Jonathan Corson-Rikert and Denis White Laboratory for Computer Graphics and Spatial Analysis Graduate School of Design, Harvard University 48 Quincy Street Cambridge, Massachusetts 02138

COMPREHENSIVE MAP INPUT SOFTWARE: Building on the Minute Man National Historical Park Data Base Project

ABSTRACT

The Minute Man National Historical Park geographic data base comprises several types of spatially distributed information about the Park and its surroundings. Soil types, elevation contours, floodplains, property boundaries, and land use/land cover data for 1984 and 1775 were digitized in vector format using the ODYSSEY Geographic Information System. Data were converted to raster format for various planning analyses using the Map Analysis Package.

The project raised the issues of data resolution, accuracy, and volume, and the relative priority of the different information types when combining data for analysis and display. The project has also motivated ongoing design and implementation of new software with a more interactive error finding approach to geographic data base construction. The software features fully graphic entry and editing of geometric and attribute data with "smart" capabilities to develop and maintain full topological consistency. Interfaces to scanning hardware and software, output to the ODYSSEY Geographic Information System and common polygon-based microcomputer mapping packages, and conversion to raster format are included in the software. Improved capabilities for updating, attribute data management, and three-dimensional displays are discussed in the context of applications in education and government.

INTRODUCTION

In the fall of 1984 the National Park Service requested that a Landscape Architecture studio group under the direction of Professor Carl Steinitz at the Graduate School of Design investigate alternative future development options for the Minute Man National Historical Park (Department of Landscape Architecture 1985). A geographic data base begun during that studio was expanded by the Laboratory for Computer Graphics under contract with the National Park Service North Atlantic Region, Office of Planning and Design.

The park itself is located twelve miles west of downtown Boston, in the towns of Concord, Lincoln, and Lexington, just west of the Route 128 circumferential highway. The park is approximately six miles in length and includes over 700 acres. It is divided into three sections, with visitor services and historic sites split along the historic Battle Road that provides the central spine of the park. The geographic data base includes a much larger area, since nearby land uses affecting the park include the Hanscomb Air Field, a regional vocational-technical school, and the rapidly growing industrial development areas bordering Route 128. Residential development bordering the park in each town has also been very active.

The data base was developed using the Laboratory's ODYSSEY digitizing module and tolerance-controlled error detection capabilities. Initial line work for each thematic overlay was converted by ODYSSEY to topologically structured chains which are checked for consistency and completeness. The data were then converted by ODYSSEY to polygon format for transfer to the SAGIS system software at the Denver Service Center of the National Park Service.

During the Landscape Architecture studio itself overlays of elevation and 1984 and 1775 land use were also converted to raster format and analyzed using the Map Analysis Package, written by Professor Dana Tomlin of the Department of Landscape Architecure at the GSD (Tomlin 1983, Tomlin 1985). Additional software written by Denis White, Associate Director of the Laboratory, and other members of the Lab was used for the vector to raster conversion itself and for display of relief shaded thematic maps (White 1984). The software performed very well with the large data files involved (as many as 60,000 vertices per map). A number of new features which would facilitate the digitizing and editing process were identified, however. Many of these features utilize software techniques and hardware capabilities which were not available at the time the ODYSSEY system was designed in 1975.

Since the completion of the project, the authors have developed new map input and editing software, known currently as TRACE. The current implementation is on an IBM PC/AT, with versions planned for the Apple Macintosh and the Digital Equipment Corporation Microvax II. New features and issues associated with projected features of that software are discussed in the latter part of this paper.

THE MINUTE MAN NATIONAL PARK DATA BASE

Content and Organization

The data base was from the beginning conceived for planning rather than engineering or cadastral purposes. This distinction influenced much of the methodology for data base construction. Neither the data available nor the anticipated uses would have justified the additional expense and time required to build an engineering or cadastral data base. The data base is being used for park-wide planning rather than site-level design work and does not serve as an archive for any legal boundary information. Adequate sources existed in paper form, and the ability to compare different thematic overlays with confidence in the general consistency of each was rated a higher priority than high accuracy for any single coverage.

Standard 1:25000 USGS quadrangle sheets were chosen for an underlying base map. A double-line street base map was digitized from 1"=400' enlargements of two quadrangle sheets. Once the full base map was available in digital form, plots were produced at the scale of the other source maps, ranging from 1"=1000' to 1"=200'. Source data could then be traced onto the plot while using the context of the source map to correct for geometric inconsistency.

The USGS quadrangle sheets served as sources for elevation contours, water bodies, and structures (with updates from aerial photographs). Floodplain designations were digitized by the method described above from the standard FEMA maps at 1"=1000'. Soil types were generalized and hand-compiled to the 1"=400' base map by the Landscape Architecture students.

