Notes for Paper 566

An ArcGIS Data Model for Archaeologists: Problems and Prospects

Ryan S. Arp
Brandeis University

Abstract

This paper discusses issues, advantages and disadvantages of creating a data model for archaeological fieldwork and analysis focused on the regional, inter-site and intra-site scales. The creation of such a model entails the awareness of issues involving scale and time. In addition, anthropological notions of settlement, household and territory are addressed in that they pose fundamental problems in both understanding and representing spatio-temporal data. The implementation of an archaeological data model could be of great utility for researchers in all phases of their project: from inception to dissemination.

Introduction

With the advent of Object-oriented data modeling (Rumbaugh et. al., 1999), there has been an insurgence in the creation of data models for various disciplines and industries (ESRI, 2003; MacDonald, 2001: 259-325). Creation of an archaeological data model for the use in Geographic Information Systems (GIS) has begun in archaeology most notably in Historic Preservation and Cultural Resource Management (ACC, 2000; Konopka, et. al., 2002 for example). Academic archaeology, on the other hand, has been slow in response to the demand for an archaeological data model. While part of this slowness to adopt a data model for archaeology may be blamed on awareness, another possibility lies in the use of mutually agreed upon analytical terms, since a data model implicitly accepts the use of analytical, or etic, terms and concepts. What exactly should be included in an archaeological data model and the relations between data sets are of concern here. This paper assumes that the archaeologists adopting a data model for their research are amenable to *some* type of data standards and that they are open to data sharing. Additionally, subsequent interpretations of archaeological data that utilize under a clear, concise data model have many advantages.

In order to create an archaeological data model for GIS data it is necessary to perform five steps as outlined in Zeiler (1999). The five steps are as follows:

1. Model the user's view of the data.
2. Define objects and relationships.
3. Select geographic representation.
4. Match to Geodatabase elements.
5. Organize Geodatabase structures.

This paper is concerned only with Step 1 (Zeiler 1999: 184-185). It is only after we model the user's view of the data can we even begin to define objects and their relationships and complete steps 2 through 4. One of the issues surrounding the creation of a data model for archaeological fieldwork and analysis focused on various scales is to create analytical units. I argue data modeling efforts become more complex as we move from measuring units to more conceptual units.

In looking at the user's view of the data we immediately see that the archaeologists working with GIS's differ in how they approach their fieldwork. For example, those privileging a site-based approach ends up with a different data model than those using a landscape-based approach (Barker & Mattingly, 1999). Theoretical concerns aside, we are constrained by the use of data structures that are available in ArcGIS. This means we are only able to model in a raster, vector (points, lines and polygon), and Triangular Irregular Network (TIN) data formats. Thus it is very important how we choose to conceptualize our archaeological data and how we model the relations between data structures.

Advantages and Disadvantages

Advantages of an archaeological data models are numerous. First, one of the clearest advantages is that end-users would have a common template from which to start their GIS-based research. Archaeologists wishing to put their existing data into a GIS would have at the very least a starting point from which to work. This could only be possible if a data model were constructed on a very basic level. The reason that the accepted model would have to be very simple is that differing research designs require different measuring units and differing demands on data collected in the field.

Second, an archaeological data model encourages data sharing. At the very least, GIS-based archaeological data could be shared because they are modeled in the same way. Conversion of multiple sources of data collected at various scales may be problematic.

Third, another advantage of the data model is our assumptions about relations between data would be explicit to the end user (via associated metadata and documentation). Making the assumptions and relationships between data explicit would tacitly place constraints on what you could do with your GIS data. Thus, explicit data models would limit the types of analysis and interpretations that could be drawn from a specific data set. This point is stressed in research that is utilizing GIS for analytical purposes.

Fourth, an archaeological data model could allow for consistent data transformations and aggregations of data by scale and time. This can only happen if Step 1 is completed. Once data structures are agreed upon along with their proper scales we would also have to come up with ways in which to aggregate and/or transform the data in order to perform either local and/or higher orders of analysis. Specifically, I am thinking of data that is collected at the site level and aggregated as a sample to the regional level where the data can be compared with other regions.

