Joining the Club: Issues, Problems, and Practices in Initiating GIS

Caroline Beebe

School of Library and Information Science. Indiana University.
Bloomington, Indiana 47405. USA
E-mail: beebe@indiana .edu

Abstract

Organization and protection of data is crucial to research and the preservation of scientific projects. Digital data collection in the field is the first step toward ensuring future access, but it does not come without problems. Archaeologists, who are digital technology newcomers, have few places to turn for information, guidance, and training. Most textbooks cover digital issues in less than a chapter, more often a page or even a paragraph. Reading the Archaeology and GIS listservs, the newcomers may encounter the technologically astute, who often underestimate their expertise and resources, when responding to technical or introductory problems. The problem of understanding solutions is exacerbated by a lack of common vocabularies, between the various domain experts: archaeologists in different culture areas, geophysical surveyors, information managers, ceramicists, lithics experts, osteologists, hydrologists, etc. The list of specialists, that may be involved in a single investigation, is ever increasing, as is the specificity of their language. Somehow, the principle investigator must provide for the digital collection, integration, and sharing of all their data. Even assuming that the archaeologist recognizes and accepts the long term advantage of preparing their data, for digital consumption, the pathways into the digital world, and GIS, are often not advantageous enough, to overcome the technological problems, and lack of information guides.

Entering the digital data realm, raises issues of standardization and accuracy that are far more problematic than in their paper counterparts. Digital data must be precisely accurate, in order to be shared, and the digital process creates an affect of truth that is dealt with more softly in paper record keeping. Fundamental problems of terminology control and identification of data units manifest in the practical problems of data design and management. These, in turn, create management problems in the selection of hardware and software, and then, in field training. In addition, the software imposes adjustments to the data collection process, in the field, followed by post-processing problems, in the laboratory.

Most of these problems become solved, once the archaeologist has joined the digital club and gone through the initiation and learning process. In reality, that takes several persevering years. This paper documents the initiation process, as an archaeology team in the Belize jungle is experiencing it. General issues and problems are identified, and suggestions are made for practices that will streamline the digital initiation process. Focus is on the geophysical survey, which forms the foundation of the digital data collection, as it migrates to the GIS environment. The data model, data codes, and processing practices are presented in the belief, that they are generic enough to be adapted to a broad range of archaeological situations.

Introduction

Traditionally the archaeologist studies the patterning of past human activities in space and time. The assemblage of their findings is reported in textual reports, with hand drawn charts and maps. This creates an access problem for secondary analysis or replication study. The raw data must first be made accessible, then documented, as to its organization and coded meaning, then manually sorted, to find the information of interest. Advances in digital technology over the last decade, promise to address these information management problems through databases and geographical information systems (GIS), that are affordable and easy-to-use. Described below are some conceptual foundations of the archaeological record, in relation to electronic record keeping, and the experiences of an archaeological project in Belize, that is attempting to make the transition from manual to digital record-keeping.

The archaeological record

The archaeologist patiently and meticulously excavates the layers of cultural material, that have been deposited through time. Thus begins the *archaeological record*, the collection of material remains that the archaeologist uses as evidence, for the ordering and describing of "ancient events, and to explain the human behavior behind those events" (Sharer and Ashmore, 1987, p. 10; Shiffer, 1987, p. 4). As the archaeologist progresses, the excavation process is thoroughly documented, to "describe what was recovered and analyzed and what procedures were used," (Shiffer 1987, p. 339) and these documents also become a part of the archaeological record. These records and the material remain:

> ... uncovered during an excavation, are a non-renewable cultural resource. Once an archaeological site has been disturbed it can never be restored to its previous condition. ... if the data about the site is incorrectly or incompletely recorded, the historical value of the excavation will be minimal. ... there is no chance to go back and do it correctly a second time. (Mcmillon,1991, p. 20)

These records and materials form the primary data for all prehistoric archaeological research. The archaeologist returns from the fieldwork, where the data have been described, then synthesizes and analyzes the data. A document is then written, that provides explanations, supported by physical evidence for the particular research questions being addressed. Those documents, describing what the archaeologist has learned, whether published or not, become additional, secondary data in the archaeological record (Silverman & Parezo, 1995; Patrik, 1985), along with analysis summaries and methodologies. And then, one day in the future - ten, fifty, a hundred years later, or more - when there are new methodologies or theoretical bases, another archaeologist may want to reconsider a particular culture, excavation, or synthesis. "How much of what had been in their [excavator's] notes will never have seen print, and how much of what had been published will demand reexamination of the primary records - if these are available?" (Silverman and Parezo,1995, p.1).