The project also included land use/land cover data from both 1775 and 1984. The historical data was digitized from maps compiled by historian Dr. Joyce Malcolm from records at the Middlesex County registry of deeds (Malcolm 1985). Wherever possible the land use, structures, and vegetation cover were compiled from property records of the time. Aerial photography and local survey maps were used to compile a similar land use/land cover map for 1984.

The individual town map series were used for compiling land ownership categories and parcel boundaries for the study area. Here the discrepancies between the many small local map sheets were readily evident, and even transfer to a same-scale plot required some interpretation and adjustment. The overlay was designed to show private land versus several categories of public land, as an overlay of jurisdictions affecting the area immediately surrounding the park. With the outline of each parcel present, the data base can be cross-checked with individual local map sheets for parcel numbering and ownership information.

Software and Methodology

The Minuteman Project was the first major data base built with the currently distributed

version of ODYSSEY at the Graduate School of Design. The software performed remarkably well with the large data files, and with few exceptions provided the necessary tools for managing and manipulating the data.

The chain structure. The fundamental data unit for ODYSSEY is the chain, made up of a starting node point and ending node point connected by one or more intermediate line segments. Each chain has a header record containing the number of points in the chain, a chain identifier, starting and ending node numbers, and left and right polygon ids. The chain structure has significant advantages for data storage over the more common polygon structure used for many display mapping programs. Common boundaries between areas need be stored only once, which saves storage and eliminates discrepancies between alternative versions of the same line. The chain structure facilitates selection, aggregation, polygon overlay processing and other analytic capabilities. The topology can be verified for completeness, and the polygon identifiers together with node points can be converted by ODYSSEY to a full polygon format when needed for shaded polygon display.

Full manual coding of chain id's, start and end nodes, and left and right polygons is a tedious and error-prone task. ODYSSEY permits digitizing lines by themselves without topology and without regard to the final assemblage of line segments into chains. Duplicate segments, overshoots, undershoots and intersections are then detected using a user-specified tolerance value in a separate verification program. When an input line file (referred to as spaghetti) has been checked and edited, the file is processed with a file of point identifiers (called meatballs) which are used to associate the user's polygon values with the coordinates of each chain. Typically with a large file some further editing is required to add missing id's or remove duplicate or conflicting identifiers for polygons. Once the polygons ids have been added to the chains the chain file is complete.

Ground control. Control point files were developed using the state plane coordinate grid ticks on the USGS quadrangle sheets. While these control points were not accurate enough for engineering work, they did enable us to produce map files in state plane coordinates, which could easily be compared with more accurate sources at individual sites later in the park planning stages.

<u>Tolerance</u>. A standard tolerance of 10 feet was used for the programs which convert line input and polygon id points into complete topological chain files. The tolerance level was set to enforce careful digitizing without overstating the accuracy of the map sources.

Once an overlay had been entered in line form, the file would be submitted for a first pass at "line conversion" -- elimination of duplicates, merging or flagging gaps and overlaps based on the tolerance value, and construction of chains. Most files required editing to correct errors greater than the tolerance, and this editing was normally done on the graphics terminal with input crosshairs rather than on the digitizing tablet.

<u>Tutorials</u>. The ODYSSEY system has a flexible command file facility that we have found very helpful for both novice and experienced users. Command files were originally conceived as macros, or groups of instructions to be run in sequence from a file for batch operation or program initialization. In addition to program commands, a tutorial can include comments and instructions while leaving the user free to issue commands and control program execution as he wishes (Brockhuysen and Dutton 1983).

The tutorial itself is a simple text file with standard ODYSSEY commands (e.g. "open input file; draw map;") with additional symbols that control pausing, skipping, indexing, and echoing. At pause points the user can execute commands completely independently of the tutorial or can branch to other parts of the tutorial. Alternatively, command execution can be hidden from the user or observer, to facilitate demo examples.

Tutorials were most useful to the Minute Man project for training new digitizer operators, for controlling the background commands that establish standard control point and window environments, and for managing the sequencing of commands for digitizing and plotting. ODYSSEY uses sequential files that allow only front to back traversal of the map file; with the tutorial a more random-access editing environment could be simulated by repeated cycling through

the current file being edited.

Updating. Replacing existing data with new data for the same area of a thematic overlay is more complex in a topologically complete data structure than with a simple line file, especially if changes to the topology as well as geometric changes are required. Updating chains is far easier than a polygon structure, however, where each common boundary is stored twice within the coordinate lists of the two bounding polygons. Since it is virtually impossible to duplicate a correction manually, polygon software must either ignore sliver gaps and overlaps or offer tools to extract and copy sections of polygon boundaries.

