InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **Walkthrough of the Chain of Preservation (COP) Model Using Case Study 14 Data, InterPARES 2 Workshop No. 10, Dun Laoghaire, Ireland, 23 June 2004**

Status: Draft (public)

Version: 1.2

Submission Date: September 2004

Release Date: September 2007

Author: The InterPARES 2 Project

Writer(s): Randy Preston
School of Library, Archival and Information Studies,
The University of British Columbia

Project Unit: Focus 2

URL: http://www.interpares.org/display_file.cfm?doc=
ip2_cs14_COP_model_walkthrough.pdf

Table of Contents

Version History:

Date	Version	Notes Regarding Modification	Changes by
7 Sept 2004	1_1	First Draft	Randy Preston
29 Sep 2004	1_2	First Draft – Minor Revisions	Randy Preston

1.0 Introduction

The purpose of this walkthrough was to assess whether the Manage Chain of Preservation (MCP) model accurately reflects the records management activities that are undertaken in a real-world situation and, if not, why not. One of the first challenges the Modeling Grouped faced was finding a case study advanced enough and with enough data with which to work. In the end, the decision was made to use Case Study 14 (Archaeological Records in a Geographical Information System), which at the time of the walkthrough was in its final stages, having already produced a draft report, an activity model and an on-line survey. As well, the case study contained over 130 pages of fully transcribed interview data.

1.1 Case Study 14 Background

Initiated in February of 2003, and directed by Richard Pearce-Moses of the State Archives of Arizona, this case study is examining the electronic record-keeping activities of archaeologists involved in research projects that utilize a Geographic Information System (GIS). Case study research data have been obtained from three primary sources:

1) An on-going GIS research project from the American Southwest;
2) A survey of GIS archaeologists worldwide; and
3) Existing archaeological and GIS research literature.

1.1.1 Coalescent Communities in Arizona Project (A.D. 1200-1540)

The case study is investigating the Coalescent Communities Database and GIS (hereinafter, CC GIS) created by the Center for Desert Archaeology (CDA) in Tucson, Arizona. The CDA is a private nonprofit organization whose primary mandate is to promote "the stewardship of archaeological and historic resources in the American Southwest and Mexican Northwest through active research, preservation, and public education." The CC GIS is a central component of the CDA's *Coalescent Communities in Arizona Project (A.D. 1200-1540)*, which is a multi-year, federally-funded project investigating the aggregation and migration of peoples in the prehistoric Southwest. The stated goal of the GIS component of this project is to assist the Center's staff and its partners in "integrating [the] growing archaeological knowledge of the region with recent computer advances in the analysis of attributes such as terrain, hydrology, and land-ownership."

The CC GIS consists of compilations of pre-recorded archaeological site data from multiple documents, repositories and researchers. These data sets came in multiple formats, predominated by digital spreadsheets and other pre-existing computer databases, with occasional paper records as well. All incorporated data sets routinely are checked for data redundancies, errors and omissions. The CC GIS was created by an archaeologist at the CDA whose primary function is the development and management of the CC GIS. This archaeologist is responsible for producing all of the outputs from the CC GIS, which include digital and hardcopy maps, graphs and statistical data for publications, digital and hardcopy maps along with their accompanying tabular data for other researchers and geospatial analyses relevant to different research projects.

The primary method of case study data collection from this source was intensive, semi-structured interviews with several individuals intimately involved in the planning, creation and maintenance of the Coalescent Communities Database and GIS, as well as other experts in relevant professional areas, such as GIS, Southwestern archaeology and information management of archaeological data sets.

Approximately one week before the interviews, the subjects were given a copy of the Peter Hirtle article, "Authenticity in a Digital Age" and a few mock scenarios that discussed issues of authenticity, reliability and accuracy in a contextualized and meaningful format for the subject. This allowed the interviewer to begin with a discussion of these issues without too much discrepancy of the meaning of the terms. This step also allowed the interviewer to witness the vocabulary that the subjects used when discussing issues of authenticity, reliability and accuracy. The interviews were recorded on analog tapes and later fully transcribed.

1.1.2 Archaeological GIS Record-keeping Survey Questionnaire

To help assess the representativeness of the record-keeping habits of the individuals interviewed within the context of the broader archaeological community, a survey questionnaire was developed and administered to GIS archaeologists worldwide via the internet in April 2004. To this end, the survey questions were written, in large part, to address specific records creation, management and preservation issues identified during the interviews. Personal invitations to participate in the survey questionnaire were emailed to approximately 900 GIS archaeologists from 69 countries worldwide. Additional invitations were posted to various archaeology- and GIS-related listserves and internet discussion groups.

This survey targeted archaeologists who currently use, or who have in the past used, a GIS in their archaeological research. Consisting of 40, primarily single or multiple choice, questions, the survey was designed to be completed in approximately 30 minutes. The survey was organized into the following seven sections: 1. Introduction (1 question), 2. GIS Experience/Background (7 questions), 3. File Management/Documentation Procedures (11 questions), 4. Digital Preservation Practices (13 questions), 5. Data Input/Output Practices (2 questions), 6. Record Quality, Reliability and Authenticity Issues (5 questions), and 7. General Comments (1 question)

Of the 186 log-ins recorded during the 32 days that the survey was available on-line, 157 completed (or sufficiently completed) surveys were received from archaeologists in 30 countries across 6 continents. A detailed written analysis of the results currently is ongoing and is scheduled for completion by early September 2004.

1.1.3 Annotated Bibliography

The creation of an annotated bibliography was another means of collecting data relating to the issues in the case study. Published resources were gathered relating to the following subject areas:

1. History and development of archaeological theory and methodology, especially in the American Southwest;
2. Information management as it relates to archaeology;
3. GIS theory and methodology, especially as applied in archaeological research;
4. Electronic records creation, use, management and preservation in archaeology; and
5. Case study methodology.

1.2 What is a GIS?

On a conceptual level, a GIS may be thought of as a coordinated, inter-related amalgamation of hardware, software, data, applications or procedures, and people (Figure 1). In more technical terms, a GIS is defined as a spatially referenced data set incorporating a database management system with a graphical display. In particular, a GIS dynamically links geospatial data and descriptive attribute data (i.e., tabular data) from a wide variety of sources such as maps, graphs, photographs, remotely sensed data, tabular data and text. Typically, data from each of these sources are stored in separate data stores and are manipulated via complex interactive queries. It is their capacity for storing, mathematically manipulating, and visually displaying spatially referenced data that distinguishes a GIS from Computer Aided Design (CAD) and Computer Aided Mapping (CAM) systems.

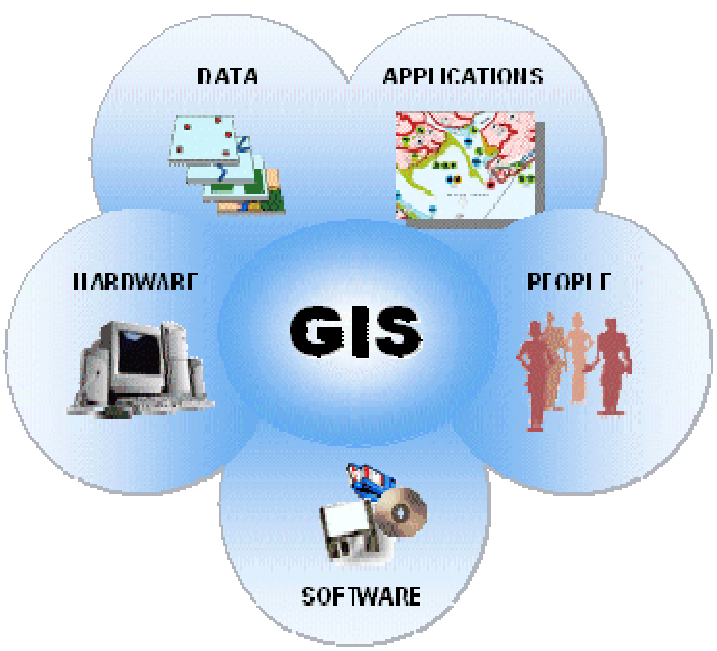


Figure 1. Components of a GIS.

The technological aspect of a GIS can be described by a model consisting of four interrelated subsystems, including:

1. A data entry subsystem that transfers both analog and digital data to a storage device;
2. A data storage and retrieval subsystem;
3. A data manipulation and analysis subsystem; and
4. A data visualization and reporting subsystem.

A data entry subsystem can employ a host of devices, such as digitizers, scanners, CD-ROMs, and global positioning devices (GPSs). A data storage and retrieval subsystem typically consists of a computer and any number of data storage devices such as hard drives, DVD drives, tape drives, etc. The data manipulation and analysis subsystem includes the software application(s) that enables a user to query the data, create new data, and perform statistical and spatial analyses. Finally, the data visualization and reporting subsystem, composed of any number of output devices such as high resolution graphic monitors, plotters and printers, facilitates graphical display and provides print results of a spatial analysis or query.

Data may be incorporated into a GIS in numerous ways. One technique is to manually enter point co-ordinates of the area or feature of interest, collected using either traditional field methods or with the aid of GPS (Global Positioning System) receivers. Another option is to convert pre-existing databases that already have spatial data embedded in them. Existing paper map data are perhaps the largest source of raw GIS data, and are converted into digital data through a process of scanning, digitizing or a combination of the two. As noted above, data incorporated into the CC GIS have largely come from pre-existing computer data sets

2.0 MCP Model Walkthrough

2.1 Manage Chain of Preservation (A-0)

Several fundamental practical issues related to the task of comparing the case study's research activity to the MCP model were identified at the outset of the walkthrough. First, there was the issue of how to reconcile the fact that the case study examined the creation, implementation and maintenance of a research database and GIS, and the records management activities associated with these processes, rather than a records management system *per se*, as is presumed in the MCP. In fact, the MCP model presupposes the existence of a comprehensive records management framework comprised of a number of inter-related systems, including record-making, record-keeping and record-preservation systems, none of which are associated to any developed degree or indeed in any formalized sense with the research activities investigated by the case study. Further compounding this disjunction is the fact that the record-keeping activities which the case study did investigate were those related specifically to one focused research activity within a larger organization, rather than the those of the organization as a whole. In other words, the case study had examined the isolated record-keeping activities of an isolated research activity within a larger research and administrative entity.

Second, there was the difficulty of identifying which entities, if any, associated with the GIS research activities constituted *records* in the sense presupposed by the MCP model. For example, despite the fluid (i.e., non-fixed) nature of the current, active GIS database and its underlying data sets, it was suggested that these entities might be construed as records inasmuch as they can be said to provide evidence of activity within and about the research project. It also was suggested that it may be useful to view the entire database as a *dynamic* record with the capacity to create other records (e.g., maps, graphs, statistical data, etc.).