Finally, we need an explicit data model(s) in order to perform consistent GIS-based analysis. Examples of spatial data analysis range from ESDA to Bayesian approaches (Fotheringham et. al., 2002; Longley & Batty, 1996; Robertson, 1999). All require either vector, raster, or TIN data structures. If data is collected in the field due to constraints of collection (GPS units, drawings, etc.) some of the data may need to be transformed and/or generalized/ re-sampled to appropriate data types or to appropriate scales. An example would be how data is often collected by survey transect but is often aggregated to a polygon's centroid and analyzed from there.

While the advantages of an archaeological data model is somewhat obvious, we need to look at the disadvantages of using a single data model versus the use of a very general, adaptive model that archaeologists can modify and/or amend. If we slip into the use and acceptance of one data model we may unknowingly accept universally accepted notions of anthropological institutions. A few examples of institutions that are under debate follow. First, let's turn to common spatial data types that are basic measurement units from which archaeologists systematically work (with the understanding that this only represents *some* of the possibilities utilized in archaeological research).

The User's View of the Data: Etic Archaeological Spatial Data Types

In this part I highlight some archeological spatial measurement units that can currently be put into a GIS with relative ease.

Datum

Datums measure the "0.0" point on an arbitrary excavation grid. Some sites, with multiple operations require multiple datum. These are usually measured as points. In creating a grid in a GIS, the datum can be tied to a script that uses the datum's x, y coordinate from which to lie out a vector-based excavation grid (An example of this functionality can be found in the script gridmaker.ave developed for ArcView 3.x.).

Excavation Units

Excavation units are usually arranged into a grid composed of squamous units usually measured in meters. Data can be aggregated into these excavation units in a database that records vertical provenience and SQL-based queries are able to elicit information by stratum, depth below datum, artifact type, and artifact density. Since stratums are almost never evenly deposited it may be necessary to model the stratigraphic units as a TIN data structure in order to get better volume measurements. As you can see, the choices between data types pose specific challenges to the GIS modeler in that they will have to adapt a particular model that suits their needs.

Multiunits

Multiunits, composed of an archaeological feature (soil stain, burial, etc.) may span many excavation units and puts many demands on archaeological data in that it is measured both horizontally and vertically. First it is necessary to map the boundaries of the multi-unit, which varies in size and depth (which inevitably sets up many more measurement demands on the fieldworker). For this reason, it may be necessary to model this data type in a TIN environment.

Features

Another data type that is often measured is an archaeological feature. In my experience, these are usually drawn in the field in sketch maps, and digitized or scanned for use in a GIS modeled either as a raster (bitmap or jpeg) or a digitized vector drawing (points, lines or polygons). Attributes are usually a descriptive type and its characteristics. The choices, then, are many and the relations between data types would need to be established via a data model.

Survey Transects

Survey transects on the other hand vary in size and shape. Most often they are polygonal in nature and can extend to very long thin polygons to systematically spaced polygonal grid. Some studies have used transects as their primary measurement unit while other surveys aggregate their data to an archaeological site (usually a point on a map).

Usually, there is a move from the recording units (excavation unit, multi-unit) to site, settlement, etc. It is this move I am concerned with, as I will hint by focusing on the aggregation of data and its implications for an archaeological data model.

Site

The use of archaeological site varies according to each archaeological study. Some studies (Kuna, 2000: 31) reject the notion of site. Rather they use a series of sampling units from which to construct the relative frequencies of artifacts found over a given area. However we conceive of the anthropological notion of site, we can conceive the data as either point, polygon with either the extent of the site or sampling units recording number of artifacts per unit when maybe a raster a raster may be more suited to modeling palimpsests. An implication for this is that we need to create models that are able to utilize both raster and vector data structures.

Now that we have laid out some of the basic etic data structures used most often in archaeological excavation, the question is raised on how we bring into higher-order analytical spatial structures as contributing to, for example: Settlement, Household, Region and State. While this by no means represents the range of etic spatial entities, they represent a sample of terms that are considered loosely as etic by the discipline but to this day have not had clear definitions. In addition to the example of what is meant by site, examples can be found when examining by what is meant by Settlement and Household: two core units of analysis when examining archaeological data.

