Conversion of Automated Geographic Data to Decision-Making Information

Stephen J. Ventura
Land Information and Computer Graphics Facility, University of Wisconsin - Madison, Madison, WI 53706

ABSTRACT: The Dane County Land Conservation Department (LCD) needs an information system to certify that farmers comply with state and federal erosion control mandates, to guide farm planning decisions, and to monitor compliance. The LCD implemented a relational data base management system on microcomputers. This system, called the "Cooperator Tracking System," contains information about farms and farmers. The system had no automated spatial data or analysis capabilities. A separate initiative created a geographic database, built on the concepts of a multipurpose land information system and assembled for county-wide soil erosion control planning. After conversion for use on microcomputers, this database provided the missing spatial components for the LCD. The geographic database was merged with the Cooperator Tracking System. Extensions to relational database management and geoprocessing software have been used to create a "push-button" system that provides information specific to LCD's mandates. The result was a microcomputer based system that provides automated spatial analysis resulting in program specific information and that updates attribute and geographic data on a transactional basis.

INTRODUCTION

SIMILAR TO THE PLIGHT of perhaps as many as 3000 local conservation offices throughout the United States, the Dane County, Wisconsin Land Conservation Department staff (LCD) has a tremendous information processing burden. The conservation provisions of the recent federal farm bill (Food Security Act of 1985) require the LCD to develop farm plans for almost 3000 farms by 1990. They determine eligibility and monitor compliance with four federal programs (Sodbuster, Swampbuster, Conservation Reserve, and conservation compliance) along with similar requirements for state programs (Non-point Source Pollution Control, Soil Erosion Control, and Farmland Preservation cross-compliance). In other words, they need to collect, maintain, and analyze data about the ownership, management, and physical resources of 243,000 hectares (600,000 acres) of farm land in Dane County.

The potential efficiencies of automation were an attractive option for dealing with this burden. In fact, two separate automated systems evolved in response to these conservation mandates. The first system (referred to as the Cooperator Tracking System, or CTS) was primarily a database management system that deals with records about farms, farmers, and farm management practices (Dane County, 1986). The CTS was a locally created equivalent of the USDA Soil Conservation Service's CAMPS software (Computer Assisted Management and Planning System). Like CAMPS, there were no automated spatial data in the CTS. It only referenced geographic data maintained on paper maps. The CTS was developed with the assistance of the County's Division of Systems and Data Processing using dBaseIII software.

The second system was a multipurpose land information system (MPLIS), which was used for county-wide soil erosion control planning (Ventura, 1988). It was developed in cooperation with the University of Wisconsin-Madison and several other state and local agencies as part of on-going research in land records modernization and rural resource planning. It contained detailed soils, land cover, and land management data (Niemann *et al.*, 1987; Ventura *et al.*, 1988; Niemann *et al.*, 1988). However, it was not designed for use on microcomputers or for transactional update.

The merger of these two systems provides most of the information needed by conservation staff for fully automated program determinations, compliance monitoring, and farm planning. The combined system is understandable to conservation staff and, at least for most queries, provides information in reasonable amounts of time (about as long as a farmer is willing to wait at the service counter). It makes the provision of conservation services more efficient and provides consistent, comprehensive, and impartial methods, hopefully resulting in more equitable administration of conservation laws.

CONCEPTS

In over-generalized terms, there are two types of automated spatial data systems - systems primarily focused on automated mapping and systems focused on spatial analyses. The first type, typified by a land records system, needs to quickly reproduce graphic representations of data from relatively large source scales and manage frequent updates. In contrast, systems built for resource decision-making are often assembled in single massive data automation efforts from relatively coarse data. Analytic capabilities are quite important, particularly the combination and cross-reference of data from various sources.

An examination of the information processing needs of the LCD suggested that characteristics of both types of systems were needed. They needed transactional update capabilities and quick graphics to respond to interactions with farmers in a timely manner. They also needed analytic capabilities, especially polygon overlay and interfaces to soil erosion models. They required information both at the scale of individual farm management units and at watershed and county scales.