Finally, there was some initial confusion regarding the purpose and utility of the case study activity model (Create CC GIS) in relation to the walkthrough. This was due, in part, to an emailing error which resulted in the inadvertent omission of the case study interview transcriptions and the draft report from the suite of case study documents intended for circulation to all members of the modeling group prior to the walkthrough. Without the benefit of the interview transcriptions and the draft report, the proper function of the activity model within the context of the case study and, especially, its relation to the model walkthrough, was unclear. Contrary to what is indicated in the modeling cross-domain meeting notes for the walkthrough, the case study activity model was *not* produced for, nor in anticipation of, the walkthrough. In fact, the first draft of the model was created in mid-February (nearly 4 months prior to the invitation extended to CS 14 to participate in the walkthrough) primarily as a means to help the case study researchers better visualize and analyze the GIS database and record-keeping activities contained in the (then) newly transcribed interview data. Thus, the activity model was just one analytical component of the case study and was never intended to serve as a template for assisting with the MCP model walkthrough.

Despite these challenges, it was decided that sufficient conceptual parallels could be drawn between the records management systems and procedures associated with the creation, use, maintenance and preservation of the CC Database and GIS and the records management systems and procedures presumed by the MCP model to warrant conducting the walkthrough. At the very least, it was suggested that the walkthrough should help identify questions directly pertinent to assessing the MCP model that, in the existing absence of a set of specific walkthrough protocol questions, were not being addressed during case study interviews. It is hoped, therefore, that the findings of the walkthrough will help facilitate the ongoing effort to devise such a protocol, which, in turn, will help lead to the establishment of a minimum set of data collection requirements for future case studies to ensure that they capture the information necessary to adequately test the MCP model in future walkthroughs.

2.2 Manage Framework for Chain of Preservation (A1)

The focus in A1 is on the design, implementation and maintenance of a chain of records preservation based on an analysis of the creator's overall records management circumstances. The outputs of A1 include the record-keeping, record-making and permanent preservation systems. Each of these outputs, together with the four A1 subactivities – Determine Framework Requirements, Design Framework, Implement Framework, Maintain Framework – were found to have rudimentary parallels within the research activities examined in the case study, as noted below.

2.2.1 Determine Framework Requirements (A1.1)

The goal here is to assess the records creator, its existing records and the available information about those records to enable identification of the requirements for the chain of preservation framework. The case study parallel involves the process of determining the GIS framework requirements. Although the actual requirements appear never to have formally been recorded, this should not be interpreted to imply that the GIS was created without any forethought as to its requirements. In fact, the GIS creator undoubtedly was sensitive to the various aspects of the technological and functional requirements, including, in particular, their potential impact on the ability of the GIS framework to adequately address the research questions. As is noted in the interview data, the creator's GIS expertise is largely self-acquired and supplemented with little or no formal training. At present, this appears to be the norm within the archaeological profession. Thus, it is not uncommon for these self-taught GIS specialists to rely on largely idiosyncratic, if not ad hoc, GIS creation procedures that often never are documented. Integral to the conceptualization of the GIS framework requirements is the development of archaeological models – consisting of sets of hypotheses put forward to help simplify and explain the complex patterning of archaeological data – as these, in many ways, determine the nature of many of the most basic elements (such as the types and sources of data, for example) that must be accommodated by the GIS to enable it to successfully address the research questions. While it is likely that some discussion of the Center's GIS framework requirements was included in the project's National Science Foundation (NSF) grant proposal, case study researchers did not have access to the proposal to confirm this.

Given the paucity of specific information about the actual procedures involved in the development of the GIS framework requirements, A1.1 of the MCP model was not broken down further during the walkthrough.

2.2.2 Design Framework (A1.2)

On the basis of the framework requirements identified in A.1, the goal here is to design a record-making system, a record-keeping system and a permanent preservation system. While there is no direct correlate in the records management activities of the creator in the case study to this component of the MCP model, it was suggested that the "Build GIS Framework" component from the case study activity model (A0 in Appendix A) provides a close parallel. Building the GIS framework involves integrating the five elements of a GIS (see Figure 1) into an overall GIS framework that best accommodates the archaeological models and the research questions. For this case study, these five elements consist of:

1. Hardware: The hardware system for the GIS is built using technology commonly used in offices, including a local area network (LAN) central file server, personal computers, printers and a flatbed scanner. The system includes no custom hardware. The project's GIS specialist has occasional need for both a plotter and a digitizing tablet (the latter to convert analogue map data into digital form), but has not direct access to either and must instead rely equipment at another facility by transferring files back and forth on CD.

2. Software: The Center's operating system software is typical of what is found in many offices. At the time of the interviews, they were using Microsoft Windows 2000, version 5.0. The primary application software used includes: ESRI ArcView, versions 3.2 and 8.2 (GIS software) -- including a number of unspecified analytical (primarily statistical) add-ins downloaded from the ESRI website, EndNote (bibliographic software), Adobe Acrobat, Topo! (topographic map software), Notepad (basic text editor software), Microsoft Office Suite -- including in particular Access (database), Excel (spreadsheet) and Word (word processor), and unspecified basic Microsoft image editing/analysis software (the default image editor that comes bundled with Windows 2000).

3. Data: As noted earlier, the CC GIS consists of compilations of pre-recorded archaeological site data from multiple documents, repositories and researchers. These data sets came in multiple formats, predominated by digital spreadsheets and other pre-existing computer databases, with occasional data from paper records that, depending on the exact nature of the data source (e.g., map vs. tabular data), either were entered manually into a computer or else digitized using a digital tablet. In addition to the more familiar file data formats associated with MS Access (.mdb), MS Excel (.xls), MS Word (.doc), Notepad (.txt and .rtf), Adobe Acrobat

(.pdf) and EndNote (.lib), ArchView generates five proprietary file data formats, including: .shp, .shx, .dbf, .sbn, .sbx and .apr. Image files are saved in JPG format.

The CC Database obtained much of its archaeological data from a large GIS database known as AZSITE, which is a compilation of all Arizona state-wide archaeological site excavation and survey data held in federal, state and tribal repositories. The AZSITE data imported into the CC Database were found to contain numerous errors that required “cleaning up.” Once the CC Database is more finalized, the Center for Desert Archaeology will share the corrected version of its database with AZSITE.

4. Procedures: The procedures used to create the GIS framework design (as well as those used in subsequent data manipulations) are difficult to characterise precisely because they are not clearly defined or delineated by the creator. In fact, as was the case with the framework requirements, the actual procedures used to create the GIS framework design appear never to have formally been recorded.

5. Persons: The original design and subsequent manipulation and maintenance of the GIS framework has been the sole authorised responsibility of the same full-time archaeology GIS specialist throughout the duration of the project.

While there are no formal, centralised or dedicated record-making, record keeping and/or permanent preservation systems associated with the activities of the records creator in the case study, at least not to the degree of integration and comprehensiveness outlined in the MCP, there are nevertheless definitely elements of such systems. In fact, taken together, the activities associated with the above noted five elements of the GIS framework result in many outputs that are at least analogous to, and at time closely congruent with, those associated with the MCP A1.2 outputs.

2.2.2.1 Design Record-making System (A1.2.1)

The four main steps involved in designing the record-making system, as outlined in the MCP model, include: 1) developing record forms and metadata schema, 2) establishing competencies for record-making, 3) designing integrated business and documentary procedures, and 4) selecting appropriate record-making technologies. The framework requirements identified in A1.2 help determine the exact nature of the particular activities and outputs associated with each of these steps. In comparison to the system design process outlined in the MCP model, design of the CC GIS was more ad hoc in execution and less comprehensive in scope with respect to its record-making capabilities. In fact, it is important to emphasise that no part of the CC GIS intentionally was designed to function as a record-making system in the sense that is assumed by the MCP model. Nevertheless, CC GIS does possess many of the same record-making system elements and outputs (or close parallel surrogates) outlined in the MCP model and is in fact capable of creating records.

2.2.2.1.1 Develop Record Forms and Metadata Schema (A1.2.1.1)

This step involves establishing the documentary forms of records and the required attributes (metadata schema) to be inextricably linked to them in accordance with InterPARES 1 benchmark requirements 1 (Expression of Record Attributes and Linkage to Record) and 5 (Establishment of Documentary Forms). In the context of the case study, the identification of formal or standardised record forms largely depends on which elements, inputs and outputs of the GIS are taken to constitute records. If the entire database is presumed to constitute a record, then one may identify the various data entry, query and report (i.e., output) templates within the Access database as legitimate record forms. As well, although there are no deliberate, creator-imposed metadata *schema* associated with either the CC Database or the GIS, there are metadata associated with both. Most of these metadata are generated automatically by the various computer programmes and the operating system with little, if any, conscious input from the creator (e.g., date of file creation, last modification, etc.). On the other hand, certain metadata are consciously determined by the creator, such as any user-defined constraints associated with the fields within the CC Database, for example. It is noted also that at the time of the interview, the GIS specialist seriously was considering creating a proprietary metadata schema to document data source information, such as author, date of recording, etc., for use with ArcCatalogue, a built-in metadata tool in the new version (i.e., 8.2) of ArcView. ArcCatalogue enables a user to create, manage and edit metadata based on the Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata, or the ISO 19115 Metadata Standard. ArcCatalogue stores metadata in XML format.[1]

If the individual datasets imported into the CC Database, or those exported into Excel, manipulated and then imported into ArcView, are seen as records, then it may be possible to speak of the specific data formats (as embodied in their file types) as record forms. It also was suggested that the *views* generated by ArcView may constitute analogous entities to records forms. An ArcView file incorporates five distinct components: 1) views, 2) tables, 3) charts, 4) layouts and 5) scripts. Together, information about these five components is stored in one index file called a *project* (with an .APR extension). At any one time there may be multiple project files residing within the ArcView file system, each representing different analyses of the same or a different set of data. Each project file tells ArcView exactly where within the computer file system the information (maps, data files, images, etc.) used within that project can be found, but does not itself contain that information. When a project file is opened, it retries this disparate information from each of five proprietary data files (with .SHP, .SHX, .DBF, .SBN and .SBX extensions) and opens each of the five components in separate windows, all of which are then co-ordinated by the project window. The project file manages information on the display status, screen location, size, etc., of its various component windows. The ArcView interface includes individual menu, button and tool bars for each component, all of which can be customised by the user. Information about the ArcView interface is stored in the 'default.apr' file.