Settlement

Settlement is always spatially bounded. Settlements may refer to spatially bounded areas where daily activities occur. Communities, as in the cases with the State, are sometimes highly correlated with these spatial arrangements. In some cases, they are not. Instead, communities may be based on age, sex, ethnicity or experience. Thus, communities do not always require settlements or spatially distinct territories. Given this definition we can see modeling a settlement can be rather complex.

Household

In defining household, two schools of thought concerning the utility of the concept exist. The first position rejects the concept of household, seeking to refine what is meant by the concept. Bender (1967) asserts households encompass a variety of concepts that need to be delimited and the lack of distinction that surrounds these terms has lead to confusion within household studies. Instead of examining households, we should be examining three distinct terms that are mistakenly thought to be part of household. They are family, co-residence and domestic functions. Following Bender, family does not always carry out economic functions together due to the fact that economic functions often take place outside of the family. Thus, residence is based on aggregates of humans rather than the presence of permanent dwellings. Co-residence refers to “living together, which is minimally characterized by a proximity of sleeping arrangements and a sentiment similar to that expressed by our own folk concept of home”(Bender 1967: 498). Domestic functions are concerned with the day-to-day activities that are necessary for living. This includes provisioning, food preparation, and childcare and is carried out by small groups—most frequently by co-residential families.

On the other hand others (Netting et. al., 1984) disagree with Bender’s redefinition of the household concept. They state that households are more conceptual units but are more cross-culturally comparable than many more frequently studied institutions. Households are task-oriented residence units composed of household members where location, shared activities and kinship need not overlap. Most people, they state, live in households that are sometimes based on kinship, descent, economic operation, reproduction, or socialization. Under this conception, we have multiple requirements for households that would have to be made explicit by the researcher and ethnographic area.

As you can see examining just two of these etic spatial entities requires us to refine what we mean by other etic terms that may or may not be spatially bounded. Clearly, more work needs to be done on the conceptual level as we move from measurement units to more abstract social entities.

Issues of Scale and Time

The Modifiable Areal Unit Problem (MAUP) has been very pervasive in sociology, geography and the environmental sciences (Openshaw, 1984; Tate & Atkinson, 2002). Basically this problem refers issues of scale and zoning effects. In looking at data that is collected at one scale and aggregated to other units at a different scale we will have to deal with the MAUP. Thus, we may need to ask what are the rules of aggregation that will be applied to data that is, say, collected from a survey?

An anthropological example that questions sampling at a broad scale may be found in Barth's work in *Balinese Worlds*. The earlier accounts would say amongst the Balinese, etc. without realizing the moasicked nature of each part of the Island. We, thus, could be in the same situation when we model--assuming that all areas are homogenous by our assignment of rules between sets from settlements to regions. We need to be careful here too in that such simple aggregations can cloud out heterogeneity thereby denying us insights on factions, groups, etc. (Brumfiel, 1992). Rules of aggregation can include rules that have to do with relationships between data. While I have no hopes of solving this problem it is something we need to be aware of when creating data models.

When creating our measurement units (excavation unit, survey transects, etc.) we need to decide how to relate lower-order units to higher-order units such as region, household, community, etc. In addition, we may need to create additional data that can be derived from ethnohistoric or historic documentation and maps of these areas as they relate to bounded spatial entities (administrative units, households, and properties).

Time

While time is something that is currently being addressed by scholars working with Temporal GIS (TGIS) applications (Johnson, 2000; Wheatley & Gillings, 2002: 242-3), it is a factor that we need to deal with in our archaeological data model. The point I want to make here is that time will become an important issue when creating relationships between datasets—depending on the data that is available to the researcher (for example, geomorphological and climatic data).

Final Comments

If we are to create a data model we must keep in mind issues of scale and time. The creation of data models demands truly etic definitions that should be understood by those who choose to use and accept the data model. Those that fail to accept (or at least understand) the accepted definitions run the risk of misunderstanding subsequent GIS-based models and interpretations of GIS-based analyses. As it stands it looks as if we *can* model some spatial archaeological units consistently.