Neither the MPLIS used for county-wide soil erosion planning nor the CTS alone was sufficient to provide all the information for routine decision-making about program eligibility, compliance certification, or farm planning. The MPLIS lacked information about ownership and management systems of individual farms and the conservation practices on farm fields. The CTS lacked automated spatial data, particularly soils and topographic data and the configuration of farm fields.

The overall goal of our project was to link the two data management systems in such a way that retrieval and analysis of both resource information and farm management information was possible on a "real time" basis. An easily learned user interface and transactional update for both geographic and attribute data were necessary components of the overall system design.

The "geo-relational" data model, as described by Dueker (1985), Morehouse (1985), Waugh and Healy (1986), and others, was

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 56, No. 4, April 1990, pp. 511–516.

0099-1112/90/5604–511$02.25/0

selected as the most appropriate for these demands. This concept calls for specialized geoprocessing software coupled to a relational database management system. The use of a topological, vector-based geoprocessing system allows for a broad range of spatial data input, management, and analyses without degrading source information. The use of a relational database management system links the geographic entities with attribute information and provides powerful tools for managing and querying the attribute data.

Budget constraints limit many local conservation offices to microcomputer hardware and commercially available software. Until recently, it has not been possible to create a microcomputer based system from commercially available hardware and software that combines topological vector data structures, analytic polygon overlay, and other geographic analyses with a relational database management system. While database management systems such as dBaseIII and Rbase5000 have been available for microcomputers for quite some time, commercially available GIS on micros were limited to CAD or raster based systems until the marketing of pcARC/INFO by Environmental Systems Research Institute (ESRI, Redlands, California) and STRINGS by GeoBased Systems (Research Triangle Park, North Carolina).

While the geo-relational concept is a good general descriptor for a data model, it provides no mechanism for modeling data flows. The concept is static – "here's a data set and the possible relations between elements" – and does not account for an interface to the changing demands of the real world. A data model needs to describe where and why data are entered and updated, in what form they are used, and how data element relations change.

In the Dane County LCD case, it bacame clear through a structured systems analysis (Gurda *et al.*, 1988) that legal mandates are driving the need to collect, maintain, and use both spatial and non-spatial data. Transactions drive the movement of data through the system. For example, when a farmer requests a farm plan to comply with 1985 farm bill conservation regulations, data are retrieved from the database, combined with additional information from the farmer, analyzed, and archived. Figure 1 presents the general concept of a mandate-driven, transactionally-maintained geo-relational data model, along with some of the specific features of the system developed with LCD.

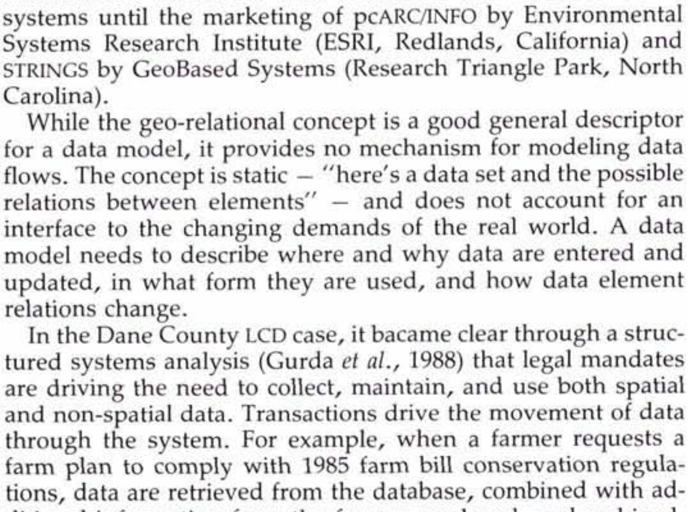

FIG. 1. Concept for a mandate driven, transactionally maintained geo-relational data model, along with some of the specific information flows of a system for local soil conservation planning and management.