A *view* essentially is an interactive map that allows the user to display, explore, query and analyse geospatial data. Views can be said to serve a function analogous to record forms in as much as it is the views which *define* the geospatial data that are used and *how they are to be*

[1] ESRI. (2003). Spatial Data Standards and GIS Interoperability, 3.

displayed. Moreover, it is important to note that the views do not actually contain the geospatial data files themselves, just the references to those data files. The referenced data within a view are organised according to *themes*, which may be defined as distinct sets of geographic features within particular geospatial data sources. It is common to think of each theme as a discrete overlay or layer which references distinct geospatial information.[2] Thus, for example, a view might have one theme representing archaeological sites, one theme representing water sources, one theme representing soil types, etc. With respect to the CC GIS, the archaeological themes are constructed from the data sets exported from the CC Database, while geographic boundary, topographic, hydrologic and related themes are constructed from data sets imported into ArcView from government DEM files. Prior to their import into ArcView, all data sets must be transformed into ARC/INFO format, as a raster image, or else created as ArcView shape files from X-Y co-ordinates. The actual themes used in a view are itemised in a table of contents associated with that view. ArcView allows the user to use both pre-defined (i.e., defined by the software manufacturer) and custom views (i.e., defined by the end-user). More importantly, some or all of the contents of a view can be locked so that they cannot be modified by subsequent users. ArcView also supports the creation of custom functions, user interfaces and applications based on views through the use of scripts.[3] ArcView, version 3.2, uses a proprietary object-oriented programming language known as *Avenue*, while version 8.2 uses *VisualBasic*. However, the GIS specialist noted that, thus far for the CC GIS project he has rarely written and implemented any customised code for use within ArcView, primarily due to his inexperience with Avenue (at the time of the interview, the project was still using ArcView 3.2). The GIS specialist is more familiar with VisualBasic, and hopes to develop customised functions, especially to help automate some of the more repetitive map-making processes, once the migration to ArcView version 8.2 is completed.

Another component of ArcView that may be analogous to a record form entity is the *layout*. In ArcView, a *layout* refers to a framework for preparing graphical output, such as maps, charts, and table records, for hardcopy printing or for saving in digital format for exporting to word processors, publishing applications, etc. The layout defines which data are used to generate an output and how those data are displayed. In addition to views, charts and tables generated by ArcView, layouts can include elements such as graphical files (photographs, images, etc.), text, frames, directional arrows, scale bars, legend boxes, etc. As with views, ArcView allows the user to use both pre-defined and custom layouts in which some or all the individual elements can be locked to prevent further modifications, or unlocked to enable the elements to dynamically be updated whenever the data sets to which they are linked are modified.

Unlike traditional paper record forms, ArcView views and layouts are both interactive and dynamic. They are interactive in the sense that the user has control over which themes and elements to include/exclude, how to display them (colour, symbol, text, scale, display resolution, etc.), and the order in which to draw (render) them. They are dynamic in that they can be designed to reflect the current status of the referenced source data such that any changes to the referenced source data will immediately be reflected in the view/layout.

[2] In fact, the data sets exported from the CC Database, manipulated in Excel and then imported into ArcView to be included in a view are included as themes, and must be prepared in ARC/INFO format, as a raster image, or created as ArcView shapefiles from X-Y co-ordinates.
[3] ArcView, version 3, uses a proprietary object-oriented programming language known as *Avenue*, while version 8 uses *VisualBasic*.

2.2.2.1.2 Establish Competences for Record-making (A1.2.1.2)

This step involves defining the access privileges for the creation, modification, annotation, relocation and destruction of records in accordance with InterPARES 1 benchmark requirement 2 (Access Privileges). All access privileges to the CC Database are instituted and controlled by the GIS specialist through the use of password protections. At present, only two individuals – the GIS specialist and the volunteer – have full access privileges. Like most aspects of the design of the case study's record-making system, it appears that delineation of the record-making competences has been largely ad hoc and has not, for example, formally been documented or captured in a procedures manual.

With respect to procedures governing access to the CC Database and other GIS files, the only procedure currently in place is a basic password access system that provides the GIS specialist and the volunteer unrestricted access to the CC Database. There are no other security or access procedures in place, including any form of monitoring or auditing system to track who accesses, modifies or deletes any of the project's files. The GIS specialist is aware of the necessity and importance of defining and establishing access privileges, and has been involved in the implementation of relatively complex and sophisticated access control systems with past projects. However, given the relatively small scale of the current project and the limited number of people who work directly with the CC Database, the GIS specialist does not feel the need at this time to incorporate more stringent or elaborate access and security monitoring procedures.

2.2.2.1.3 Design Integrated Business and Documentary Procedures (A1.2.1.3)

The goal here is to develop the procedures necessary for documenting the creator's record-making activities and the resulting records, and linking the documentation to those records. This step in the design of the record-making system is very poorly represented in the case study. With just one full-time employee dedicated to creating and maintaining the CC Database and GIS, coupled with the fact that this is a project of a small, non-governmental organization, it is not surprising that the procedural context for records creation and documentation is neither robust, systematic nor predetermined.

2.2.2.1.4 Select Technology (A1.2.1.4)

This step in the design of the record-making system involves assessing the technology currently available on the market, and choosing and reporting on the appropriate technology for the system. As noted above in the A1.1 discussion, it is presumed that the GIS specialist, despite being largely self-taught, possesses the necessary technological knowledge and expertise required to make informed choices with respect to the GIS-related technologies that are the most appropriate within the context of the project's framework requirements. However, given the absence within the project of any formal records management technology, apart from that which *ipso facto* is associated with the various operating system and GIS-related software applications, as well as the near absence of any formal records management procedures, a similar presumption does not seem warranted with respect to the GIS specialist's knowledge of, or experience with, dedicated records management technologies.

2.2.2.2 Design Recordkeeping System (A1.2.2)

The five main steps involved in designing the recordkeeping system, as outlined in the MCP model, include: 1) developing a classification scheme, 2) developing a retention schedule, 3) establishing procedures to meet InterPARES 1 benchmark requirements, 4) designing a retrieval system, and 5) selecting the appropriate recordkeeping technology. The framework requirements identified in A1.2, together with the outputs identified in A1.2.1 (acting as constraints), help determine the exact nature of the particular activities and outputs associated with each of these steps.

Aside from the basic, proprietary recordkeeping functions associated with the various operating system and GIS-related software applications, there is no separate dedicated recordkeeping system associated with the CC GIS project. While the GIS specialist does intentionally save certain project files, the recordkeeping "system" used is ad hoc in its formation, implementation and maintenance. There is, for example, no formal classification scheme or retention schedule. The closest parallels in the case study to the recordkeeping functionality intended by these two system components are found in the GIS specialist's idiosyncratic approach to naming, organising and "scheduling" project files (see below), and in the file extensions that automatically are created within the various applications. Consequently, further analysis of the individual sub-components of A1.2.2 was not conducted during the walkthrough.

2.2.2.3 Design Permanent Preservation System (A1.2.3)

As outlined in the MCP model, design of the permanent preservation system involves the integration of three distinct systems, including: 1) a records selection system, 2) an archival preservation system, and 3) an archival descriptive system. Aside from the undocumented, ad hoc backup and data migration procedures that are used to save and/or update periodic versions of the CC Database and related GIS files (see below), there is very little in the way of any component of the case study that could be construed as a formal permanent preservation system as defined in the MCP model. Consequently, this component of the MCP model was not investigated further during the walkthrough.

2.2.3 Implement Framework (A1.3)

The procedures involved in this step include acquiring, testing and activating all the components of the record-making, recordkeeping and permanent preservation system outlined in A1.2. Once again, while there is no direct correlate in the records management activities of the creator in the case study to this component of the MCP model, it was suggested that the "Implement CC GIS" component from the case study activity model (see A0 in Appendix A), in conjunction with the "Build GIS Framework" component noted earlier, provides a close parallel. In fact, this component involves all of the same basic activities – software/hardware acquisition, testing and activation – that are associated with A1.3 of the MCP model; albeit not for the express purpose of implementing a record-making, recordkeeping and permanent preservation system, however.

2.2.4 Maintain Framework (A1.4)

This activities associated with this component of the MCP model involve assessing information about the record-making, recordkeeping and permanent preservation system and making recommendations on the revision of the system's framework design. These activities are roughly parallel to the various revision activities associated with A3.3 (Revise CC GIS) from the case study activity model (see Appendix A). As defined in the activity model, the goals of these revision activities are:

1) to evaluate the current analyses and data (base, derived and interpreted) and make any necessary corrections, additions or completely redo certain portions of the analysis;
2) to rid the CC Database of repetitious and inaccurate data that are identified during the analysis of the database;
3) to monitor the changes made to the formative parts of the GIS; and
4) to identify any changes that might need to be made to the formative parts of the GIS.

In general, this process involves scrutinising various analysis results and/or outputs, such as maps and spatial and tabular data, as well as evaluating the CC Database itself, to ensure that the GIS is functioning as expected and is producing accurate, sensible and useful results with respect to the data being used and the research questions being asked. With the exception of the data auditing activities carried out by the volunteer, all of these activities are performed by the GIS specialist. Aside from any discussion of preliminary research results that may be included in interim reports (e.g., performance reports to granting agencies), it appears that few, if any, of these activities formally are documented.

2.3 Manage Records Creation (A2)

This component of the MCP model refers specifically to activities designed to control the making and receipt of records, and the storage of records in a dedicated recordkeeping system. As has already been discussed, there is no formal, centralised or dedicated recordkeeping system associated with the case study, so there is no formal records capture activity. Nevertheless, there are various record-making and recordkeeping activities that are congruent with, or at least parallel to, certain record-making and recordkeeping activities identified in the MCP model. In fact, there are numerous dispersed, application-driven, capture activities within the CC GIS. Data within the CC Database are captured within the Microsoft Access database application, for example. However, aside from other applications within the Microsoft Office Suite, there are no collective capture tools or procedures for the information generated within the CC GIS. Instead, as was discussed above for A1.2.1.1, groups of data are captured temporarily and organised or stored in what are called "projects" within the GIS application, ArcView, while geospatial analyses of those data are being conducted. Once these analyses are completed, however, the data and results are then exported (or output) to their appropriate areas outside of the GIS application, or, in some cases, simply deleted.

To date, no procedures manuals have been developed to address the creation and management of the CC Database and other GIS components and outputs. In fact, most creation and maintenance procedures are entirely undocumented. Even the basic file naming process that the GIS specialist uses to keep track of file versions, although somewhat systematic at times, is idiosyncratic and entirely undocumented. On the other hand, some degree of incidental documentation does occur during the course of creating certain, often transitory, records within the CC GIS. For example, the GIS specialist will document certain steps within various likely-to-be-repeated analyses, primarily to speed the time it takes to replicate them.

> I don't have a [procedures] manual, but when I do a complicated process, like some of these that I have to make a whole bunch of different files and do a whole bunch of manipulations to get to an end result, I usually try to write that down, just so that I can remind myself more than anybody else because I'm gonna have to do it again in a couple of months. And I'll spend a day or two trying to figure out how do something sometimes and if I don't and I forget it, then I have to do it again.

This example underscores the largely ad hoc nature of the documentary process and emphasizes the fact that it is conducted for other reasons than the traditional need for procedures.