It is often more efficient with ODYSSEY tools to revert to the simple line file and regenerate a revised chain file with full topology after editing the geometry. Two possible editing strategies can then be considered. One approach is to make changes one by one, using the operator's knowledge of the map context to ensure accuracy and completeness. Alternatively, one can devise tests that can determine where change occurred and make systematic corrections, using chain left and right polygon identifiers or other information. The ODYSSEY polygon overlay processing can also be used to identify areas of change between two files, using cross-reference files that show the history of each polygon created. When procedural methods for editing data rather than direct change are involved, care must be taken that all the corrections and only the intended corrections have been made.

Conversion to raster format. One critical issue in the planning and design of Minute Man National Historical Park is the visual quality consequences of any restoration of the historical landscape. In recreating the generally cleared fields and pastures of 1775, the multi-family housing, office parks, and utility structures of 1984 might easily obstruct the visitor's appreciation of the historical conditions. The Map Analysis Package has the the flexibility and capability for studying these problems by spatial analyses such as viewshed determination. For this reason we converted the ODYSSEY created vector data base to MAP's raster format.

The conversion to raster format made use of a modification of part of the basic ODYSSEY WHIRLPOOL algorithm. The program MATRIX, written by James Dougenik, processes chains in the usual ODYSSEY sort order and generates successive rows of raster cells as it moves through the chain file. Because it uses the polygon value at the center point of the cell to assign the cell's value, a supplementary program was written to allow area-weighted, averaged, or priority-based cell assignment. The land use/land cover overlays used priority-based assignment and the elevation data base used averaged assignment. In addition to the smoothing of the elevation contours achieved by averaging the elevations within a cell, a binomial smoothing algorithm was applied to the second order neighborhood about each cell.

As an example of the ensuing analysis in the Map Analysis Package, we created a MAP overlay consisting of the 1984 land use/land cover in the area surrounding the park and the 1775 landscape inside the park to simulate a reconstruction of the historical landscape. With the assistance of the studio designers and planners we then identified fifteen locations of visually obtrusive contemporary structures. By radiating lines of sight from the tops of these structures over the land cover (including the effect of elevation) an overlay of all points visible from one or more of these structures was created. As expected, the open areas in the "reconstructed" landscape of the park were mostly visible, implying that the structures in turn could be seen from the open areas of the park. Reconstruction could emphasize modern intrusions rather than restore the colonial landscape if not undertaken selectively and carefully.

NEW SOFTWARE TOOLS

In the course of the project we identified several major improvements to incorporate in the next generation of software beyond ODYSSEY, which we have now implemented for map input. Many of the changes reflect generally higher expectations for an interactive graphic environment, especially since the advent of programs with the fluidity and accessibility of the Apple

Macintosh-style graphic interface.

A second and ongoing goal is to provide a set of tools that utilize advances in hardware and software to make geographic information system processing more accessible and managable for government and educational organizations. Expanded hardware capabilities combined with competitive pricing strategies are bringing more sophisticated computer systems to smaller agencies, regional and local offices, and academic departments. GIS software has not yet adapted to these advances either by similar cost reductions or by fully using the capabilities available, however. Experience gained at the Laboratory with GIS implementations to date indicates that current approaches to data structures and project organization may not be adequate (White, 1985).

The remainder of this paper will discuss several current and proposed software features under implementation at the Laboratory. These features are not necessarily new or unique, but in aggregate the advances in capabilities and ease of use are significant. A number of questions on future directions for software are also raised.

Major areas of software development effort at the Laboratory are 1) facilitating map input by improving interactive graphic control and automated topological verification; 2) managing data consistency and priority; 3) accommodating scanning hardware and software; 4) providing multiple forms of cartographic output; 5) facilitating ready conversion between vector and raster data formats; 6) managing multiple data attributes and attribute-based queries; 7) designing data structures to accommodate flexible construction and updating across multiple layers; and 8) moving into the third dimension to bridge the gap between CAAD (Computer-Aided Architectural Design) and GIS.

Graphic Feedback. Continuous graphic feedback is the most immediate improvement we have implemented. Our current workstation is a minimal configuration of an IBM PC (XT or AT) with a graphics screen (monochrome or color) and a simple 11" by 11" digitizing tablet with a pen cursor. The map input program is controlled 99% from the tablet, with keyboard entry required only for file names and attributes. The current position of the cursor is echoed to the screen so that the location of new points or lines can be seen in relation to existing data <u>before</u> actual entry, by "rubberbanding". Any changes in the data file are echoed continuously to the screen so that the user feels confident that he or she knows and controls the process of map entry rather than working blind. Editing commands are located on the same menu as digitizing commands to facilitate correction of errors as soon as they are recognized, using the familiarity of the operator with the data at the time of entry. Windows on the data base are controlled graphically also so that the user can pan or zoom in on specific areas with maximum clarity.