SYSTEM DEVELOPMENT

The system developed to automate LCD transactions was based on data about farms and farmers maintained with dBaseIII in the CTS, and geographic data maintained and analysed using pcARC software. Farm fields and tracts provided the primarly linkage between the two databases. These data come from the identification system and delineations used by USDA Agricultural Stabilization and Conservation Service (ASCS). DBaseIII and pcARC are "loosely coupled" through the use of DBaseIII programs and pcARC macros (Tandias, 1989). Temporary files are used to pass parameters and data back and forth. Though the CTS and MPLIS database existed prior to this effort, there were several major tasks to create the combined system. These tasks included

- automate farm and farm field boundaries and link with soils, land cover, and management data;
- convert data in MPLIS to ARC/INFO format; divide large coverages into pieces manageable on microcomputers; download from minicomputer;
- develop dBaseIII applications to incorporate pcARC attribute data in the CTS;
- develop dBaseIII applications to incorporate user supplied information in the database and in various analytic models; and
- develop pcARC macros for spatial query, polygon overlay, farm field reconfiguration, and linkage to other analysis software.

FARM AND FARM FIELD AUTOMATION

Individual farm fields are the smallest unit of reporting for most of the programs administered by the LCD. The boundaries of these entities are needed as a unit of aggregation for analyses of other geographic data. Farm fields have several associated attributes, including land cover (crop) and crop history, management factors for soil loss calculations, and owner and manager of the field. Project planners initially thought they would be able to use digitized 1:4800-scale parcel maps from the County Land Regulation and Records Department to define farm boundaries. Field boundaries would then be digitized on recently acquired 1:12,000-scale orthophotographs. Unfortunately, neither of these records were available in time to meet deadlines imposed by the federal farm bill.

The best records available with farm and field delineations were ASCS photobases. The photography was flown in 1980 and received in local offices in 1985 as 1:12,000-scale rectified prints. Farm boundaries and fields as described by farmers and/or derived from plat maps were delineated with relatively coarse pencil lines. Figure 2 compares ownership boundaries digitized from this source with ownership boundaries from 1:4800-scale tax parcel maps for four sections (approximately 3.2 by 3.2 kilometres or 2 by 2 miles) in northwest Dane County. The discrepancies between the two versions of farms are a complex mixture of error from several sources; including

- identification of section corners (section corners were used as control to transform digitized data into UTM coordinates); The algorithm used to transform digitized data to the map projection (an inverse distance weighted affine) insured that control points were transformed exactly to their homologous points in the projection, but in many cases, section corners were scaled on the photobase rather than identified from ground evidence.
- relief displacement; A "worst case" analysis suggested that a ground distance of 10 metres was the greatest displacement expected at the corner of a photo on a 60 metre hill.
- drafting and digitizing errors; The drafters' placement of lines on the photobases was not precise nor was the digitizing. Some of the photography was acquired fairly early in the morning, making shadows a problem.
- interpretation differences; The photointerpreters did not have access to deed information, so such errors as drawing boundaries on the wrong side of roads (see upper right of Figure 2) occurred