In addition, what is the dispensation of the primary records, the physical objects, and the evidence on which their publication and notes are based? Archaeology collects the material evidence of culture. The research of the archaeologist includes text documents, as well as natural phenomena and cultural material, and results in written reports of analyses. These mixed data might include: artifacts, architecture, human and faunal remains, and geographic and geologic information. The only primary data that is available to other researchers is in the form of printed catalog tables or images in the paper based publication of analysis reports. The objects themselves are in repositories, either universities or museums, with an individually organized paper trail to their location. These information storage and access methodologies of the archaeologists are not ensuring the preservation of the archaeological record and its availability to future generations. And, as noted by Fagin (1995) and Rice (1996), the length of elapsed time to publication can be detrimental to intellectual discourse, in common problem areas and the synthesis of new knowledge.

Electronic technology and the archaeological record

The early introduction of the computer in the 1950's, into the archaeological process, was met with acknowledgment for feasibility, when applied to artifact sorting tasks, but disdain for its introduction into the methodological process (Gardin, 1989, p.6). Gardin quickly lost interest in the artifact sorting aspects of computer applications, and went on to explore computer-assisted methodologies for discourse analysis, which continues today in the domain of artificial intelligence. Unfortunately, the potential uses of the computer for artifact sorting, succumbed to the disdain, for the computer in general, and resulted in little use, within the discipline. Further research regarding the use of electronic information systems in archaeology, tended to be focused on specific software products, technologies, or statistical analysis methodologies (Carr, 1984; Chenhall,1975; Gaines, 1974), for application in the creation of analysis reports, as opposed to mechanisms for representing diverse information types in an electronic systems design. The results of these forays into computer technology were either too specific in application, or in technology, to generate a broad appeal for the advantages the technology might hold, in general, for the discipline. Currently, all electronically available archaeological data are held by individual archaeological projects, cultural resource management repositories (including museums), or emerging examples of WWW-based, data archives (see ADS and ADAP).

Methodologies for creating the archaeological record, generally give little guidance on the creation of the an electronic record (Alexander,1979; Kenworthy, 1985; McMillon, 1991; Sharer & Ashmore, 1987; Silverman & Parezo, 1995), which would form the data bank of an archaeological information system. Permanence and labeling are two of the highlighted characteristics of all data banks, but there is no guidance on methodologies, which might facilitate indexing capabilities for intra-site analysis. Most methodological texts provide samples of paper recording forms, but no guidance on either the construction of those forms, or the conversion to electronic format. When electronic methods are mentioned, the emphasis is on the need for consistency in the data (Flood, Johnson, & Sullivan, 1989; Sharer & Ashmore, 1987), or on analytical techniques to be applied once the electronic data is available (Carr, 1984; Hirschheim, Klein, & Lyytinen, 1995; Michalski & Radermacher, 1992). Each archaeologist must create their own recording forms, or adapt or adopt forms from other sites.

The text *Sites and Bytes* (Flood, Johnson, & Sullivan, 1989) shows the most promise, in providing data description or structure guidance by presenting multiple site recording forms and exemplars of individual approaches to the creation of electronic data collections. The focus however is turned, first, on actual data capture methodologies, e.g., how to perform a land survey and systematic collections of surface finds, or, second, on the advantages and disadvantages of specific software applications. In discussing similarities across the recording forms, three components are identified: detailed information using site specific methodologies, checklists of general information, and a "... set of index data for the site, e.g., type of site, recorder's name, data map sheet, grid reference, cadastral information, etc. Presumably is it this type of index which will increasingly be accessible through the computerized systems being discussed." (Flood, Johnson & Sullivan, 1989, p.3) This description of an index is more similar to the meta-data, that is being proposed and provided for in the new WWW offerings of archaeological data (see ADS and ADAP), or by the Dublin Core of meta-data standards for images.