This philosophy of "what you see is what you get" has become a standard for word-processing and spreadsheet programs. Earlier text editors that used embedded formatting commands needed to be submitted to a processer for analysis, error reporting, and generation of an output file in the desired text format. ODYSSEY and other mapping systems still similarly require graphic data to be checked and edited in separate processes before a file is complete. The transition to "smart" map input can eliminate the separate error-processing step while greatly improving visual feedback to the user, just as "smart" text editors for pascal and other programming languages can find errors in syntax as the program is typed in. (Teitelbaum and Reps, 1985).

Continuous Structuring. A digital map needs to contain more information than the coordinates needed to reproduce the graphic image (see the discussion of chain topology in 'Software and Methodology' above). Building the required topological structure on the fly ensures file consistency so that the cartographic file is complete and correct when the initial digitizing has been finished. The user needs only minimal familiarity with topological structures, and is required to provide only the necessary attribute information to properly code output polygons or lines as desired. Using a tolerance sensitivity optionally controlled by the operator, the software checks for point convergence, line intersections and line duplications. Polygons are assembled and split as lines are added without any user interaction. Polygon labels are added

when the lines are complete, with each polygon lighted up individually and an attribute requested. Polygon labels are applied to the chains as left or right identifiers, and a center line attributes may be added as an option for selective display or as an alternative to polygon output.

The topological relationships are also maintained during editing. If a point is moved, all lines connected to it move. If a new line crosses an existing line, the topology of both lines and the adjacent polygons are corrected to reflect the addition. The user works with the graphic image and the software maintains the structure; when graphic editing is complete the user requests an update of attributes that flags polygons or lines without attributes or checks for a desired attribute value.

Priority. One area where we have found ODYSSEY's control structure lacking is in the management of spatial priority. To maintain data base consistency it is important that new data and changes to existing data should not inadvertently lower the quality of existing data in a file. Similarly, when combining two data layers by overlay analysis or when merging two adjoining files, higher quality data should retain precedence over data known to be less accurate. If a coordinate has been established and checked against its source, and the source has a known accuracy level, then that data should not be arbitrarily changed.

In ODYSSEY any points within the current tolerance distance of each other are merged during the data base verification step, when raw line files are converted to topologically complete chain files. The merging method assigns as a representative of a cluster of points within the given tolerance setting that single point whose distance to all other points of the cluster is least. While this method is reasonable if there is no a priori knowledge of data quality, the user cannot implicitly or explicitly hold one value fast and merge all other close values to it. Control over spatial priority is lost, since a well-established point may be moved to a new location anywhere within the tolerance distance.

When several overlays are involved that share common points, such as a study area boundary or road network, slight meaningless sliver inconsistencies can develop between all the alternate versions of the same lines. In most cases the discrepancies are not noticeable, but deviations along a straight map edge can become quite visible and require corrective editing.

One solution could assign to every point a measure of its expected accuracy, so that more accurate information would always have precedence over sketchy or generalized map coordinates. The accuracy measure would serve as a separate tolerance parameter for each point, limiting the distance any point could be moved to accommodate another close value. The extra storage requirements of such an implementation would have to be weighed against the importance of data consistency and control for a project. Explicit priority management is most important for data of legal or archival significance and for data bases including overlays of widely varying quality.

We have implemented a simpler, implicit solution in our new digitizing software. Any point in the file is stored only once, either as a vertex along a chain or as a node if more than one chain intersect at it. The first instance has priority, and all new points are checked against all existing points within the current coordinate window. If a new point is within the tolerance distance, a reference is made to the first point occurrence rather than creating a duplicate of the point. New data always has a lower priority than data existing in the file. No explicit priority coding is involved, and no additional storage is necessary. The system will not help with the overlay of two files, however, since no tolerance value is carried with the file.

Data Partition. The additional processing load of checking new input against all data in the file can become burdensome, especially on the personal computer versions. Consequently we are currently designing a method for adaptive data partitioning that will come into effect when the the number of points and lines exceeds a threshold value. Searches will subsequently be limited to a currently active primary area and adjoining areas as lines extend outside that area.

For very large databases more complex data partitioning methods may be required (Keating, 1986). Several current geographic information systems are experimenting quadtree structures for data queries, data storage, or both, such as quadtrees (Samet et. al., 1985) or TID's (Scott and Iyengar, 1986).