While the use of mutually agreed upon terms is being used in current professional archaeological applications, we need to realize our models in the light that they should be not only spatial data management systems but as analytical systems where spatial units are able to be analyzed confidently and rigorously. Furthermore, when we move from measurement units to higher order units we realize we will have to be very careful how we create relationships between spatial data sets. An archaeological data model that is adaptive in nature can be of great utility to archaeologists interested in various types of spatial analysis at various spatial and temporal scales. That being said, the prospects for such an archaeological model are many but only after serious attention to the problems surrounding the creation of a *single* data model are given.

Acknowledgements

I would like to thank the participants of the ESRI Archaeology Discussion Conference, Bob Booth of ESRI and Sally Wyman of the Brandeis GIS Services for comments on my paper ideas, and the Brandeis Department of Anthropology and Graduate School of Arts and Sciences for covering travel expenses.

References

ACC (Air Force Air Combat Command located at Langley AFB)
2000 Cultural Resources Geospatial Data Integration. URL: http://tsc.wes.army.mil/contacts/groups/FWG/Natural-Cultural/edit/proposed_cultural_entity_set.htm

Barker, Graeme and David Mattingly
1999 *Geographical Information Systems and Landscape Archaeology*. Oxbow Books: Oxford.

Barth, F.
1993 *Balinese Worlds*. University of Chicago Press: Chicago.

Bender, Donald
1967 A redefinement of the concept of household: families, coresidence, and domestic function. *American Anthropologist* 69: 493-504.

Brumfiel, Elizabeth M.
1992 Distinguished Lecture in Archaeology: Breaking and Entering the Ecosystem Gender, Class, and Faction Steal the Show. *American Anthropologist* 94:551-567.

ESRI (Environmental Systems Research Institute
2003 Data Models Homepage. URL: http://support.esri.com/index.cfm?fa=downloads.dataModels.intro

Fotheringham, A.S., Chris Brunsdon, and Martin Charlton
2002 *Geographically Weighted Regression: the analysis of spatially varying relationships*. John Wiley & Sons, Ltd.: Hoboken, NJ.

Johnson, Ian
1999 The TimeMap Project: Developing Time-based GIS Display for Cultural Data. URL: http://www.timemap.net/documents/publications/2000_j_gis_arch_esri/timemap_article/index.html

Knopka, E., K.E. Clark, and L. L. Wiggins
2001 What's Eligible Here? Paper presented at 2001 ESRI User Conference, San Diego, CA.

Kuna, M.
2000 Surface Artefact Studies in the Czech Republic. In *The Future of Artefact Survey in Europe*. Edited by J. Blintiff, M. Kuna and N. Venclová. Sheffield Academic Press: Sheffield.

Longley, P. and M. Batty
1996 *Spatial Analysis: Modelling in a GIS Environment*. GeoInformation, Int'l.: Cambridge.

MacDonald, Andrew
2001 *Building a Geodatabase*. ESRI Press: Redlands, CA.

Netting, Robert McC., Richard Wilk and Eric Arnold
1984 *Households.* University of California Press, Berkeley.

Openshaw, S.
1984 'The modifiable areal unit problem', *Concepts and Techniques in Modern Geography* 38: 41.

Robertson, I.
1999 Spatial and Multivariate Analysis, Random Sampling Error, and Analytical Noise: Empirical Bayesian Methods at Teotihuacan, Mexico. *American Antiquity* 64(1)

Rumbagugh, James, Ivar Jacobson and Grady Booch
1999 *The Unified Modeling Language Reference Manual*. Addison-Wesley: Reading, MA.

Tate, N. J. and P.M. Atkinson (eds.)
2002 *Modelling Scale in Geographical Information Science.* John Wiley & Sons: New York.

Wachowicz, Monica
1999 *Object-Oriented Design for Temporal GIS.* Taylor & Francis: New York.

Wheatley, David and Mark Gillings
2002 *Spatial Technology and Archaeology.* Taylor & Francis: New York.

Zeiler, Michael
1999 *Modeling Our World: The ESRI Guide to Geodatabase Design*. ESRI Press: Redlands.

Author Information

Ryan S. Arp

GIS Graduate Assistant
Brandeis University GIS Laboratory
MS, 070 Gerstenzang Science Library
415 South Street
Waltham, MA 02454
(781) 736-4722
arp@brandeis.edu