: Grew out of the former Arizona Division of the Institute for American Research, a not-for-profit organization founded in 1982. The not-for-profit aspects of this organization became the Center for Desert Archaeology, legally incorporated in 1989. The for-profit aspects became Desert Archaeology, Inc.

Specific pertinent legislation includes:
- National Historic Preservation Act 1966
- Archaeological Resources Protection Act 1979
- Native American Graves Protection and Repatriation Act 1990
- Environmental Policy Act 1966
- Wilderness Act 1964
- Arizona Antiquities Act 1973
- Arizona Historic Preservation Act 1982
- Arizona Burial Protection Act 1990
- Arizona Revised Statutes

Archaeologists may also find their actions governed by professional codes of conduct and standards of research performance guidelines promulgated by the professional associations to which they may belong, as well as by the organizations by whom they may be employed.

Activity: The CDA project "Coalescent Communities in Arizona" (A.D. 1200-1540), which included a database and related Geographic Information System (GIS) application relating to

the aggregation and migration of prehistoric peoples in the American southwest. Researchers investigated the modeling of interactions between archaeological sites in the region. Among other issues, they looked at cost surface analysis, which calculates the least "cost distance" between two points (in this case, the cost is energy expanded by prehistoric peoples). Population density analysis, archaeological site preservation and stewardship were also observed.

Specific activities include research, data input and data analysis. Related to the activity studied in this case study, though distinct from it, the CDA has an active publications program to disseminate information to a wider audience.

Nature of Partnership
The project has several strong ties to AZSITE, a similar database in the region.[1] "A large amount" of the initial data for the Coalescent Communities GIS database was provided by the AZSITE database. (FR 2) The records created in the Coalescent Communities project are therefore highly related to those of AZSITE. In addition, the development of the AZSITE database greatly affected the creation and maintenance of the Coalescent Communities database. "Issues that have arisen during the creation and management of the AZSITE database have affected the way the Center for Desert Archaeology designed its CCD-GIS [Coalescent Communities database GIS]." (FR 2)

In return for "taking" from AZSITE, the Coalescent Communities database will also give back to it. "Once the Coalescent Communities Project is at a somewhat completed state, the Center for Desert Archaeology will share the database with AZSITE. Thus, AZSITE's very large GIS is connected to the Coalescent Communities GIS. *This connection is not formally documented* and not extremely significant except as a comparison of methods of approach to geospatial data management within archaeology." (FR 24, emphasis added)

In addition to the apparent influence and input of AZSITE, the CDA refers in their grant proposal for the Coalescent Communities project to "the process of incorporating data from diverse institutions into a single database" and to their plan to "meet with many archaeologists across the state to review these maps and to obtain input regarding "missing sites"" (FR 2) The final report adds, "The Coalescent Communities database is also connected to the core data sets of three researchers…These three datasets are very crucial to the development and refinement of the GIS. The documentation of this relationship is in its formative stages. The Center for Desert Archaeology is beginning to link the archaeological data to its source (repository, researcher, project, etc. for each archaeological site entry in the database)." (FR 25) The "database is also linked to federal spatial data that the GIS Specialist obtains mainly from the World Wide Web." (FR 24)

Bureaucratic/Organizational Structure
The center is governed by a four-member Board of Directors: President, Vice-President, Treasurer and Secretary. In addition, there is a six-member Advisory Board. The President is

---

[1] "AZSITE (http://www.azsite.asu.edu/) is a consortium made up of the Arizona State Museum, Arizona State University Department of Anthropology, the Museum of Northern Arizona and the State Historic Preservation Office. These four agencies have created a statewide GIS regarding archaeological sites, surveys and archaeological districts." (FR 22, footnote 11)

also the CEO. In addition, there is a Programs Manager, an Office Assistant, and a Content Editor (for *American Southwest* journal). Employees consist of seven Preservation Archaeologists (who carry out the day to day operations) and three Preservation Fellows (who carry out research). The CDA employs one archaeologist whose primary function is ""to develop and manage a Geographic Information System...incorporating a variety of archaeological and environmental data from the southwestern U.S. and northwestern Mexico."" (FR 7)

The center receives private funding in addition to significant state and federal funding. Federal funding includes a 1997 Arizona Humanities Council grant, a 1999 National Science Foundation grant, a 2001 National Endowment for the Humanities grant and the Arizona Heritage Fund. The final report states that "the acceptance of state and federal funding carries with it various administrative responsibilities," (FR 7) which are outlined in Appendix A in the Final Report.