There is no formal, systematic and documented records classification scheme. Instead, the GIS specialist employs an idiosyncratic, undocumented and somewhat ad hoc computer filing system for organising the individual files, along with equally idiosyncratic, undocumented and often ad hoc strategies for naming the files. The files typically are organised into folders by "project" broken down into 50-year time periods (e.g., 1250-1300AD, 1300-1350AD, etc.), a format which does mirror the general Coalescent Communities research process. This filing "scheme" is not significantly altered once a project has been completed and typically is preserved intact in those instances where all the files of a project are backed up or transferred into "long-term" storage (i.e., burned to CD-ROM or transferred to the CDA's LAN file server).

At least three key factors have been identified which influence the nature of the overall records creation and management processes. First, as has already been noted, there are administrative constraints related to available financial and personnel resources.

Second, there are technological considerations and limitations, one of the most significant of which appears to be the relationship of file size to available file storage capacity and data processing speed. As the following excerpt from the interview transcript demonstrates, the time required simply to transfer very large files influences the GIS specialist's work habits:

> I work on my station and, in fact, even with a lot of the GIS stuff. Even though I keep it backed up on the network, on the server, I actually prefer to copy the files onto my computer. I keep [the] working files on my computer just because the GIS stuff is a lot of times pretty time intensive, pretty labour intensive stuff that just makes it work that much faster if I keep it on my computer. So I'll do everything I need to do on my system …some of these things like these terrain models, for example, for the entire Southwest it's hundreds of megabytes, and I can't move that back and forth on the network very easily every time I want to do something with it.

The third key factor influencing the nature of the overall records creation process is training. When asked about the potential for procedural differences in the way he creates and manages GIS files relative to what other GIS archaeologists do, the GIS specialist cited lack of formal GIS training as one of the underlying reasons for his idiosyncratic, often ad hoc, records creation procedures. In particular, he suggested,

> …there probably is a difference between the way I do it and the way other people do it, but I don't really know what that is and I think that…I kind of learned how to do a lot of this stuff. I was sort of self-taught but with the guidance from a couple of – there were two or three kind of important people who helped me figure a few things out and then I patched together the rest myself, so I never really came out of a GIS program where there was a lot of formal instruction, where people might be more inclined to do things the same way. And I've always kind of worked on my own. I've always been working on my own research project or I've been on a project where I was "the GIS guy" and so there wasn't really anybody else to talk to about it. So I don't really know how other people do things quite as much as I should probably, or if I'd come from a program that had a whole cohort of GIS people working together.

Given the largely idiosyncratic, ad hoc and undocumented nature of the GIS specialist's records creation and management activities, further analysis of the individual sub-components of A2 was not conducted during the walkthrough.

2.4 Manage Records in a Recordkeeping System (A3)

Within the context of the MCP model, this component refers to the activities required to maintain, make accessible and carry out the disposition of records. The absence of a formal, dedicated recordkeeping system designed to manage the CC GIS records limited the degree to which case study data could be applied to this component of the MCP model. Nevertheless, various aspects of the GIS specialist's recordkeeping activities and procedures do parallel (and at times, in fact, closely mirror) certain of the procedures outlined in the MCP model enough to warrant further discussion.

Overall responsibility for, and management of, the CC Database and GIS records rests solely with the GIS specialist. However, to a varying degree, both the GIS specialist and the volunteer carry out record maintenance activities with respect to the CC Database. In fact, it is the volunteer who typically makes the "small day-to-day changes"[4] to the CC Database as a result of his data auditing activities, while the GIS specialist periodically performs all large-scale changes, such as incorporating new data sets. The GIS specialist also currently performs all file deletion, relocation and database back-up activities. The volunteer, despite lacking "official" authority (i.e., competency) to do so, nevertheless does possess the necessary access privileges, in the form of full, unrestricted access to the CC Database, to potentially perform most or all of these same recordkeeping activities. As noted earlier, there is no formal monitoring or auditing system in place to track who accesses, deletes, moves or otherwise modifies any of the project's files.

[4] Changing incorrect archaeological site numbers, deleting duplicate data, etc.

Given its dispersed nature, the recordkeeping "system" does not provide ready, systematic access to all relevant digital entities and, when available, related metadata. Instead, the GIS specialist must act as intermediary whenever access to these entities are needed. As noted earlier, full, unrestricted access to the CC Database is limited to the GIS specialist and the volunteer, and is controlled by the GIS specialist using a basic password access system. Requests to access information in the database by individuals other than the GIS specialist or the volunteer are assessed, by the GIS specialist, on a case-by-case basis, taking into consideration such factors as the nature of the request, the credentials of the requester and legal restrictions on the dissemination of archaeological resource location data. To date, successful requesters are only provided with a locked (i.e., read-only) version of the database.

There are no formal procedures or policies in place governing disposition of the CC GIS records. All disposition-like activities that do occur, such as deleting or naming/renaming files, are conducted by the GIS specialist following an idiosyncratic and largely ad hoc approach. For example, with respect to his procedures for naming and renaming files, the GIS specialist notes that,

> [They are] not very standardized and to tell you the truth, no one really messes with my files very much, I guess, that they need to know what my [filename] code is. When I get into a situation, I guess I always try to use names that are sort of meaningful. I don't write code [names], I just try to make it as obvious as I can and especially if I'm writing a name for a file that I'm going to send to someone or that I'm going to put in a public place. We have a sort public GIS directory here, so if someone else is working on a project they can go in and get a land ownership coverage. So I'll try to name them in some sort of obvious way, but it's not systematic.
>
> I sometimes change [a file's name] because a lot of what I do is kind of exploratory and I'm not sure what it's going to turn into and I have my own pet name for files like test or junk or something like that. And if it turns out to be something that I think is useful, or I want to send it to someone, then I'll rename it, but my temporary folder is full of test1, test2, test3, stuff like that.

Similarly, with respect to file deletion and backup activities, the following excerpts from the interview transcripts provide a good sense of the ad hoc, idiosyncratic nature of the GIS specialist's approaches and some of the records management problems that these approaches have created:

> I start off with a file, site locations, and I have a file of digital elevation. Then in order to figure out how far people would walk, I have to take that elevation and turn it into slope. And then I have to take that slope and I have to get the sine of the slope. And I have to get the cosine of the slope, and before I can get the sine and cosine of the slope, I have to turn the slope into radions. Then once I get that I have to multiply the slope times something else and multiply the cosine times the sine and I end up with all these files and everyone of them is enormous and it

would be completely prohibitive to try to keep. So in a way, those would be the [documentation] of the process of me getting at the final product which is a map where people would have walked in the past. But I delete most of them.

…when you start off with 2 files and you want to get clear down here to point B, you've got 50 files in between. I sort of make subjective decisions along the way, 'well that one might be useful sometime, so I'll save it.' But it's only useful to me and it probably wouldn't be worth me going to the trouble to document that and make it available to somebody else at a future date because I'm not sure they would have the same sense of its value or they wouldn't be wanting to get to point B or something. So a lot of it – it would be hard to reconstruct all of it. I try to keep the basic stuff, the site files – basic things like the topographic data, the land ownership data, the vegetation data – I try to keep those things kind of separate, in their own [folders] sort of. Well, I even kind of refer to these things as archives. If I ever need to go back to the data, that's where it is. All this other stuff [i.e., the numerous intermediary files created during the course of transforming a dataset from the CC Database, via various algorithms, prior to import into ArcView], these are kind of my working files and if I [leave the project], no one else is going to be working with them.

I've created so much stuff…and even though I kind of try to clean up after [each analysis], I don't know how to describe it, but it's starting to get kind of confusing to maintain, and I'm getting to a point now where [the volunteer] is sending me a whole bunch of corrections and so I want to start using the new data but I don't want to get rid of the old data just yet. I just start a new folder, so I have "GIS files" and now I have a "GIS gen2", so now everything I do I'm starting to put it in generation 2 and eventually, "GIS files" will become obsolete and I'll probably delete it.

File management is a big part of GIS, you just end up with so much junk. I just try to clean up after myself as I go. And a lot of times when I'm just working on things, if I'm just doing dummy stuff, if I'm just making things to see what they look like, or I'm just doing something that is a relatively unimportant thing, I will save those files in a temp file that I'll just clean out periodically. So that way, if I need to back things up, I periodically take things off my computer and put them on a server. I just take the whole GIS folder and transfer it [to CD-ROM or the CDA's LAN file server].

We haven't developed a systematic process for [versioning when backing up files], but for example, we'll make versions and call them underscore 1 or if I archive something like the Coalescent Communities database, every now and then, _______ will send me 500 corrections to it and so I'll take the one that I've been working on and archive it and take a copy of it and make the changes in it so when I archive it, I'll call it "Coalescent Communities 171703."

With respect to macro level records management, the various CC GIS files tend to be organised according to the individual projects that utilize the CC GIS for purposes of archaeological research. When assessing the creator's recordkeeping system at a micro level it is important to remember that the CC GIS is a problem- or project-specific GIS that answers temporally-defined research questions. Thus, within these projects, the raw datasets, intermediary files, analysis results and outputs typically are aggregated into individual folders distinguished by temporal divisions (1250-1300AD, 1300-1350AD, etc.). In the words of the GIS specialist:

> I try to organize things by project. And then within project – within the Coalescent Communities project, I'll have subfolders. For example, with the GIS stuff I'll have several different files for each time period, for example, the interval from 1250 to 1300 is in a different file than 1300 to 1350. And so it's much easier, instead of coming up with creative names for every single file. Instead of calling the file 1300, 1350, blah, blah, blah – I'll just call them all the same thing – they'll all be 'population contours' or they'll all be 'site location' or they'll be 'cost surface' or something like that. And the only way they'll be distinguished is that they're in the folder that's named 1250 or 1300.

It is worth emphasising that the function of this records organisation system is to help facilitate analysis in terms of patterns observed over time, rather than to facilitate recordkeeping, *per se*.

As these largely ad hoc management procedures clearly demonstrate, the recordkeeping "system" currently in place does not, and cannot, routinely capture all the digital entities within the scope of the activity it covers. Instead, there are voluminous intermediary files that are created during the course of the project that are discarded once the immediate activity for which they were created is completed or the research question they are helping to address has fully been answered. The reality is that numerous digital entities go unaccounted for due, in part, to the manner in which they are created, in part to the purpose for which they are created, and in part to limitations inherent to this type of ad hoc, idiosyncratic and dispersed recordkeeping "system." Related to the latter is the issue of technological practicality discussed earlier, in which the creator's recordkeeping activities are influenced by technological considerations and limitations, such as file sizes.

Given the rudimentary, dispersed and largely ad hoc nature of the recordkeeping "system" associated with the activities of the records creator in the case study, further analysis of the individual sub-components of A3 was not conducted during the walkthrough.