The new WWW archaeological archive services provide information about the information (meta-data), in a specific collection or excavation. The meta-data is not the primary data of the excavation or collection, but provides the description of a given data collection in order to facilitate the discovery of data that might be useful for a given research direction. When archive services are provided, there are guidelines for the file format descriptions of the data banks, but no guidance on the data representation methodology that could be used to facilitate a given archaeologist's information retrieval system design. For example, one data bank might represent potsherds, by lot number counts and weights, while another data bank might represent the potsherds, by type variety, followed by counts and weights. In order to retrieve similar data from these two data banks, the researcher needs to know the relationship between the lot numbers and type varieties, as well as the specific method of type variety that was used to classify the potsherds, or, alternately, the relationship of real world coordinates to lot numbers. The disparity in representation may result from the focus of the researcher's problem domain. A ceramicist may represent the data in their analytical report as type varieties, which are the foundation of analysis for sequencing. An

excavator, interested in monumental architecture, may report potsherds by lot number, as the ceramic sequencing is part of the post-analytical process for architecture. Additionally, each ceramicist may subscribe to a different schema of type varieties, or have created personal variations. Type varieties, then, become problematic for cross-data comparisons. This leads to the question of whether or not different problem domains highlight different attributes of the data, when the data bank is created, and perhaps these are reconcilable through the use of a generic data model.
The potsherd situation, described above, is similar to that, in the text based information systems of libraries:

> ... prior to the development of computers, a library that arranged its holdings according to Dewey was not able to use LC classification. But a computer is multilinear and hence not limited to any particular set of relations. It allows books to remain in Dewey order while simultaneously making them accessible via different classification systems. (Quinn, 1994, p. 142).

Thus the use of a generic data model would allow different archaeological researchers to compare each other's classification, analysis methods for any given type of material, providing a larger frame of reference for the development of analysis and the extension of knowledge. Carr (1984, p2) highlights this need when he concludes:

> ... the pace of progress along both theoretical and methodological lines of advance has been constrained by the limited effort that has been devoted to integrating them. Until very recently, little attention has been given to formally developing and maintaining, during analysis, *logically consistent* relationships between the theoretical developments, technical developments, and the data and phenomena of interest. *Yet it is precisely this concordance between theory, technique, data, and phenomena that is required for analysis, theory building, and technical development to be relevant accurate, meaningful and efficient.*

Carr continues with an explicit distinction between categories of information, in which the broadest, or most basic category, is the reference to the real world, that the researcher selects for study. These real world referents have many facets or phenomena, that could be studied from different perspectives and problem sets. He uses the term "data structure" to "refer to all the variables, observations, and the relationships among them within a data set (bank) regardless of whether or not they reflect the phenomena of interest." (P11) He, therefore, refutes the position proposed by Scholtz and Chenhall (1976), that the archaeological data banks are worth the effort "only if they are created to satisfy realistic and precisely defined needs ..." (p96) In other words, Chenhall "organizes his operations of system development around the *research activities,* in which the system is to participate ... in contrast, Parker et. al. organize their operations of systems development around the *nature of the data* ... data modeling is fully complete before the consideration of the files." (Carr, 1984, p. 88) ; Hirschheim, Klein, & Lyytinen (1995, p155) describe data modeling as the process, by which the nature of the data is examined for its structure and then represented in the data bank. It "involves the design of a knowledge representation schema," i.e., a generic, standardized, data model.

Chau Hiix goes digital

In archaeology, real-world coordinate data (location) is the most basic physical representation of an artifact. Artifacts acquire a physical location in the excavation, the moment they are uncovered. The site of the excavation itself, has a physical location in the real world. It is the most objective fact available, even if it contains subjectivity, on the part of the surveyor, to select the level of acceptable error in measurement, as in judgments on the location of a specific wall, within a pile of rock tumble. When an artifact is uncovered, an excavator must decide what to record about the discovery. The sensory inputs, focus of attention, and noise in the situation affect the excavator's judgment and record keeping. Nonetheless, once the artifact is removed, its real-world situation is destined to be based on the records made by the excavator. Most GIS (Geographical Information Systems) projects report on the digitization of geophysical data, through the scanning or manual digitizing of existing maps (McManus 1997) and the hand entry of artifact location data. A notable exception is the work of Steve Nickerson (1997) and the Humeima site in Jordan, where they experimented with digital measurement of the standing architecture. Recognizing the value of creating a digital portion of the archaeological record for analyzing, publishing, and sharing data, Dr. Anne Pyburn forayed into the realm of digital data capture as the foundation for developing an electronic data bank. In 1995, Pyburn's Chau Hiix project was still in the early stages of creating the site map. She chose to attempt to capture the geophysical survey data, directly to the computer, for site map generation, mound architecture evaluation, and later use in GIS based intra-site feature analysis.