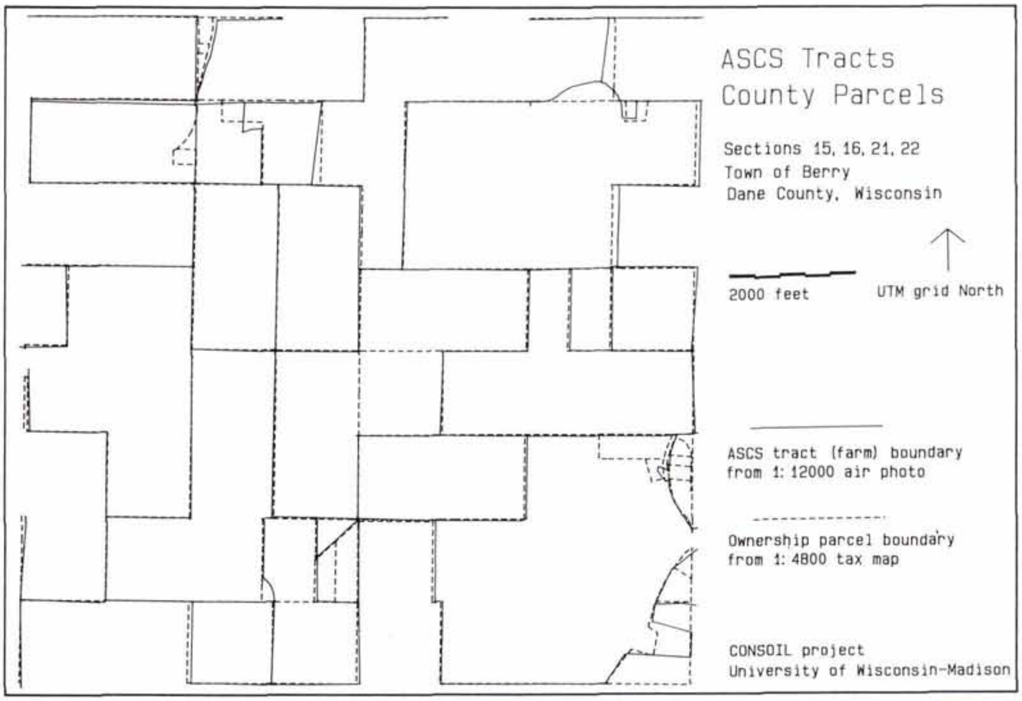

FIG. 2. Comparison of farm boundaries derived from 1:4800-scale county tax parcel boundary maps and from 1:12,000-scale Agricultural Stabilization and Conservation Service photobases.

frequently. Because ASCS photointerpreters were most concerned with farmland, the delineations around non-farmland are the least accurate (see lower right of Figure 2).

- logical differences; Although ASCS generally split separate ownerships into separate tracts, there were cases of longstanding land contracts or adjacent owners with family ties where ASCS treated separate ownerships as one tract (see center of Figure 2).

Although the spatial accuracy of the field and farm delineations was compromised by all of these factors, historically it has been good enough for commodities payments. Local conservation staff have assumed that, if the delineations were acceptable to farmers where payments are involved, they will be acceptable in a less demanding calculation such as a soil loss determination.

The compounded effect of overlaying two coverages with unquantified spatial inaccuracies, such as the farm field boundaries and the soil boundaries, has only been determined in a program specific sense. For example, one of the initial screenings for federal programs required a "highly erodible lands" determination. The areas of all soil types within a field were measured to see if more than 30 percent of the soils are considered highly erodible by the USDA Soil Conservation Service. In Dane County, this was done by overlaying the ASCS based farm field delineations with soils data. For an area of about 31 square kilometres (12 square miles), we compared this method with results from overlay of a more accurate delineation of field boundaries (1:12,000-scale orthophotographs from 1987 imagery). In four of 135 fields, different determinations were made using the less accurate ASCS data. However, even the use ASCS data was more accurate and time efficient than traditional manual techniques (dot gridding or visual estimates). Visual estimates of the area of highly erodible soils within a field varied by as much as ± 20 percent compared to the value determined by overlay of soils and orthophoto delineated fields.

Farms and fields digitized from the ASCS photos will only serve as the initial source of this information. It is hoped that, over time, land ownership information from the county Land Regulation and Records Department can be incorporated in the system. This requires resolution of long-standing institutional problems. It is also hoped that, over time, ASCS will begin using the 1:12,000-scale orthophotographs for field delineations. Traditionally, ASCS has favored working with 1:7920-scale prints because it is too hard to write field identification numbers in delineations of contour strips at smaller scales. We hope that we can demonstrate that an automated approach will eliminate such problems, and so allow all conservation agencies to work from the same orthophoto base. For example, automation will obviate the need to write field identifications directly on source maps.