Scanning. The potential advantages of map scanning over manual digitizing have been

evident for several years. Some expensive systems are already in use, particularly in national mapping agencies with huge processing loads (Moreland, 1986). Even more promising is the advent of at least one low-cost hardware and software system that could make scanning a viable option in academic and smaller government settings.

In the Boston area several firms can provide scanned files in raster format from map sheets ranging up to three or four feet on a side, typically using rotating drum scanners. The resulting files require significant processing both to clean up the raster image (principally line thinning and text removal) and for conversion to a vector format. At least one system is available as a bundled package including scanning (with a high resolution camera), image processing, and conversion to vector format. The system has been tested by the Wisconsin Land Records Project (Chrisman 1986) and appears very promising.

The likely increase in use of scanning technology in no way limits the need for good manual digitizing and editing software, however. Many source maps are too complex for effective scanning. An operator's ability to select and pre-edit linework from a map sheet with many different layers or themes will continue save time and effort over available automated approaches. Good cartographic practice can simplify conversion to scanning methods, but there will always be a need for digitized map input from less than ideal source materials.

Issues of raw input aside, however, scanned data still will need topological and attribute coding and probably a certain amount of linework editing. Good editing software will simplify this latter phase of input that often consumes more than half of the time for data base construction. Our new software is designed to read in raw vectors output from scanning software, to build underlying topology, and to support graphic editing and attribute coding in the same manner as manually digitized data.

<u>Conversion to raster format</u>. Both vector and raster formats have distinct advantages for geographic information system applications. Vectors generally serve better for applications requiring the faithful representation of linework or a comparison of boundaries between areas. If a landowner wants to see what part of his land is unbuildable because of poorly drained soils or floodplain boundaries, generalized cell maps will not adequately portray the overlaps and intersections of layers. Vector software can both display actual boundaries and perform polygon overlay processing to extract polygons with different combinations of attributes or identify areas of change between the same coverage at two different time periods. Vector representations also typically require less storage space for overlays that are irregular or sparse, since a standard grid representation requires a value to be stored for each cell whether it is different from its neighbors or not. (Run-length encoding, quad-tree data representations, and other grid data representation formats can reduce the storage problem, as can increasingly larger computer memory capacities).

Raster or grid-cell encoding offers different advantages for geographic data processing and display. By standardizing the geometric spatial unit, attribute data can be processed independently and more efficiently. In addition, the regular spatial neighborhoods of a grid can be utilized for generating new attribute coverages from one or more original values. Elevations on a grid can be used to generate separate slope, aspect, and drainage layers. Distance functions can be applied to create an overlay of values indicating the distance of each cell from a road, stream or other feature. Certain cell values can be treated as full or partial barriers to passage, so that overlays indicating the net friction or cost of moving from one area to another can be generated for use in applications such as timber sale layout or erosion modeling. (Tomlin 1983, Tomlin 1985) Grid cell based systems also lend themselves to processing along a time axis, either by simple comparison of two time periods for an overlay or by more sophisticated spatial modeling techniques which use relationships among sets of current cell values to create a new generation of values, simulating change in the natural environment.

Raster display options extend beyond thematic mapping of one variable to include relief-shaded displays, where each color value at a cell is altered by lightening or shading depending on the exposure to an artificial sun as determined by the elevation of that cell. (White 1984). A raster data format also interfaces more easily with remotely sensed data and other digital imagery. Low-cost image processing systems can now better register and integrate satellite data with traditionally developed map sources. New inexpensive video digitizing cameras and paint systems offer many possibilities for integration of video images of landscapes or photographs with more traditional cartographic representations.

Ready conversion between vector and raster data formats will be increasingly important for effective data management in a GIS. New vector-to-raster conversion software written at the Laboratory supports grid encoding from vector data at user-specified grid resolutions to support both large and small scale application work from the same vector data. The grid resolution is previewable from the vector input software so that the effect of rasterizing at a given cell resolution can be predicted. The conversion software supports attribute assignment to cells using a priority system, area-weighting, or averaging of the attributes of polygons which occur in each cell.

Attributes. Attribute support should be as open-ended and flexible as possible. Identifiers used to label structures such as chains or polygons internally in the software need not ever be seen by the user, who should be free to deal only with thematic attributes or codes. Ideally an attribute should link to other data base software for maximum flexibility. A key value such as a parcel identifier can be used as a link to a complete record of all availabe attributes for the parcel in a non-graphic data base. Standard data base queries in the attribute data base need to be linked to graphic data so that all areas meeting certain attribute criteria be displayed. Conversely, a user looking at a map should be able to retrieve any available attributes about an area or set of areas that he or she identifies graphically. Data base query capabilities should also be flexible enough to combine spatial and attribute processing for selection, as in identifying all residential lots within 300 feet of a stream that have been granted building permits in a given time period.