Digital Entities Studied
The digital entity studied is a database and related GIS called "Coalescent Communities." The database and GIS consist of "compilations of pre-recorded archaeological site data...compiled from datasets contributed by individual researchers and various repositories, as well as from data extracted from published sources held by libraries and related institutions." (FR 2) The core dataset is represented in both text and numeric characters, while the outputs are textual and graphic in nature [map(s) alongside tabulated data]. The GIS database is created within Microsoft Access and tables; other aggregations of data are sometimes exported into Microsoft Excel.

The primary file formats used include: Access database (MDB), Excel spreadsheet/flat files (XLS), word processing files [Word (DOC) and ASCII text (TXT)], maps, images (JPG) Adobe Portable Document Format (PDF), EndNote (LIB) and files related to the ArcView software (SHP, SHX, DBF, SBN, SBX and PRJ files). ArcView, an ESRI product, is the proprietary software platform for the GIS.

Documentary Practices Observed
There are no set business or documentary **procedures** within the organization—no procedural manuals have been written down for the creation, management or preservation of the Coalescent Communities GIS. "The documentary procedures mirror the business procedures." (FR 23) This situation is mostly, if not entirely due to the size of the business. "Due to its size, structure and organizational culture, the organization does not rely on procedures in a formalized sense—they are inferred." (FR 29)

Records Creation and Maintenance
The Center for Desert Archaeology has no existing formal **records management** program, mainly due to the fact that one individual is responsible for the creation and maintenance of the GIS database. "It is [the GIS Specialist's] idiosyncratic procedures that form the procedural context. These procedures are usually decided ad hoc and are not documented except through notes that are occasionally created or the transitory documents that are created during the course of creating records within the CCD-GIS. On occasion, the GIS Specialist will document certain steps within various analyses to speed the time it takes to replicate the analyses, especially for 'generic' analytic processes that are likely to be required for other analyses in the near future.

This reiterates the fact that this type of documentary process is ad hoc in nature, and is done for other reasons than the traditional need for procedures. With only one full-time employee dedicated to the creation and maintenance of the CCD-GIS, formally documenting every procedure is viewed by the GIS Specialist as both excessive and unrealistic in relation to other business needs." (FR 7-8) The final report adds that "If the database and resulting GIS grow larger in the future, then the record creator will create a procedures manual. Currently, there are only two people that use the data regularly, so a procedures manual does not seem necessary at this point." (FR 24)

Although there is no formal records management program, many of the records are kept, and the CDA acknowledges that "the central element of [the Coalescent Communities] project is information management." (FR 2)

There are specific **roles** regarding the creation and maintenance of records at the CDA, though it is uncertain if these roles are formalized by documentation or are merely accepted as common practice. The Coalescent Communities GIS is created under the authority of the GIS Preservation Specialist within the organization. This CDA preservation archaeologist who maintains the database is responsible for producing all of the outputs from the GIS, which include maps for publications, printouts of maps and tabular data for other researchers and analysis relevant to different research projects.

No **procedures** manuals have been created for the creation, management, or preservation of the Coalescent Communities GIS. The notion of a lack of human and financial resources as the reason for a lack of formal procedures is re-iterated later in the final report. "Two key things to note with respect to the...relative absence of formal and consistently utilized or applied elements, attributes and behaviours associated with the digital entities in question are that the procedural context for their creation is neither rigid nor always predetermined, and that the overall level of systematization of the GIS Specialists' activities is low, resulting in a process for creating and maintaining these entities that is most accurately characterized as both *idiosyncratic and ad hoc*. These procedural realities are largely due to the fact that only one individual (with limited assistance from a volunteer) is responsible for the creation and day-to-day maintenance of the CCD-GIS, coupled with the fact that *this activity takes place within the context of a relatively small, private, non-profit organization* with identifiable financial and human resource constraints." (FR 18, emphasis added)