Chau Hiix is an unlooted Maya site in Belize, that shows continuous occupation from 1200 BC to 1500 AD. It rests on the edge of 5 kilometers of seasonal lagoons, that show evidence of an extensive water control system. It also rests on the center of a 40 kilometer transect, between two other previously excavated sites. As stated by Pyburn, the research goal is to explore inter-site, political relations with the two sites on the transect, and to test competing models of ancient Maya political economy, with several complimentary lines of evidence: area populations, agricultural features, post-classic settlement, and a long-lived, civic center.

Early pace and compass maps provided a sense of the site center and extent of the immediate occupation area. Fortuitous funding provided for the purchase of a total station for the 1996 season. The purchase process took many months and suffered from vocabulary problems, between the vendors and the archaeologists. The vendors are generally civil engineers. The archaeologists are not sophisticated in that style of mapping, and its attendant vocabulary and accuracy requirements. Take for example the subtle differences between a geophysical survey and an archaeological survey. The former may stake out structural data points, the latter may contour map mounds in existing terrain, or make a collection of surface samples at regular intervals. Searching an information database on the term *survey* will result in very little information, about archaeological methods and applications of geophysical survey. Long after the fact, the magazine P.O.B (Point of Beginning) was discovered, with its explanations of technical specifications, equipment comparisons, and market updates. Survey software and a data recorder were purchased, unfortunately not from the same vendors. Each vendor could only explain how his or her own piece of equipment functioned. Total configuration took weeks of work, and training could only occur piecemeal. Again, software and data recorder vendors are used to civil engineers, and the vocabulary differences caused communication problems in the archaeologist's learning process. In addition, it was discovered that many vendors of total station equipment have never used a transit, so they have difficulty in training the surveyor to move from the manual system to the digital system. This proved problematic once the electronic equipment was moved into the field.

Chau Hiix is nestled in a cahune palm jungle, seven miles by boat or truck from the nearest village. Electricity is provided by 12-volt batteries charged by a solar panel, with a gas generator for back-up. The laboratory often reaches 115 degrees, causing the DC/AC inverter to shut itself off, thus limiting the hours of computer time available for map making. The mapping software can only run on one computer at a time, and can be directly connected to the data recorder for data downloading. There is one Windows 95-based laptop for mapping, and three 286 DOS-based portables for inventory and text editing. Two Mexican surveyors were hired for the season, one to work with a transit and map the outlying regions, the other to use the total station and map site center. In the training situation, there already existed communication problems *within* the English language. Since English was not the surveyors' first language, communication problems were exacerbated, when trying to train them to move between transit and total station.. This is when another problem manifested, in regard to process changes.

Evidently many surveyors use the total station as a stand-alone device without a data recorder. They functionally know how to shoot points, using the laser instrument, or 'gun' alone, manually recording data points in their paper field book. They may not understand, however, the relationship of the data recorder and electronic mapping as a total process. This problem is similar to moving from a typewriter to a word processor. There are conventions to be learned, in order to be fully functional, such as the concept of the cut-paste buffer. Similarly, there are total station conventions that must be internalized, before the surveyor becomes efficient with electronic equipment. For example, unlike a transit, this particular gun, itself, must always be turned in a clockwise direction, and, descriptive codes must be exactly consistent in order to be electronically retrievable. In addition, it is possible to enter plotting codes as the points are shot. These codes will facilitate the lines and polygons of the digital plot map, saving the surveyor time in the post-processing, cartography stage. Use of plot codes, however, requires a firm grasp on the interstices of the whole digital process. These learning problems were somewhat alleviated in the 1997 season, by hiring a surveyor, who had not only used both a transit and a total station, but had personally generated electronic maps. New problems became evident, however, once the season was over, and the post-processing began, back at the university's laboratory.