2.5 Select and Preserve Records (A4)

The key activities associated with this component of the MCP model involve selecting, acquiring, preserving and outputting records in the permanent preservation system developed in A1.2.3. At present, there is no formal, systematic and comprehensive permanent preservation system in place to preserve records generated by the activities of the creator in the case study. As well, there currently are no formal or systematic procedures in place for the long-term preservation of the CC GIS, or any of its components or individual records. However, certain short-term preservation strategies are employed, albeit on a limited and largely ad hoc basis. For example, the GIS specialist creates periodic 'archival' backups of the CC Database which he saves to his personal computer, the CDA's LAN server and/or burns onto CD-ROMs. There are no formal or systematic scheduling procedures in place for determining when such backups should be created. Instead, the backups typically are created immediately prior to any "significant" database updates or changes and serve primarily to facilitate roll-back to an earlier version should any of the new data added to the database later be found to be inaccurate or corrupt.[5] Exactly what constitutes a "significant" database update is an entirely ad hoc and undocumented decision made by the GIS specialist on a case-by-case basis. The GIS specialist also periodically migrates data and system files to newer versions of software, primarily in an effort to address immediate software obsolescence and current file useability issues, rather than in response to long-term preservation concerns.

There are no formal records appraisal procedures in place to assist the GIS specialist in deciding which records to preserve for the long-term. It appears that all such appraisals are based entirely on idiosyncratic criteria and are conducted in an ad hoc manner with little, if any, documentation. In the words of the GIS specialist,

> I sort of make subjective decisions along the way [about which records to preserve], 'well that one might be useful sometime, so I'll save it.' But it's only useful to me and it probably wouldn't be worth me going to the trouble to document that and make it available to somebody else at a future date because I'm not sure they would have the same sense of its value or they wouldn't be wanting to get to point B or something.

However, as the following excerpt from the interview transcript demonstrates, there is an underlying rationale that helps guide the GIS specialist's appraisal decisions with respect to the long-term preservation of the intermediary files that he generates:

> I guess I don't ascribe a lot of value to [the] intermediary [files] for the most part, because they're easily recreated and I think that they're kind of, in many cases, they're a function of the process that I have to go through, and the process will change. Whenever I'm trying to teach somebody about GIS, people will say, here's the data I have and here's what I want to get to. And the…first caveat that I offer them is, 'if you ask 10 different GIS people, they'll tell you 10 different ways to get there.' Each one of those different ways will produce different intermediary files, and so I don't think, for the most part, that those things have much value.

There are no procedures in place to verify the transfer of records or to monitor the efficacy of the limited preservation strategies that currently are used, especially in terms of the ability of those strategies to serve their intended function(s) and to adequately meet the base needs of the project. Nor are there any procedures in place to monitor the continuing integrity of the media on which

[5] Given the dynamic nature of GIS research and analysis in which only small subsets of all the data in the system are utilized at any one time, data errors often are not discovered until well after their initial introduction or propagation. Consequently, most GIS practitioners routinely save *all* previous backups in case there is a need to rollback to a version of the database several generations prior to the current version.

the backups and migrations are kept. This situation stems, in part, from a lack of personnel and financial resources required to implement a more robust permanent preservation system. However, the GIS specialist's views on the importance of the long-term preservation of GIS projects also is a significant factor contributing to the lack of a robust and systematic permanent preservation system and formal preservation procedures. In fact, the GIS specialist believes that within perhaps as few as 20 years hence nobody will be interested in or need his data or his results, and especially not the specific analytical techniques and data manipulations he used to arrive at those results. He bases this belief, in part, on the assumption that GIS technology will have changed so significantly within that time that his data and results will have been rendered completely "obsolete," and, in part, on his observations of how infrequently today's researchers rely on the data and results of archaeologists from 100, 50 or even 20 years ago. The underlying assumption here is that everything about the research that is worth preserving long-term will be contained in the publications of the results, so that there is little, if any, need to preserve anything else.

Given the lack of any formal preservation system and systematic preservation procedures associated with the records management activities in the case study, A4 of the MCP model was not broken down further during the walkthrough.

3.0 Summary

The overall process of designing, implementing and maintaining the record-making, recordkeeping and permanent preservation systems associated with the case study largely has been ad hoc, idiosyncratic and undocumented. There are, for example, no procedures manuals, metadata and classification schema, or retention schedules to help guide and systematise the records management activities of the records creator. The limited records management procedures that are instituted tend, on the whole, to serve more expedient and transitory purposes than those presumed by the MCP model (i.e., establishing, maintaining and preserving the authenticity, accuracy and reliability of records over the long term). Several key factors are responsible for this situation, including the constraints imposed by the availability of financial, personnel and technological resources, as well as the GIS training and experience level of the records creator. Compounding the situation is the fact that case study examined the creation, implementation and maintenance of a research database and GIS, and the records management activities associated with these processes, rather than a records management system *per se*, as is presumed in the MCP. As well, there is the difficulty of identifying which entities, if any, associated with the GIS research activities constitute *records* in the sense presupposed by the MCP model. In fact, the degree to which the records management systems, procedures and outputs associated with the case study even can be satisfactorily isolated and compared to those of the MCP model is, in many cases, debatable. Consequently, the records management "systems" identified in the case study are at best characterised as dispersed and rudimentary, and, in all instances, lacking many of the specific components and much of the functionality presumed by their MCP model counterparts.

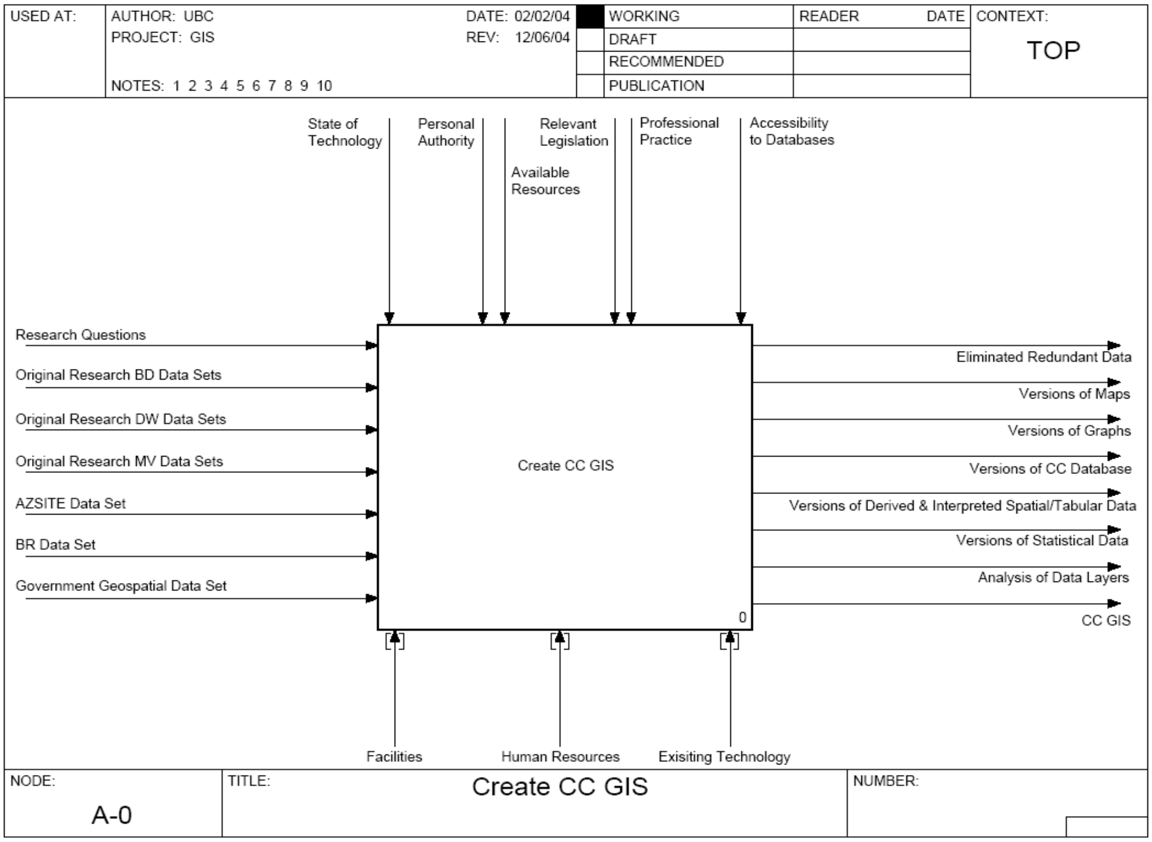
4.0 Conclusions

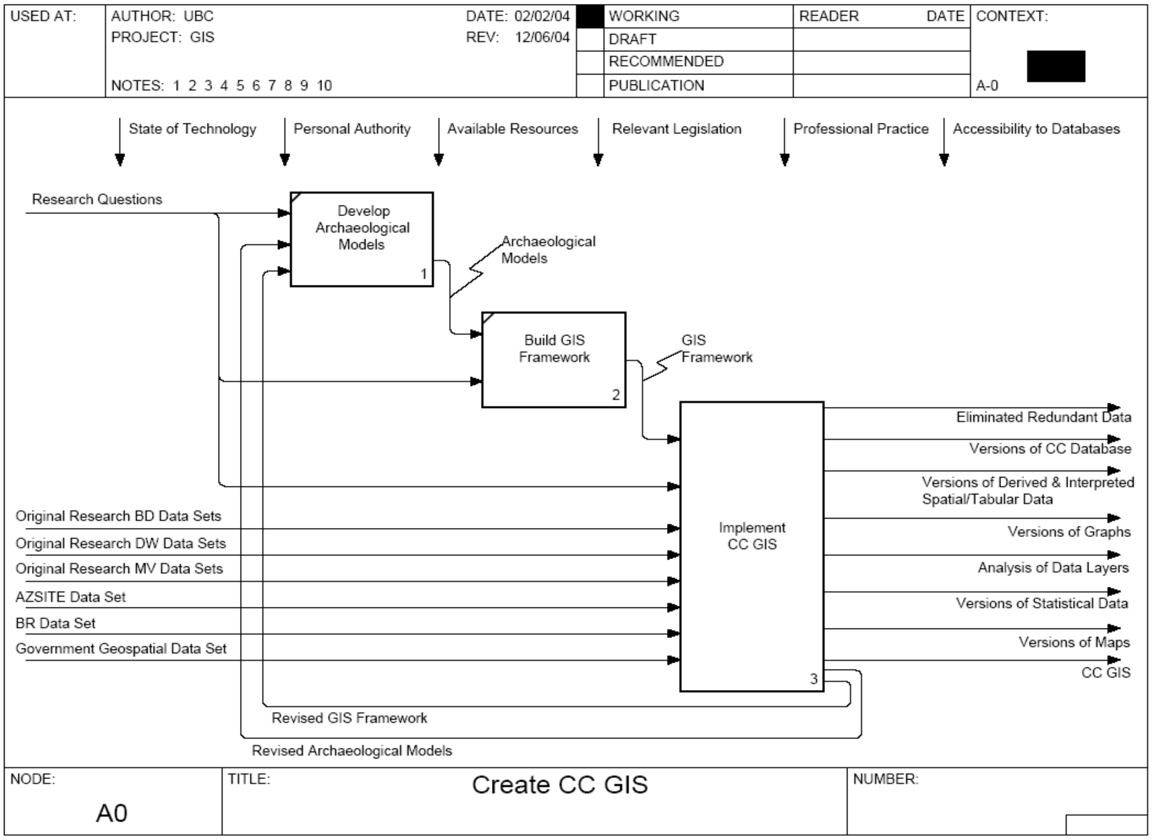
The MCP model presupposes the design, implementation and maintenance of a formal, dedicated records management system comprised of integrated record-making, recordkeeping and permanent preservation sub-systems. A concern raised during the walkthrough was that the current MCP model is too heavily prescriptive and, as such, does not adequately take into consideration the fact that records creators often work from their business process needs when designing their records management systems. In as much as the records management systems and procedures identified in the case study satisfactorily can be compared to those of the MCP model, it appears that the results of the walkthrough do indeed provide some support for this concern.