There is, however, an entry form for the database and this control measure also imparts a notion of **authorship**. "Within the archaeological site entry form in the Access CCD, there is a special sign that takes the form of the CDA's organizational logo. The GIS Specialist views this special sign as one that denotes authorship and intellectual ownership of the unique collection of data brought together in the CCD.[2] The creator and author is the CDA, but the name of the originator is the GIS Specialist (who is also the writer)

---

[2] "It is important to emphasize that this attribution of authorship and intellectual ownership refers specifically to the totality of the CCD, as opposed to the individual data values or datasets within the CCD, since it is the individuals and institutions associated with the original creation/collection of the data who are still considered the authors and intellectual owners of those data even after they is imported into the CCD." (FR 17, footnote 7)

The final report states that "The records in question are part of the CDA's fonds." (FR 8) The "majority of the fonds" is **organized** according to special project and grants (series). Within these projects, the raw data is organized by 50-year time intervals. For example, all of the site location files are organized into folders by their corresponding time interval. "This temporal organization eases analysis in terms of patterns observed over time." (FR 21) In addition, the organization by project "mirror[s] the majority of the creation process, which is project- or problem-focused. Data are usually created in relation to a project and the filing schema is not significantly altered once the project has been completed." (FR 28)

The organization of entities is done using **archival principles**. The Coalescent Communities Project is a series within the Center for Desert Archaeology fonds. This series consists of the Coalescent Communities Database, the original researcher's data sets, the AZSITE data set, the BR data set, government geospatial data, NSF grant records, records documenting analysis, and administrative records. Digital entities are identified through file naming conventions.

Within the series, there are **aggregations** of files related to specific analyses or projects. "Aggregations usually form at the file level and relate to specific analyses or projects. The Coalescent Communities Database is also an aggregation itself at the series level." (FR 21)

Groups of data are captured temporarily within the GIS application (ArcView), where the user can create, manage and edit **metadata** based on standards. "The GIS Specialist is in the process of creating metadata relating to the source of the data, including the original author, date or recording, etc." (FR 28) The creator is interested in using ArcCatalogue, a metadata tool that is in the new version of ArcView. Within ArcCatalogue, the user could create, manage and edit metadata based on the Federal Geographic Data Committee Content Standards for Digital Geospatial Metadata or the ISO 19115 Metadata Standard. This metadata would be stored in XML.

Recordkeeping and Preservation
In the scientific community, "the long-term **preservation** of [complex datasets and outputs that are commonly associated with a GIS] has become a critical issue with regard to enabling the kind of multidisciplinary research crucial to modern scientific knowledge." (FR 1) However, other than burning data to CD-ROM, there are no preservation strategies being used by the creator in this case study. "The recordkeeping environment is dispersed and so some entities may go unaccounted for due to the nature of the system." (FR 27)

"There is no recordkeeping system external from the applications, so there is no formal **capture** activity." (FR 27) Nonetheless, "there is a basic **versioning** process that is used on the Coalescent Communities Database, but it is not entirely systematic. Once there is a large upload of corrections or newer data from researchers, the creator will save a version of the old database and start working with the newer database. There is no audit trail that tracks who enters what data into the database (there are only two people that perform this activity). In their own words, they are "too low-tech" for those means of tracking input." (FR 25)

Data are **stored** on a Local Area Network and are also sometimes saved on "the creator's personal computer...and/or copied onto CD-ROMs." (FR 26) However, most records remain

active or semi-active. While final versions are stored as described above, interim working copies of the GIS files typically are downloaded from the server onto the GIS technician's computer hard disk where he manipulates them before uploading the modified (updated) versions back onto the LAN server. The Center does not presently use any form or tertiary storage and not every record is kept.