Five easy pieces

In theory, there are five easy pieces to the digital capture of geophysical data: shoot, edit, and cartography, analyze, and publish. Each is worth a hefty chapter in a textbook, on digital survey for archaeologists. The following discussion will only focus on the highlights of each piece, as it relates to the digital archaeological record, and preparations for analysis using GIS.

Shoot: Problems have already been described, regarding the purchase of hardware and software, as well as the need to train the surveyor in the specifics of the purchased technology. Additional problems were discovered in post-processing the data. They began with the understanding that every point recorded must have a unique identifying number, otherwise the mapping software and the archaeologist get confused, as to which points are being referenced. Point number tracking then becomes a logistical problem. If the surveyor uses a single file to collect all data points, the file becomes very large and is slow running on the data recorder. If the surveyor moves around to various excavations during the day, then a given excavation has a collection of non-sequential points for the map. It is then more difficult to ask the computer to draw the map of the given excavation area, as all the points must be entered individually, rather than as a single series. On the other hand, having the surveyor maintain multiples files for general excavation areas, creates problems in the overlap of the files. If a feature is uncovered in the overlap area, its surrounding area is then split between two different files. This problem cascades, as you map and excavate more and more connected areas. The most efficient solution to the question of single versus multiple files has yet to be identified.

In good practice, the surveyor must maintain clear communication with the excavators and with other surveyors. Chau Hiix is a very large site, with potential outlying, building groups anywhere from 1-5 kilometers. There have been as many as three surveyors, working simultaneously on any given day: one on a transit, one on the total station, and one plotting maps in the lab. Choosing the multiple file route at Chau Hiix, each surveyor has his or her own set of data collection files, constructed under a systematic naming scheme. They also have been assigned a unique range of point numbers, as well as a range of feature numbers, for assigning to new buildings and surface collections. Transit data must be included in this coordination, process as it will be hand-entered in post-processing. Poor communication can result in point number overlaps, or more than one feature with the same identifier. When overlaps occur, they must be reassigned in post-processing, causing conflict with the excavator's notes. Thus the surveyor must communicate with the excavators, both the unique point number *and* the Northing and Easting for any given feature. This facilitates linking artifacts and notes with the survey data, and guarantees correct referencing, given any errors. In addition, as surveyors encounter new situations, they may develop new point description codes. These must be shared with everyone involved in the excavation and maping data, both for referential understanding and future consistency in usage.

Edit: No matter how good the surveyor, there will be mistakes. These might be simple key press errors, or double shots taken, or forgotten rod height changes, or a counterclockwise turn of the gun, resulting in rotated point plots. The general rule at Chau Hiix is to make no digital corrections during the actual survey, but wait until the data has been downloaded in the lab. The data recorder, itself, has limited editing functionality, and is physically a small device. This makes it difficult to file edit or to grasp the whole picture of the day's work. Once the data is downloaded, edits on the larger computer are easier and safer, as there is already a back-up. This represents a change in process for the surveyor who is used to keeping a survey record as points are shot. Traditionally the surveyor writes down all the data point information, rod height and readings, angles, etc. The digital data recorder now collects all that information. The surveyor uses their traditional paper survey record to note errors, or questions about areas to re-shoot, to note terrain details or vegetation, or to record a point number overview for communication with excavators and other surveyors.

Back in the lab, the surveyor downloads the day's files into the computer for back-up, editing, and planning. The raw data can be printed out or edited on screen, for correction of errors made when the shot was taken, or for point description refinements. Due to the specific nature of the survey software, these edits occur in the *raw* file, and not the system *archive* file. The raw file must be carefully backed up once the surveyor is done editing. These edits represent the most correct data, while the *archive* file maintains the data as it came from the field. This exemplifies a file management problem, that must be diligently attended to from this point on. Developing the initial file naming scheme, for the surveyor's data files, it must be taken into consideration that there will be many files associated with a given data file. These might include: file types (archive, raw, converted, pictures or maps, etc.); concatenated files for overlap features; and multiple, edited versions of any of the associated files by either the surveyor or future analysts. For the foreseeable future, these file names must all be maintained in eight characters, so they remain readable on a broad base of computer equipment. This is not to suggest that all versions of all files need to be maintained for posterity. Data management requires diligence, in the deletion of working versions, back-up of completed versions, and documentation about everything in the archive. There needs to be a naming scheme that allows surveyors and archaeologists to work with the files in their analysis while maintaining some sort of referencing to avoid confusion. In addition, whenever possible, there should be an ASCII version of all final files.