A data base may require attributes for more than one spatial unit, as in the case of line attributes associated with boundaries in a polygon file. Line attributes could be used to select out town or county boundaries in a parcel file for a heavier line width, for example, or could serve as the basis for selection or aggregation functions not determined by the attributes of a full polygon. Nodal attribute support can also make input software adaptable for applications such as transportation networks or structural diagrams for which node attributes are critical.

As memory restrictions become less and less of an issue and as relational data base software tools become more sophisticated and accessible, the distinctions between spatial and attribute data will become less important. Queries relating the two types of data will be performed more naturally and powerfully under an umbrella structure that can support the degree of variability present in spatial data within the logical framework of a textual data base manager.

Updating: relations between layers. Building a geographic data base that can efficiently accept changes and additions is much more complicated than a one-time project which can remain in a frozen state once completed. Too often the initial hurdle of mass digitizing masks the need to plan for data updating until the first version of a database is finished. Thereafter, managing data consistency given a proliferation of changes across a multi-layered data base can be a major headache. Both careful data base planning and new software capabilities are needed to limit the difficulties involved in updating.

We have found it helpful to organize the data components of a multi-layer geographic data base by each component's quality level and by the relationships between different layers or parts of layers. A highly accurate source, usually large-scale photogrammetrically derived engineering maps, can serve as the master spatial reference for a project. This source, whatever it is, should provide the framework of control points that will be used for map registration, and should also provide what will be accepted as definitive boundaries for rivers, roads, municipalities, and perhaps other features. When the same feature occurs in another coverage, as on a zoning map or property map, the coordinates of that feature should be replicated in the new layer rather than redigitized with slightly different coordinates. A link may also be established between the original line and its clones so that when the original is changed a the change is flagged in the dependent coverages. This approach follows the organizational structure of relational data base systems, which can link several files in order to avoid duplication of information that is common to each. The spatial nature of the proposed links prevents a direct referencing of coordinates -- if a road from a project street base map also serves as a property boundary, that boundary may be cut several times during digitizing of the property overlay by parcels which abut the road. While software can ensure that the boundary will continue to overlay the original exactly, it may have been split into several topologically separate units. A change in the road in the base file cannot therefore be automatically replicated in the property file; a link flagging the change could, however, alert the operator to the need to edit the road boundary as it exists in the property file to match the new coordinates from the original street base file.

The nature of the interrelationships between data layers can range from very little to a strict hierarchy. A contour map of elevation, for example, may have little direct relationship with other coverages except where contours bound rivers or highways where cut and fill has taken place. In contrast, zoning and property boundaries are often closely linked, with zoning boundaries in some jurisdictions always following property parcel lines. In some cases the relationship is one of relative data quality -- a stream may be inaccurately drawn on an old property map, so that the stream boundary should be replicated from a more modern and accurate map. In other cases the relationship may be based on administrative procedures or on legal requirements. Wetlands maps often have a legal authority that transcends issues of accuracy in compilation or change over the time since a wetlands map was officially adopted.

Software tools can be constructed to manage these relationships to whatever extent desired. A simple approach incorporates the first-in priority described in the data priority section above. Data is organized so that the best base map information is digitized first, and relevant boundaries from that map are copied to form the start of various new layers. For example, streets and water features could be copied to start a property parcel overlay. As parcels are digitized, the software priority system will ensure that the original street and water lines are not distorted as new lines intersect them. When a zoning overlay is constructed, appropriate property boundaries are selected and copied to a new file, which is then further edited or added to as a separate zoning overlay.

An simple enhancement which we have implemented is to display a reference file in a different color from the currently active file. The reference file does not interact spatially with new data, but serves as a graphic reference for the placement of new information. This feature is also very useful in pre-editing adjacent sheets so that endpoints along the edge will fall within tolerance of each other when the two files are later merged.

The reference file concept could be expanded so that the coordinate structure would be accessible as new data is digitized. Lines from the reference file could be picked and copied to the new file on the fly, or lines meant to match a line in the reference file could be tested against the endpoint coordinates. One commercial system claims to be able to check new lines agains features in other layers (McLaren and Brunner, 1986). A further enhancement might construct a formal link between the two files at common points or lines, so that a later change in the parent file might be flagged for change in the file containing the clone.

The question remains whether the additional complexities of a data structure incorporating spatial references outweigh the advantages, and whether new management problems are not created. It is nearly impossible to anticipate all the circumstances which will occur with spatial data. If the software overhead limits free use of a data base or is unreliable, simpler approaches which rely on a good understanding and on well-planned sequential updating may be more effective. In our experience, good graphic feedback and internal support of topological consistency can help to guide a relatively untrained user through complicated editing using primarily good intuition and judgement.