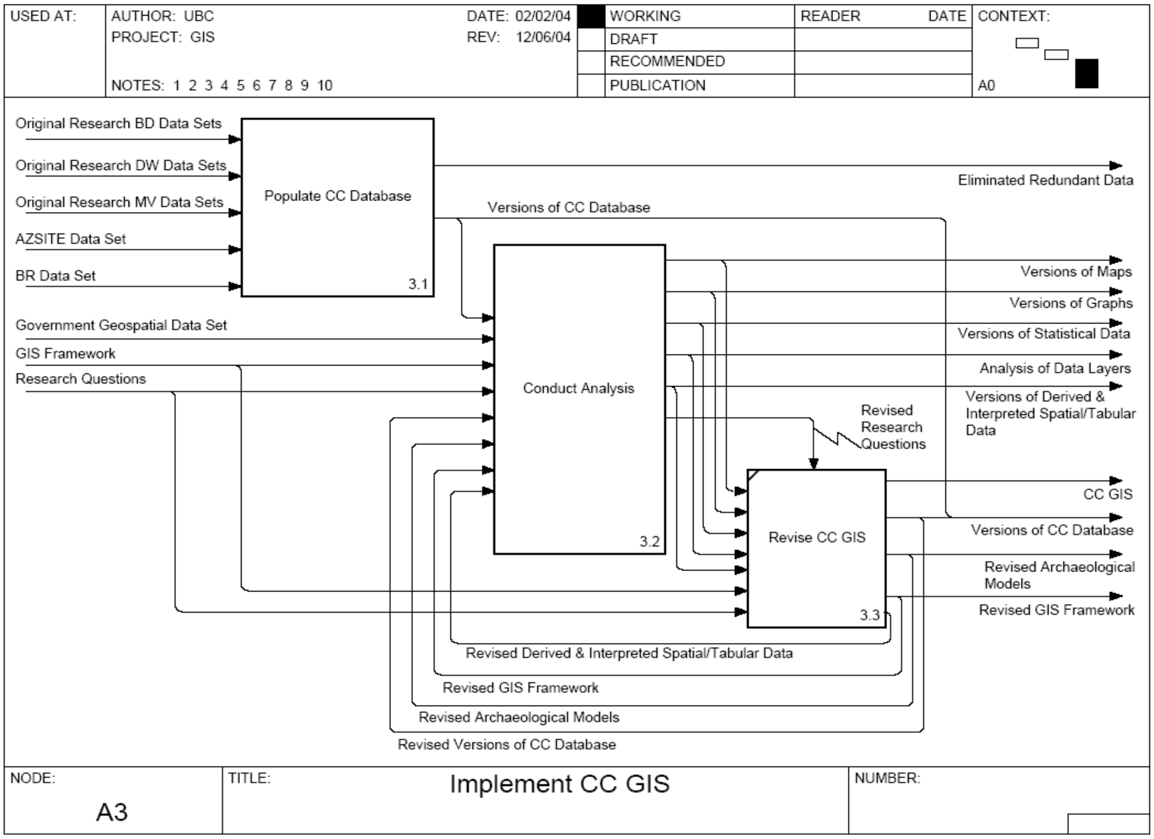
One of the major difficulties encountered throughout the walkthrough was the relative incompatibility of the nature of much of the case study data that were collected with those required to adequately test the MCP model. Clearly, what is needed is a model walkthrough protocol consisting of a set of questions designed to guide the walkthrough process. These questions should, in turn, be used to help establish a minimum set of data collection requirements for future case studies to ensure that the case studies are capturing the information necessary to adequately test the MCP model in future walkthroughs.

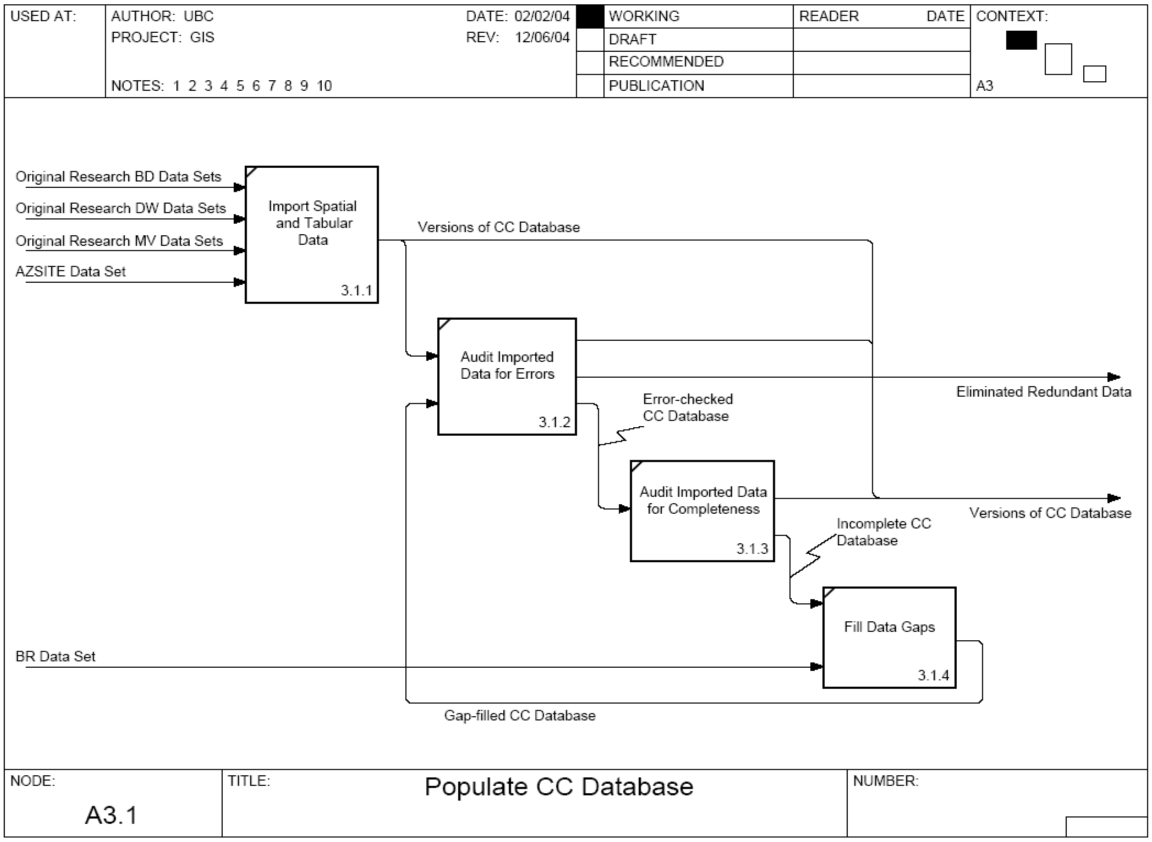
Appendix A

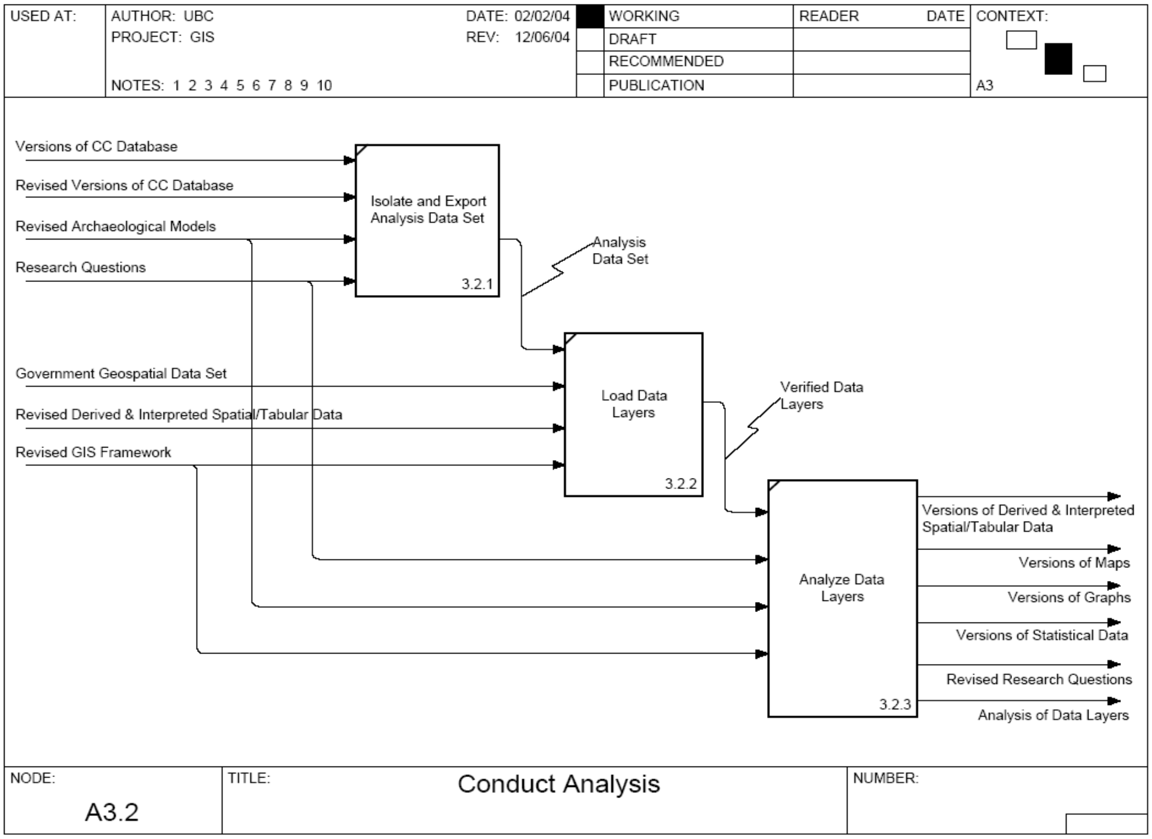
Create CC GIS:
Activity Model and Definitions