Cartography: This is accomplished using either the transit or pace and compass method, to survey results in a surveyor's sketch map, in the field. The data points and sketch map are then used to hand plot the map on graph paper, back in the lab. These plot maps may then be turned into publishable maps of the site. Somewhere in this process, the surveyor, acting now as cartographer, has drawn in an estimated size and shape of the buildings and has calculated some contour lines in strategic locations. The final publishable map is a creative and interpretive effort on the part of the cartographer. The actual points, that were collected to indicate the building locations, have been rectified in the best estimation of the surveyor, to be reported as rectilinear buildings, the "mahlerization." For this transit-based method, this level of accuracy is adequate, because its goal is to present a map, that clearly conveys the basic site information. At the site level map, a few meters accuracy is relatively unimportant. For the excavation level maps, the accuracy is reduced to millimeters. The transit-based surveyor, then, has generated three maps: the field sketch, the plot map, and the publishable map, with contours and mahlerized buildings.

The total station surveyor is in a different situation; he generates different products, and has a more comprehensive goal that provides data, for the digital, GIS analysis process. A hand drawn sketch map is not necessarily part of the recording process in the field, since survey software can easily generate the point plot map. Once the edits have been made, to the raw data file, the survey software converts the raw data to coordinate data, and the computer generates a point plot map. When doing this by hand the transit-based surveyor connects the appropriate points, to create features, both cultural and natural. The digital system can only plot points and must be told how to connect the points appropriately. This can be done with more or less ease, depending on the specific mapping software in use and the expertise and experience of the one, using the software. Point descriptions and connections have the greatest impact on the ease, with which data can be moved into a GIS program.

In order to get the computer to connect data points and draw cultural or natural features, the points in most cases, must be in a connect-the-dot order. The points are not shot in this order, however, as it is not convenient or efficient. If the descriptive codes are well defined and consistent, the software can often collect the correct groups of points, for any given feature, but the drawing order must still be arranged by manual input. Additionally, for any specific polygon, such as those that form part of a building, there must often be 5 points: one for each corner, and a fifth that replicates point one so that the connection can be made from point 4 to point 1. If the field data points are not satisfactory for the mahlerized version of the buildings, then new data points must be hand digitized on the computer. These drawing capabilities vary, according to software, but must be accounted for, in preparing to transport digital data to a GIS, for analysis.

Field generation of complex contour maps is an added feature of some digital mapping software, but digital contour generation raises its own problems and questions. At Chau Hiix, contour maps were generated at the end of each day. This provided a ready view of the terrain, places that needed more points shot, and errors in existing points, such as bad elevations, due to typos in instrument heights. However, contours are only useful for the central portions of each individual file, as the edges have no points of reference and tend to fall away rapidly. For a publication quality map, a lot of editing is required because it may be most useful to only show one or two strategic contours. The software cannot differentiate these, and plots only at regular intervals. The cartographer must delete each unwanted contour line. In addition, regular interval contours and line-polygon features, all on the same map, are difficult to read visually. In the 1997 season, the coordinate data were copied into two separate files: one for topographic points and one for feature points. This facilitated the ease, with which specific questions could be addressed, but created problems for file management and archiving. Since these were working files in the field, corrections, modifications, and refinements were constantly being made, by either the surveyors, the excavators, or the principle investigator. File naming and version coordination required constant attention. It is particularly essential at the end of the season, to be sure the field archive contains no more, than the final and essential versions.

A final word is in order, about the accuracy of the digital survey data, "the map that was understood to be of dubious quality on paper becomes inviolate once reproduced on screen," (Miller 1995, 321) and "perceived or affected accuracy can arise simply from the representation of an archaeological feature in the computer ..."(Harris & Lock 1995, 358). Point data from the total station is recorded to four decimal places. There is no consideration for the fact, that a given total station may only be accurate to a few centimeters, or that the surveyor of mounds may be making measurement judgements, that could be a meter or more off, in any given situation. Nonetheless, even with this level of error, the recorded point is still the most accurate representation, of the primary feature before its destruction. The 'truth' of four decimal places, as the data is carried into GIS and statistical analysis, systems should not be overrated.