Complex interrelationships between files may also force a spatial rigidity that undermines long-term database maintenance. Adjustments to the national horizontal datum will require significant coordinate changes is some areas of the country. Existing rubber-sheeting methods may also serve to bring initially incompatible files within reasonable distance tolerances of each other (Beard and Chrisman, 1986). If primary data analysis is to be accomplished in raster format, then precise graphic alignment in vector files may not serve any useful purpose. Applications requiring only graphic references between information, so that data can be overplotted but not spatially intersected or integrated, may belong in standard overlay drafting systems without topology that are typical for engineering and utility applications.

Expanding into three dimensions. GIS systems will increasingly need to incorporate terrain modeling, texturing, and full three-dimensional objects to meet the demands of users. Many CAAD systems now support complex three-dimensional objects and even certain solid modeling operations, but buildings are displayed either on flat planes or on extremely simplified models of terrain. Conversely, advanced GIS systems which permit perspective views of terrain with surface texturing to simulate landscape vegetation cannot support more than simple prismatic structures if structures are supported at all. To some extent these shortcomings derive from typical working scales, where CAAD systems work at site or development project level and GIS systems often incorporate terrain models of a large region. As users see the sophistication of each separate type of system, they will naturally want to be able to integrate, particularly for work at the scale of a single city or park where buildings and terrain interact very importantly.

Effective terrain depiction through perspective views of grid elevations has become relatively common today; some systems offer features such as overlays of contours or other data layers in the perspective model, or textural enhancement for the image. However, grid cell models are difficult to merge with fully three-dimensional systems which contain arbitrarily-shaped object faces in arbitrary orientations. One solution has been to model terrain with a triangulated irregular network (called a TIN) of faces, so that the faces produced can be added to the display list of 3-d object faces for a single perspective viewing operation using hidden surface removal techniques. TINs are not very easy to interpret, however, and produce perspective views dominated by the graphic appearance of the triangular faces.

Contour pancakes (analagous to terraced hillsides) offer more intuitive modelling forms with closer links to the cardboard models traditionally used to represent sites. (Van Norman, 1986) One research area we are exploring is the ability to convert topologically structured and compact contour lines from our map input program into three-dimensional contour pancake objects for terrain representation in a three-dimensional modelling system. Each object must be completely independently represented so that it can be moved freely without affecting other objects unless it intersects them.

The addition of buildings to landscape models offers similar utility for understanding either current landscape conditions or proposed changes. If the detail level required for building representation does not warrant conversion of a complex terrain model into a three-dimensional modelling system, then approaches which add additional detail at certain cells of a grid-based terrain model may be useful. Typically the cell size for a regional terrain model would be too large for effective representation of buildings; superimposition of smaller block models on top of the perspective view of terrain may prove a useful approach to explore.

CONCLUSION

The ODYSSEY system performed very well for the Minute Man National Historical Park Data Base. We anticipate continuing to build data bases in ODYSSEY format using new input software which significantly shortens the time required to digitize and edit map files. By keeping the same data format, vector-based ODYSSEY thematic mapping programs will be available for output, as will polygon overlay processing and utility functions such as selection, generalization, or aggregation. The ability to maintain natural spatial data units for checking and for site-level analysis is often important, and data bases built initially in raster format can not provide the same resolution without using unworkably small cell sizes.

When the ODYSSEY family of vector-based programs are combined with flexible conversion procedures to grid cell format at any user-specified cell resolution, the benefits of raster data processing can also be realized. Relief-shaded raster displays are an effective alternative to standard ODYSSEY shaded maps and oblique views, and the simpler geometric grid cell format allows much more efficient processing of multiple overlays.

Many anticipated benefits of computer mapping have not yet been realized because of continuing high costs for data input. Our work in streamlining input procedures with a modern user interface and automatic processing is aimed at a broad goal of more accessible mapping technology. Interfaces to scanning input systems, to raster-based map analysis and display programs, to image processing software, and to three-dimensional modeling programs will further extend the range of options available for geographic data input and processing.

Large commercial systems will continue to meet the needs of larger organizations with the financial and personnel resources to devote to computer mapping as a primary effort. Our goal is to develop software which will help to define the directions that commercial systems should move while meeting many educational, research, and government or institutional software requirements. Increased accessibility of mapping technology depends on continued improvement and integration of software and data management techniques so that users with a minimum of training can understand and control what they wish to accomplish, at a cost they can afford.