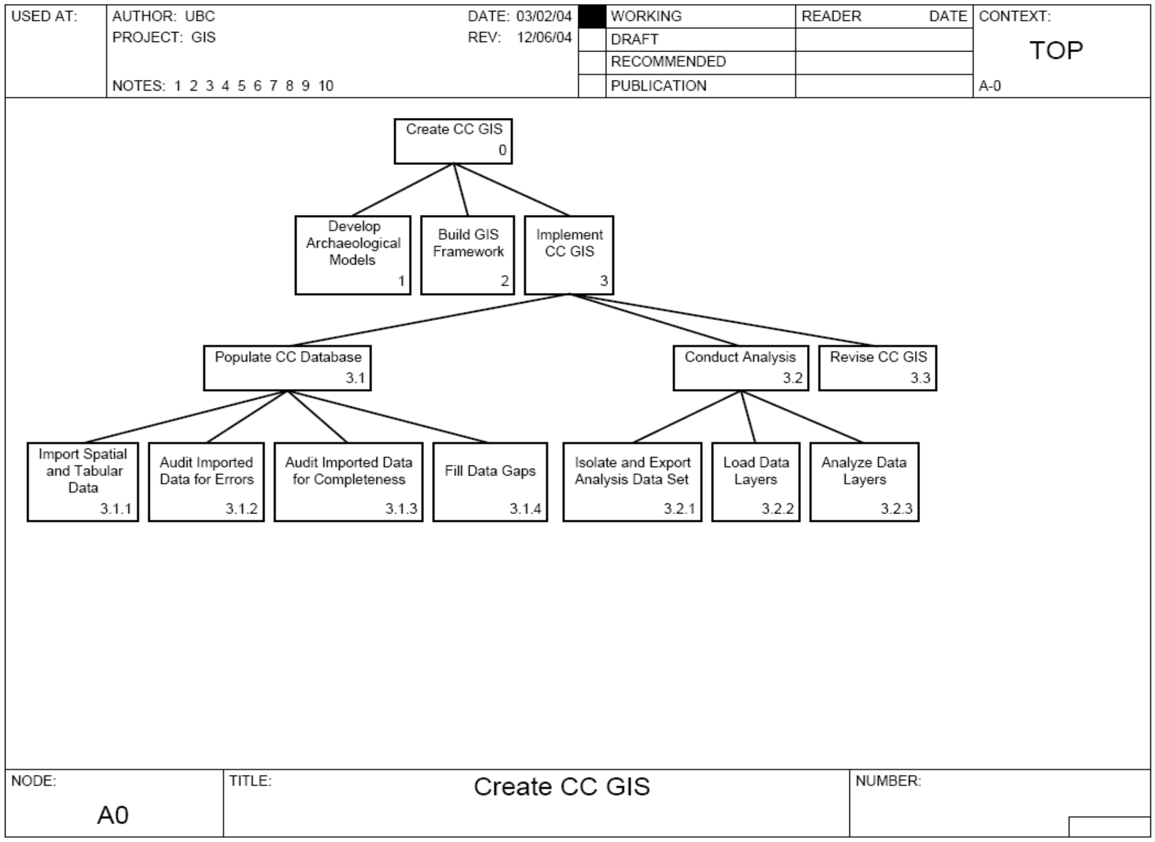

USED AT:
AUTHOR: UBC
PROJECT: GIS
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 02/02/04
REV: 12/06/04
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
TOP
State of Technology
Personal Authority
Available Resources
Relevant Legislation
Professional Practice
Accessibility to Databases
Research Questions
Original Research BD Data Sets
Original Research DW Data Sets
Original Research MV Data Sets
AZSITE Data Set
BR Data Set
Government Geospatial Data Set
Create CC GIS
0
Eliminated Redundant Data
Versions of Maps
Versions of Graphs
Versions of CC Database
Versions of Derived & Interpreted Spatial/Tabular Data
Versions of Statistical Data
Analysis of Data Layers
CC GIS
Facilities
Human Resources
Exisiting Technology
NODE:
A-0
TITLE:
Create CC GIS
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: GIS
DATE: 02/02/04
REV: 12/06/04
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A-0
State of Technology
Personal Authority
Available Resources
Relevant Legislation
Professional Practice
Accessibility to Databases
Research Questions
Develop Archaeological Models
1
Archaeological Models
Build GIS Framework
2
GIS Framework
Implement CC GIS
3
Original Research BD Data Sets
Original Research DW Data Sets
Original Research MV Data Sets
AZSITE Data Set
BR Data Set
Government Geospatial Data Set
Eliminated Redundant Data
Versions of CC Database
Versions of Derived & Interpreted Spatial/Tabular Data
Versions of Graphs
Analysis of Data Layers
Versions of Statistical Data
Versions of Maps
CC GIS
Revised GIS Framework
Revised Archaeological Models
NODE:
A0
TITLE:
Create CC GIS
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: GIS
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 02/02/04
REV: 12/06/04
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
Original Research BD Data Sets
Original Research DW Data Sets
Original Research MV Data Sets
AZSITE Data Set
BR Data Set
Populate CC Database
3.1
Eliminated Redundant Data
Versions of CC Database
Government Geospatial Data Set
GIS Framework
Research Questions
Conduct Analysis
3.2
Versions of Maps
Versions of Graphs
Versions of Statistical Data
Analysis of Data Layers
Versions of Derived & Interpreted Spatial/Tabular Data
Revised Research Questions
Revise CC GIS
3.3
CC GIS
Versions of CC Database
Revised Archaeological Models
Revised GIS Framework
Revised Derived & Interpreted Spatial/Tabular Data
Revised GIS Framework
Revised Archaeological Models
Revised Versions of CC Database
NODE:
A3
TITLE:
Implement CC GIS
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: GIS
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 02/02/04
REV: 12/06/04
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A3
Original Research BD Data Sets
Original Research DW Data Sets
Original Research MV Data Sets
AZSITE Data Set
Import Spatial and Tabular Data
3.1.1
Versions of CC Database
Audit Imported Data for Errors
3.1.2
Eliminated Redundant Data
Error-checked CC Database
Audit Imported Data for Completeness
3.1.3
Versions of CC Database
Incomplete CC Database
Fill Data Gaps
3.1.4
BR Data Set
Gap-filled CC Database
NODE:
A3.1
TITLE:
Populate CC Database
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: GIS
DATE: 02/02/04
REV: 12/06/04
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A3
Versions of CC Database
Revised Versions of CC Database
Revised Archaeological Models
Research Questions
Isolate and Export Analysis Data Set
3.2.1
Analysis Data Set
Government Geospatial Data Set
Revised Derived & Interpreted Spatial/Tabular Data
Revised GIS Framework
Load Data Layers
3.2.2
Verified Data Layers
Analyze Data Layers
3.2.3
Versions of Derived & Interpreted Spatial/Tabular Data
Versions of Maps
Versions of Graphs
Versions of Statistical Data
Revised Research Questions
Analysis of Data Layers
NODE:
A3.2
TITLE:
Conduct Analysis
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: GIS
DATE: 03/02/04
REV: 12/06/04
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
TOP
A-0
Create CC GIS 0
Develop Archaeological Models 1
Build GIS Framework 2
Implement CC GIS 3
Populate CC Database 3.1
Conduct Analysis 3.2
Revise CC GIS 3.3
Import Spatial and Tabular Data 3.1.1
Audit Imported Data for Errors 3.1.2
Audit Imported Data for Completeness 3.1.3
Fill Data Gaps 3.1.4
Isolate and Export Analysis Data Set 3.2.1
Load Data Layers 3.2.2
Analyze Data Layers 3.2.3
NODE:
A0
TITLE:
Create CC GIS
NUMBER:

Arrow Definitions – Create GIS (20040613)

Arrow Name	Arrow Definition	Arrow Note
Accessibility to Databases	Availability of archaeological databases with sufficient information relevant to the GIS and the proposed research questions.	CONTROL Not all data relevant to the GIS and the proposed will be accessible to the Center for Desert Archaeology, due to limited knowledge of the existence of, and/or restricted access to, researchers' archaeological data sets. As well, databases identified as relevant and available for use with the GIS may nevertheless contain insufficient archaeological data to adequately address all proposed research questions.
Analysis Data Set	A particular portion of the CC database needed for a project or to address a research question that is isolated and exported from the CC Database.	OUTPUT Excel format. Analysis data sets consist of spatial and tabular archaeological data isolated and exported from the CC Database and linked to government geospatial data for manipulation, analysis and interpretation.
Analysis of Data Layers	The process of visually and/or statistically analyzing the analysis data sets when linked to the government geospatial data and applied to the research questions.	OUTPUT May result in the creation of one of more of the following: maps, graphs, derived and/or interpreted spatial and tabular data, or statistical data. Output may be in the form of hardcopy printouts (either individually or in compiled reports) and/or digital files.
Archaeological Models	Preliminary conceptions of the GIS that encompass the ability to answer specific research questions.	OUTPUT
Available Resources	Means, whether financial or otherwise, available to the Center for Desert Archaeology.	CONTROL Resources including human, technological, financial and time resources. Being a non-profit organization, the Center for Desert Archaeology's resources are somewhat limited for the development of the GIS in question. Limited financial assistance has been provided for the development of the CC Database and resulting GIS in the form of grants, memberships, NSF grant and endowments. The Center provides an annual report to its board and the stipulations are not very tight regarding the use of financial resources from the board.
AZSITE Data Set	Comprised of a large data set of archaeological site information from numerous archaeological repositories in the state of Arizona and elsewhere.	INPUT In either Microsoft Access database .shp file. Consists of spatial and tabular data.
BR Data Set	Comprised of a large data set of archaeological site information from publications and archival repositories.	INPUT Hand transcribed from sources and then data entered into CC

Arrow Definitions – Create GIS (20040613)