Analyze: Once the geophysical data has been collected, the mahlerizations resolved, the cartography completed, then the intra- and inter-site features analysis can begin. This type of analysis, however, is dependent on the breadth of the archaeological record, and the interrelationships of the various aspects, that have been collected or documented. Identification or creation of the general structure, of the archaeological data itself, is not addressed in archaeological methodologies, except for a few comparisons of excavation reporting forms. The development of a generic data model, of an archaeological data bank, that includes non-textual primary data, and can coordinate seemingly disparate data sets, for the purpose of reconciliation of data in the query process, is essential to the future of the archaeological record in the digital age. The primacy of the data model, in data base design, is explained by Parker, Limp, & Farley (1985, p.91):

> An effective computerized system, however, can be derived only from an intensive examination of the general structure of archaeological data itself. Careful, top-down consideration of the nature of the archaeological data, must logically precede the design of the database to be preserved. The data base design, in turn, leads to the actual development of data processing software.

The advantage of a generic data model is its ability to be flexible. Its value is well stated in the business domain, "... generic data models conserve resources because organizations do not have to reinvent the wheel. Instead, organizations can use generic data models to increase the precision and availability of knowledge about the enterprise and to assist in uncovering areas of omission and commission. ... (and) can assist in communicating and integrating the different ... views ..." (Sanders, 1995, p. 99). As technology advances, new techniques are developed, as well as new ways to approach problems, and new problem domains. Prior to the use of the computer for data storage and analysis, a single investigator was responsible for all aspects of analysis of a given excavation: ceramics, lithics, human and faunal remains, geology, etc. New technologies and techniques, however, create more specialized analysis methodologies, in each sub-domain of archaeology. For example, GIS grew out of the land survey domain, but is applicable across a wide range of problems (Michalski, 1992). Today, in line with the 'new archaeology' (Hodder, 1985; Patrik, 1985; Schiffer, 1995), the principle investigator has their own specific problem domain, but cannot neglect the recovery of data, not directly linked to their problem. And, a specific excavation may involve multiple investigators, with multiple problem domains. The building of the archaeological record, for a given excavation, must therefore, accommodate the needs of each investigator. Thus, the data bank must reflect multiple problem domains and representational methodologies. This also facilitates the growing interest in regional analysis (Fish and Kowaleski, 1990), to allow inter-site comparisons of data. The generic data model, that organizes the data bank, has the potential of addressing the need for multiple views or multiple problem domains, rather than independent and disparate system designs. However, this introduces yet another, new specialty for the principle investigator, that of data manager. Knowing that the data bank and its model are the foundation of GIS, Allen, Green, & Zubrow, (1990, 384), comment that "despite their elegance and relative 'user friendliness', GIS do require significant study. Their potential rests to a great extent on the user's experience." A generic data model would facilitate inter-site analysis, and alleviate the problem of each investigator, re-inventing the data model wheel. Chau Hiix has a data model, available for sharing with any interested investigators.

Publish: Chau Hiix is still resolving many of the above problems, and will then publish a print document, that includes a CD-ROM of the digital data, as described above. The existence of a generic data model, and shared digital data, would facilitate the preservation and coordination of the archaeological record.

Bibliography

ADAP: Archaeological Data Archive Project. CSA, the Center for the Study of Architecture. Http://csaws.brynmawr.edu:443/web1/adap2.html

ADS: The Archaeological Data Service. http://ads.ahds.ac.uk/ahds

ALEXANDER, E. P. (1979), *Museums in motion.*

ALLEN, K.M.S., GREEN, S.W., & ZUBROW, E.B.W. (1990), *Interpreting space: GIS and archaeology*, Taylor & Francis, London

CARR, C. (1984), "The nature of Organization of intrasite archaeological records and spatial analytic approaches to their investigation", in: SCHIFFER, M. (ed.), *Advances in archaeological method and* *theory*, vol. 7 , Academic Press, Orlando, pp. 103-222.

CHENHALL, R. (1975), *Museum cataloging in the computer age,* American Association for State and Local History, Nashville

FAGAN, B. (1995), "Archaeology's dirty secret", *Archaeology 48*(4), pp14-17.

