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

Domain 3 Research Questions

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

Natalie Catto, UBC

June 2006

1. What types of entities does the diplomatic analysis identify in this case study? (i.e., records, publications, data, etc.)

The diplomatic analysis identifies the various versions of the Coalescent Communities (CC) Database to be records.

1a. If there are no records, should there be records? If not, why not?

Not applicable.

1b. If there should be records, what kinds of records should be created to satisfy the creator's needs (as defined by an archivist)?

Not applicable.

1c. What characteristics of records (as defined by an archivist) are missing yet necessary to preserve these entities?

Not applicable.

2. Are the entities reliable? If not, why not?

While procedures guiding the creation and maintenance of the database are ad hoc and involve little or no documentation, its reliability is ensured through the small amount of people involved in its creation. The GIS is created directly under the authority of the GIS Specialist. Only he can enter and manipulate datasets and obtain outputs, and only he and a volunteer have access to the actual application. The research data that populate the database are considered to be reliable because they are based on reliable sources of information such as government spatial data, and also because they are based on professional authority and verification from experts in specific areas of archaeology.

Their authority is guaranteed through legal regulations, professional codes of conduct, and standards of research performance guidelines.

3. Are the entities accurate? If not, why not?

The data held within the CC Database are considered to be as accurate as possible. Due to the nature of the archaeological record, data are not 100% accurate to actual fact, so the datasets imported into the CC Database must be probed as a whole to ascertain their overall reliability and authenticity; the data in the datasets are more-or-less "weeded" of any information that seems to be inaccurate or redundant. Spot checks are also conducted on the datasets to make sure they are being entered accurately.

4. To what degree can the entities be presumed to be authentic, and why?

Benchmark Requirements Supporting the Production of Authentic Copies of Electronic Records (these apply to the creator**)**:

The entities are presumed to be authentic by the creator to the degree that they can carry out the research activities for which they were created.

1. **Capture of identity and integrity metadata:**
   Files are indexed by their specific time period and project, and aggregations of files within certain folders can also create an associative identity of their own. Although the GIS Specialist is in the process of creating metadata relating to the source of the data (for example, the original author, date or recording, etc.), they are not yet in existence. In addition, a special sign in the form of the CDA logo exists within the site entry form of the CC Database, which confers authorship and intellectual ownership of the data.

2. **Enforcement of access privileges:**
   Within the organization, everyone has a right to view the data stored in the database, but only the GIS Specialist can enter and manipulate datasets and obtain outputs, and only he and the volunteer have access to the actual application. External researchers can also view the data, but only as a fixed source of information. Access is determined on an ad hoc, case-by-case basis.

3. P**rotection against loss and corruption:**
   Versions are determined on the basis of major changes or additions to the system, and all versions/backups are retained and cannot be changed. No audit log is in existence, as it is beyond the technological means of the CDA.

4. P**rotection against media and technology obsolescence:**
   Data are burned onto CD-ROMs, and software and data are actively migrated.

5. E**stablished documentary forms:**
   No established documentary forms are identified in the final report.

6. A**bility to authenticate records:**
No formal authentication of records is mentioned in the final report.

7. P**rocedures in place to identify the authoritative record:**
The creator considers the authoritative record to be the versions of the CC Database that have been set aside.

8. P**rocedures in place to properly document removal and transfer of records from the creator's originating system:**
Although versions of the CC Database have been set aside, they still remain within the originating system. The data within the database are organized according to project. Once a project is finished, the data related to that project are significantly altered.

Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records (these apply to the preserver):

Due to the lack of formal recordkeeping procedures, the following benchmark requirements cannot be answered.

1. Controls over Records Transfer, Maintenance, and Reproduction
2. Documentation of Reproduction Process and its Effects
3. Archival Description

5. For what purpose(s) are the entities to be preserved?

The entities are to be preserved to fulfill the CDA's mandate of promoting "the stewardship of archaeological and historic resources in the American Southwest and Mexican Northwest through active research, preservation, and public education."

6. Has the feasibility of preservation been explored?

The feasibility has been explored in terms of maintaining usability, but no long-term preservation methods have been considered due to the opinion that future GIS technology will render data, analyses and results obsolete.

6a. If yes, what elements and components need to be preserved?

Ultimately, the entire GIS system will need to be preserved, including its database, files, compilations of pre-recorded archaeological site data, and outputs.

7. Which preservation strategies might most usefully be applied, and what are their strengths and weaknesses, including costs and degree of technical difficulty?

Currently, files are actively migrated to newer versions of software (i.e., Preservation Strategy B1.4. Conversion). Data are also burned to CD-ROMS.

7a. Which alternative preservation strategies might be applied? What are their strengths and weaknesses, including costs and degree of technical difficulty?

The following alternative preservation strategies could be applied to the CC database. It is difficult to estimate the cost and degree of technical difficulty this project would entail:

B1. Use of standards
B1.2. Encapsulation
B2.3. Software re-engineering
B2.5. Emulation

8. What additional information does the preserver need to know to facilitate appraisal and preservation?

The preserver needs to obtain an understanding of how the entities were created and the archaeological and technological processes involved in their creation.

8a. If required information is missing, where should it come from and how should it be made manifest?

Creation and maintenance processes need to be fully documented by the GIS Specialist and his assistant.

9. Are there any policies in place that affect preservation?

No, there are currently no procedural policies in place.

9a. Are there any policies in place that present obstacles to preservation?

Not applicable.

9b. Are there any policies that would need to be put in place to facilitate appraisal and preservation?

A formal recordkeeping system needs to be put in place to ensure the accuracy, reliability and authenticity of the data within the system, and also to ensure that organized access to the entities is provided and that all entities within the system are accounted for. In addition, audit trails need to be implemented so that all changes to the entities are recorded.