Arrow Name	Arrow Definition	Arrow Note
		Database to fill in gaps identified in existing archaeological data sets. Consists of spatial and tabular data.
CC GIS	Coalescent Communities Project Geographical Information System	OUTPUT A multi-component acquisition, storage, manipulation, analysis, retrieval and visualization system for geospatial data and their attributes (i.e., tabular data). Comprised of hardware, software, archaeological and other public geospatial and tabular data, procedures/applications and persons (operators).
Eliminated Redundant Data	Duplicate versions of data found in the multiple data sets imported into the CC database.	OUTPUT Deleted from CC Database.
Error-checked CC Database	Versions of CC Database in which data redundancies and errors have been removed and/or corrected.	OUTPUT Prior to major data redundancy and error audits, archival versions of CC Database will be created.
Existing Technology	Technology available to the Center for Desert Archaeology at any given time.	MECHANISM
Facilities	Resources available to the Center for Desert Archaeology in terms of building space and location.	MECHANISM The Center for Desert Archaeology currently is located in an historic building in the central historic corridor named University District in Tucson, Arizona. The current facilities allow for the benefits of a central location, but also create challenges relevant to an historic practice.
Gap-filled CC Database	CC Database that has been augmented with supplemental archaeological spatial and tabular data following a data completeness audit.	OUTPUT Supplemental data are imported from the BR Data Set.
GIS Framework	System design in both a technological and conceptual manner.	OUTPUT Would include description of hardware, software, data and persons (operators) that constitute the GIS.
Government Geospatial Data Set	Comprised of public geospatial data needed for the analysis of the CC Database.	INPUT Consists of Digital Elevation Models (DEMs), jurisdictions (county, state and land jurisdiction boundaries such as Bureau of Land Management (BLM) land, Arizona State Land Department (ASLD) land, private land, etc.), and locations of current towns and cities.
Human Resources	Amount of persons and person-time devoted to creating and maintaining the GIS.	MECHANISM There are two people currently working on the creation of the GIS, BH and BR. The archaeologists who have devoted their collective data sets are indirectly working on the GIS, but are not involved in

Arrow Definitions – Create GIS (20040613)

Arrow Name	Arrow Definition	Arrow Note
		the development of the database and GIS structure and are not conducting the actual analysis. The researchers can ask for certain analyses to be conducted but do not directly interact with the GIS or CC database.
Incomplete CC Database	Versions of CC Database that have been audited for data completeness and found to contain certain areas in the existing data set that need to be augmented with additional archaeological data.	OUTPUT Auditing of data completeness is conducted by the volunteer (BR).
Original Research BD Data Sets	Comprised of the research conducted by BD regarding North American Southwest archaeology.	INPUT Microsoft Excel format. From multiple sources, which include original field work, archival research and literature searches. Data's origin is mainly from research conducted at Desert Archaeology and the Center for Desert Archaeology. Consists of spatial and tabular data.
Original Research DW Data Sets	Comprised of the research conducted by DW regarding North American Southwest archaeology.	INPUT Microsoft Excel format. From multiple sources, which include original field work, archival research and literature searches. Data's origin is mainly from research conducted at the Museum of Northern Arizona. Consists of spatial and tabular data.
Original Research MV Data Sets	Comprised of the research conducted by MV regarding North American Southwest archaeology.	INPUT Microsoft Excel format. From multiple sources, which include original field work, archival research and literature searches. Data's origin is mainly from research conducted at Crow Canyon Archaeological Research Center. Consists of spatial and tabular data.
Personal Authority	An influence exerted on an idea based on experience of an expert on a certain subject.	CONTROL Personal authority can control what is decided to be the authoritative interpretation of archaeological assemblages or data that are received from multiple sources.
Professional Practice	Opinions and notions specific to archaeology relating to professional and ethical practice.	CONTROL See Society for American Archaeology's code of ethics (www.saa.org)
Relevant Legislation	Pertinent legal requirements relating to archaeological activities.	CONTROL Includes, but not limited to Native American Graves and Repatriation Act (NAGPRA), Arizona Antiquities Acts and the Arizona Revised Statutes.
Research Questions	Questions that investigate the issue of "coalescence"	INPUT

Arrow Definitions – Create GIS (20040613)

Arrow Name	Arrow Definition	Arrow Note
	during the prehistoric Southwest.	Questions include issues of migration, population, archaeological site sensitivity (in terms of present land development), predictive modeling to locate potentially undocumented sites, cost distance analysis, interaction between sites, etc.
Revised Archaeological Models	Revisions to earlier conceptions of the GIS that encompass the ability to answer specific research questions in light of new insights gained through analyses.	OUTPUT
Revised Derived & Interpreted Spatial/Tabular Data	Derived and interpreted spatial and tabular data that have undergone further modification(s) and have been monitored through changes.	OUTPUT Consists of derived and interpreted spatial and tabular data used in analyses that are subsequently modified and re-applied to the same and/or different analyses.
Revised GIS Framework	GIS Framework which has been re-evaluated in light of any changes to any of the individual components of the GIS.	OUTPUT Revisions to research questions and CC Database might incur change to the GIS Framework.
Revised Research Questions	Research questions that have been changed or added due to analysis findings.	OUTPUT
Revised Versions of CC Database	Versions of CC Database that have been monitored for any changes resulting from analyses.	OUTPUT Microsoft Access format. Contains spatial and tabular data imported from original BD, DW and MV archaeological research data sets, as well as from AZSITE and BR archaeological data sets.
State of Technology	Level of technology relating to processing power, software capabilities, GIS and relational database design.	CONTROL
Verified Data Layers	Analysis data layers in which spatial and tabular data have been error-audited and properly linked.	OUTPUT
Versions of CC Database	Versions of CC Database, a relational database that consists of pre-existing archaeological data on the prehistoric Southwest between A.D. 1200-1540.	OUTPUT Microsoft Access format. Contains spatial and tabular data imported from original BD, DW and MV archaeological research data sets, as well as from AZSITE and BR archaeological data sets. Prior to major error or completeness audits, backup and/or archival versions of CC Database are created. May consist of hardcopy printouts and/or alphanumeric data saved in digital form.
Versions of Derived & Interpreted Spatial/Tabular Data	Versions of spatial and tabular data resulting from algorithm manipulation and visual analysis of base (i.e., original) spatial and tabular data exported from CC	OUTPUT By-product of algorithm manipulations of base spatial and tabular data exported from CC Database and linked to government

Arrow Definitions – Create GIS (20040613)

Arrow Name	Arrow Definition	Arrow Note
	Database.	geospatial data. Typical manipulations include cost-distance analysis, site sensitivity analysis, and analysis of site density and interaction between sites. May consist of hardcopy printouts and/or alphanumeric data saved in digital form.
Versions of Graphs	Versions of diagrams depicting the quantitative relationship between two or more base, derived and/or interpreted spatial/tabular data variables.	OUTPUT Originate in ArcView. May consist of hardcopy printouts and/or images saved in digital form.
Versions of Maps	Versions of visual representations of geographic areas delineated by government geospatial data onto which selected spatial and tabular archaeological data from the CC Database are plotted or otherwise visually represented.	OUTPUT Originate in ArcView. May consist of hardcopy printouts and/or images saved in digital form.
Versions of Statistical Data	Versions of numeric data resulting from statistical analyses of spatial and tabular data.	OUTPUT By-product of statistical analyses of: 1) base spatial and tabular data exported from CC Database and linked to government geospatial data, and/or 2) derived or interpreted spatial and tabular data. Typical example of a statistical analysis is the autocorrelation of archaeological site distributions. May consist of hardcopy printouts and/or numeric data saved in digital form.

Activity Definitions – Create CC GIS (20040613)

Activity Name	Activity Number	Activity Definition	Activity Note
Create CC GIS	0		
Develop Archaeological Models	1	To develop archaeological models relating to the research questions of the Coalescent Communities Project. To brainstorm and notate representations of the nature and make-up of the GIS in relation to these models.	These models include sets of hypotheses that simplify complex observations within the archaeological record (NB, as used in the archaeological profession, archaeological record refers to cumulative archaeological knowledge, not a physical entity in an archival sense of the word record). Modelling is used for organizing and structuring the data and data collection priorities for the GIS. As such, they help increase the accuracy and precision of functional, temporal and spatial qualifiers within the GIS.
Build GIS Framework	2	To develop the framework of the GIS into a system that can be implemented.	The framework is designed by the GIS Specialist in a hypothetical, rather than physical sense because it is never formally constructed on paper or delivered to anyone in a formal product. This framework includes

Activity Definitions – Create CC GIS (20040613)

Activity Name	Activity Number	Activity Definition	Activity Note
			implementing the archaeological models that were created into a system that can be successfully implemented in relation to the research questions and goals of the organization.
Implement CC GIS	3	To populate and use the GIS.	
Populate CC Database	31	To import BD, DW, MV and AZSITE data sets into the formative CC Database and to audit those data for error, redundancy and completeness.	
Import Spatial and Tabular Data	312	To combine the BD, DW, MV and AZSITE Data Sets from various formats into one Microsoft Access Database.	It is during this stage that the GIS Specialist has to manage within the database the multiple fields that different researchers used or did not use.
Audit Imported Data for Errors	312	To rid the CC Database of redundant and erroneous data.	Involves eliminating the duplicate data within the numerous data sets that were brought together to form the CC Database. Also involves searching the database and eliminating errors relating to spatial coordinates, site description, temporal range, cultural affiliation, etc.
Audit Imported Data for Completeness	313	To identify data gaps within the CC Database.	These gaps include data relating to geographic areas/regions and sites relating to specific cultures and/or time periods. Involves looking at the CC Database and the related maps that can be produced by the existing data. This process is conducted by a volunteer (retired professor) and the project's associated researchers.
Fill Data Gaps	314	To augment the formative CC Database with data gathered by the volunteer (BR) for the purpose of filling identified data gaps in the CC Database.	Process of gathering these data involves research at various repositories (the Arizona State Museum, in the organization's fieldwork and related publications and other publications within the libraries at the University of Arizona and the Arizona State Museum). Once the data are gathered, they are entered into the CC Database (collectively, these data are called the BR Data Set).
Conduct Analysis	32	To visually and statistically analyze portions of the CC Database through ArcView and other software.	
Isolate and Export Analysis Data Set	321	To isolate and extract certain portions of the CC Database and export them into an ArcView Project file.	Usually carried out through transforming the data into Excel files.
Load Data Layers	322	To initiate ArcView software; open a Project within ArcView and import	Involves an iterative process of standardizing and preparing the data layers for analysis that typically involves editing, linking and verification of the

Activity Definitions – Create CC GIS (20040613)

Activity Name	Activity Number	Activity Definition	Activity Note
		isolated data set exported from CC Database into the Project.	data layers.
Analyze Data Layers	323	To conduct analysis (spatial and temporal) on specific data layers within ArcView with respect to the research questions.	The GIS Specialist establishes spatial relationships between data layers within ArcView through map algebra. This process may include visual analyses and/or calculations relating to population statistics, predictive modeling, archaeological site sensitivity, cost distance analysis, statistics to calculate interaction between sites, etc.
Revise CC GIS	33	To evaluate the current analyses and data (base, derived and interpreted) and make any necessary corrections, additions or completely redo certain portions of the analysis. To rid the CC Database of repetitious and inaccurate data that are identified during the analysis of the database. To monitor the changes made to the formative parts of the GIS. To identify any changes that might need to be made to the formative parts of the GIS.	This evaluation process includes looking at maps and spatial and tabular data resulting from analyses, as well as evaluating the CC Database itself. If any new research questions arise, revisions are made accordingly.