ACKNOWLEDGEMENTS

The development of the Minute Man National Historical Park data base was supported by the National Park Service, North Atlantic Region, Office of Planning and Design, under Cooperative Agreement #CA-1600-4-0006 with Harvard University. We would like to thank Rolf Diamant and Isabel Mancinelli of the National Park Service and Kelly Chan and Bojana Bosanac of the GSD for their assistance in this project.

Development of new map input software has been funded in part by a grant entitled "A Cartographic Modeling System for the IBM Personal Computer and Courseware in the Fields of Architecture, Landscape Architecture, and Design", Internation Business Machines Corporation.

REFERENCES

- Broekhuysen, M. and Dutton, G. 1983, "Conversations with ODYSSEY." In <u>Proceedings of the</u> <u>1983 Harvard Computer Graphics Conference</u>, Volume I. Graduate School of Design, Harvard University, Cambridge, Massachusetts.
- Beard, M.K. and Chrisman, N.R. 1986, "Zipping: New Software for Merging Map Sheets", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 1, pp. 153-161.
- Chrisman, N.R. 1986, "Effective Digitizing: Advances in Software and Hardware", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 1, pp. 162-171.
- Chrisman, N.R., Mezera, D.F., Moyer, D.D., Niemann, B.J. Jr., Sullivan, J.G., Vonderohe, A.P. 1986, "Soil Erosion Planning in Wisconsin: An Application and Evaluation of a Multipurpose Land Information System", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 3, pp. 240-249.
- Corson-Rikert, J. and White, D. 1985, "Issues in Map Digitizing", Laboratory for Computer Graphics and Spatial Analysis, Report Number LCGSA-85-17.

- Corson-Rikert, J. and White, D. 1985, "The TRACE Program for Map Digitizing", Laboratory for Computer Graphics and Spatial Analysis, Report Number LCGSA-85-18.
- Corson-Rikert, J. and White, D. 1986, "Building Large Data Bases for Planning: The Minute Man National Historical Park", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 3, pp. 142-151.
- Department of Landscape Architecture, Graduate School of Design, Harvard University 1985, Alternative Futures for Minute Man National Historical Park, Cambridge, Massachusetts.
- Keating, T., Phillips, W. and Ingram, K. 1986, "Database Design Specific to Geographic Information Systems", Kork Systems, Inc., Bangor, ME.
- Malcolm, J. 1985, The Scene of the Battle, 1775. Historic Grounds Report, Minute Man National Historical Park, Cultural Resources Management Study #15, Division of Cultural Resources, North Atlantic Regional Office, National Park Service, U.S. Department of the Interior.
- McLaren, R. and Brunner, W. 1986, "The Next Generation of Manual Data Capture and Editing Techniques - The Wild System 9 Approach", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 4, pp. 50-59.
- Morehouse, S. and Broekhuysen, M. 1982, <u>ODYSSEY User's Manual</u> (2 volumes). Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, Harvard University, Cambridge, Massachusetts.
- Moreland, Deborah K. 1986, "Development of an Automated Laser-Based Data Capture System at the U.S. Geological Survey", Proceedings, American Congress on Surveying and Mapping - American Society for Photogrammetry and Remote Sensing 1986 Annual Convention, Volume 1, pp. 189-198.
- Samet, H., Rosenfeld, A., Shaffer, C., Nelson, R. and Huang, Y.-G. 1985, "Application of Hierarchical Data Structures to Geographical Information Systems: Phase III", Technical Report #1457, Center for Automation Research, University of Maryland, College Park, MD.
- Scott, D. and Iyengar, S. 1986, "TID Translation Invariant Data Structure for Storing Images", <u>Communications of the ACM</u>, Association for Computing Machinery, Vol. 29, No. 5, pp. 418-429.
- Teitelbaum, T. and Reps, T. 1985, "Immediate Computation, Or How to Keep a Personal Computer Busy", Engineering: Cornell Quarterly, Vol. 20, No. 2, Autumn 1985.
- Tomlin, C. Dana. 1983, "Digital Cartographic Modeling Techniques in Environmental Planning", Ph.D. dissertation, Yale University.
- Tomlin, C. Dana. 1985, "The IBM Personal Computer Version of the Map Analysis Package", Laboratory for Computer Graphics and Spatial Analysis, Report Number LCGSA-85-16.
- Van Norman, M. 1986, "A Digital Modelshop: the Role of Metaphor in a CAAD User Interface", <u>Design Computing</u>, Vol. 1, No. 2.

White, D. 1984, "Relief Shaded Thematic Mapping By Computer": <u>The American Cartographer</u>, Vol. 12, No. 1, pp. 62-67.

White, D. 1985, "Geographic Information Systems on Microcomputers", Laboratory for Computer Graphics and Spatial Analysis, Report Number LCGSA-85-11.