FISH, S. K. AND KOWALESKI, S. A. (1990). *The archaeology of regions: A case for full-coverage survey*, Smithsonian Institution Press, Washington, D.C.

FLOOD, J., JOHNSON, I., & SULLIVAN, S. (1989). *Sites and bytes: Recording aboriginal place in Australia,* Australian Government Publishing Service, Canberra.

GAINES, S. W. (1974), "Computer use at an archaeological field location", *American Antiquity 39*(3), pp. 454-462.

GARDIN, J-C. (1989), "Artificial intelligence and the future of semiotics: An archaeological perspective", *Semiotica 77*(1/3), pp.5-26.

HARRIS,T.M. & LOCK,G.R. (1995), "Toward an evaluation of GIS in European archaeology: the past, present and future of theory and applications", in: LOCK G. & STANCIC Z. (Eds.), *Archaeology and geographical information systems,* Taylor & Francis, London, pp.349-365.

HIRSCHHEIM, R., KLEIN, H. K.; & LYYTINEN, K. (1995), *Information systems development and data modeling: Conceptual and philosophical foundations,*: Cambridge University Press, Cambridge.

HODDER, I. (1985) "Postprocessual archeology", in: SCHIFFER M. (ed.), *Advances in archaeological method and theory, vol. 8,* Academic Press, Orlando, pp. 1-22.

KENWORTHY, M.A. (1985), *Preserving field records: Archival techniques for archaeologists and anthropologists,* University Museum, Philadelphia.

McMANUS, J.M. (1997), *Geographic information system applications for data input, pattern recognition, and data visualization in archaeology*, Master's thesis, North Caroline State University Department of Parks, Recreation, and Tourism Management, Raleigh, NC.

McMILLON, B. (1991), *The archaeology handbook: A field manual and resource guide,* John Wiley and Sons, New York.

MICHALSKI, R., & RADERMACHER, F.J. (1992), "Challenges for Information Systems: Representation, Modeling, and Metaknowledge", in: SCHADER M. (ed.), *Analyzing and modeling data and knowledge*, Proceedings of the 15th Annual Conference of the "Gesellschaft fur Klassifikation e. V.", Springer-Verlag, Berlin, pp. 199-216.

MILLER, P. (1995), "Thoughts on the design and interpretation of an archaeological GIS", in: LOCK G. & STANČIČ Z. (eds.), *Archaeology and geographical information systems,* Taylor & Francis, London, pp.319-333.

NICKERSON, S. (1997, September) Available: http://nickerson.icomos.org/steve/ [1998, June 1]

P.O.B. Point of Beginning. Post Office Box 3221, Northbrook, IL 60065. Business News Publishing Company.

PARKER, S., LIMP, W. F., AND FARLEY, J. A. (1985), "Archaeological data base management as conceptual process", in: CARR C. (ed.), *For concordance in archaeological analysis*, Westport, Kansas City, pp. 91-113.

PATRIK, L E. (1985), "Is there an archaeological record?", in: SCHIFFER M. (ed.), *Advances in archaeological method and theory, vol. 8,* Academic Press, Orlando, pp. 27-62.

RICE, P. M. (1996), "Recent ceramic analysis: 2. Composition, production and theory", *Journal of Archaeological Research 4*(3), pp.165-202.

QUINN, B. (1994), Recent theoretical approaches in classification and indexing", *Knowledge Organization,* 21(3), pp.140-147.

SANDERS, G. L. (1995). *Data modeling,* Boyd & Fraser, Danvers, MA.

SCHOLTZ, S., & CHENHALL, R. (1976), "Archaeological data banks in theory and practice", *American Antiquity 41*(1), pp.89-96.

SHARER, R, & ASHMORE, W. (1987), *Archaeology: Discovering our past,* Mayfield Publishing, Mountain View, CA.

SHIFFER, M. (1987), *Formation Processes of the archaeological record,* University of New Mexico Press, Albuquerque.

SILVERMAN, S., & PAREZO, N. (eds.) (1995), *Preserving the anthropological record,* Wenner-Gren, New York, 2 ed.

ZUBROW, E.B.W. (1990), "Contemplating space: a commentary on theory", in: ALLEN, K.M., GREEN, S.W., & ZUBROW, E.B.W. (eds.). *Interpreting space: GIS and archaeology*, Taylor & Francis, London, pp.67-72.