The technological environment seems at a great risk for **obsolescence** or software **dependence**. "The GIS is a system that is not comprehensive in its management of information. The GIS consists of a database within Microsoft Access, image files, word processing files, flat files in Microsoft Excel, and files created in the GIS application software that are all *managed separately due to their separate computing environments*, file formats and formal differences in content, context and structure." (FR 22, emphasis added) In an attempt to deal with this issue, the CDA actively **migrates** electronic records to newer versions of the application software (i.e., Microsoft Office Suite software and ESRI ArcView) and hardware to avoid obsolescence. However, "this is the most that they are addressing in terms of software and hardware obsolescence" (FR 28) and this migration is strictly performed for reasons of usability, not for preservation of data or recordkeeping.

The creator is the intermediary between the files when **access** is needed, especially because the majority of them are in the file directory or on the hard drive of his computer. "Due to the sensitive nature of the data (archaeological site location), use is limited to competent persons that need to use the data for purposes relevant to the dataset. Currently, no one else has access to the actual application, only hard copies of the data…If someone in the organization wanted to see the dataset(s), they would be allowed to, but only the two authors end up viewing and manipulating the data. Access privileges are also ad hoc and, therefore, are decided on a case-by-case basis." (FR 25-26)

Accuracy, Authenticity and Reliability

Accuracy
Since the Coalescent Communities database was built with three existing data sets, "once the datasets had been placed into the Coalescent Communities database in Microsoft Access, the issue of accuracy, redundancy and patchiness needed to be addressed." (FR 22) There was, therefore, the question of introduced error in the Coalescent Communities database. "This introduced error consisted of gaps in data, duplicates, incorrect locational information and assemblages and analyses that were inconsistent with those of the Center." (FR 23) As a result, the pre-existing datasets were "routinely checked for data redundancies, errors and omissions by a volunteer who is a retired archaeology professor" (FR 2) to verify accuracy, fill in gaps in the data and re-purpose them for use in the Coalescent Communities GIS. "The volunteer also weeded out any information that seemed inaccurate. Once the database was acceptable, the analysis of the site information could begin." (FR 22)

Authenticity
Authenticity seems to be a notion that can only be partially met in the archaeological field and something that is *assumed* to exist, in large part. The creator in this study believes their data to be authentic "to an acceptable degree for the research that is being conducted. There is more

concern over the reliability and accuracy of the records than their authenticity. Authenticity is assumed to a reasonable degree since records are coming from a state repository or from datasets contributed by researchers who are trusted as professionals to maintain their data." (FR 25) This response is somewhat ironic, since authenticity is assumed based on the fact that professionals "maintain their data," yet the creator itself does not have a records management program.

Authenticity is difficult to verify, given the varied sources of the data used and the fact that once in the database, the source may be difficult or impossible to determine. "In general, it is difficult to deduce if the record is authentic in some sense of the word. Archaeological data that are used are usually coming from trusted sources where the records are held in a repository. Other times, the records are coming directly from a researcher's personal computer. There is usually a preliminary audit stage where the GIS Specialist checks the dataset as a whole to see if it is generally reliable and authentic." (FR 25)

Reliability

In their grant proposal for the Coalescent Communities project, the CDA expresses concern regarding the reliability of the data that they will use for analysis. "The central element of this project is information management. The current AZSITE program has very large amounts of legacy data that were collected long ago. Furthermore, the process of incorporating data from diverse institutions into a single database can be a source of introduced error." (FR 2)

Reliability of the database seems to be based on the reliability of the data that go into it. "The creator attempts to use the most reliable sources of information possible, such as published government spatial data and research data that are the most reliable in the field. The research data's reliability is based on professional authority and verification from experts in specific areas of archaeology." (FR 24)

Reliability seems to be an ideal that the creator knows will never be met, although a certain minimum degree must be achieved. "Archaeologists know that their data are not 100% reliable to fact, due to the nature of the archaeological record, so there is a degree of reliability that needs to be met before the data are considered "usable."" (FR 21)

The use of metadata-type information to document the data sources "is not based on any metadata standard." (FR 29) Nevertheless, current "in-house" strategy that is used for recording archaeological site source information "is seen to improve the reliability of the database as a trusted source of archaeological information." (FR 29)