The database was designed for the gradual replacement of spatial data with more accurate data as transactions occur. For example, when fields are reconfigured, a tract and its fields will be delineated on the most accurate photobase available, and then digitized and inserted in place of the old data. It is assumed that the orthophoto-derived data is more accurate, so the tracts around the new data will be forced to conform to the inserted data. Preliminary testing has indicated that fields delineated on the orthophotos will nest in tracts derived from the county tax parcel maps, so the system will be able to use more accurate tract data when Dane County can provide automated land records data.

DATA CONVERSION

There were two major computing tasks to make soil erosion control plan data available to the LCD - the conversion of the data format from Odyssey to ARC/INFO, and the downloading of the data from the University's VAX minicomputer cluster to LCD's IBM-AT system. It was necessary to develop data conversion software and develop a tiling system for dividing the data into manageable pieces.

Although Odyssey and ARC/INFO have very similar topological vector data structures, there are subtle differences that prevented a direct conversion. In Odyssey, the internal representation and the user's identification of polygons are the same. In ARC/INFO these identifiers are separated and the internal identifier must be unique and sequentially enumerated by the topological processor. Polygons are identified by a label point in ARC/INFO. In the case of soils information, the identification number of soil mapping units (polygons) in Odyssey were not unique numbers. It was necessary to bring unstructured coordinate data into ARC/INFO, topologically structure it, then go back to Odyssey to get the user identification of the labels for those polygons - a more computationally expensive conversion because data already topologically structured in Odyssey had to be structured again with the ARC/INFO "BUILD" and "CREATELABELS" commands.

Even though seamless county-wide coverages were assembled for the Dane County soil erosion control plan, that analysis was done township by township (approximately 36 square miles at a time for each of 35 townships). The size of the data sets for a township were burdensome even on a VAX 8600, requiring up to a couple hours CPU time and 10s of megabytes per township for analyses, processing, and plotting. It was clear that data would need to be partitioned for management on LCD's microcomputers.

To some extent, the photos from which farms and fields were digitized represent a *de facto* tiling (spatial partitioning) scheme. The county attempted to edgematch this coverage in only a few cases. This was a very labor intensive process, because there was no edgematching of the analog photobases. In fact, there are gaps and overlaps in what is delineated on adjacent sheets and there are often discrepancies between sheets. In the long run, LCD will replace these data with delineations from more accurate photobases (that do not spatially correspond to these bases), so these bases were not used as the tiling structure for other coverages. In the short run, they have cut apart the photobase derived tracts and fields into individual tracts and re-assembled them into townships, ignoring gaps and overlaps until better data are available.

Township sized units were chosen as a partitioning for several reasons. First, the data were split at that size for the soil erosion control plan. Second, one township of the largest data set, the soils data, just fit on a 1.2 megabyte floppy disk. Third, it tended to minimize the times that more than one tile of data was needed for any given tract. As with any tiling scheme, there are tradeoffs between the number of times it is necessary to retrieve more than one tile to see a feature of interest and the length of time it takes to access a single tile. In other words, with larger tiles, more than one tile is needed less frequently, but it takes longer to retrieve that piece. Farm tracts are the features of interest most often needed by LCD. Because of the way land was homesteaded in Dane County, farms end on township boundaries more often than any other feature, especially more often than on arbitrary features such as quadrangle boundaries.

Data conversion (about 100 Mb in Odyssey format, about 250 Mb in ARC/INFO format) took about 175 minutes of VAX 8650 CPU time. Downloading data to an IBM-AT took a little over 2 hours connect time per township, or about 80 hours for the whole county.

Macros and Programs for LCD

Applications specific to LCD's operation have been developed in three general areas: to incorporate pcARC attribute data in the Cooperator Tracking System and generate reports specific to various conservation programs; to create command files of DOS and pcARC instructions for manipulation of coverage data; and to incorporate user supplied information in the database, including updating pcARC coordinate and attribute data.

The macros for incorporating geographic data in the CTS have been developed by the same county personnel (Division of Systems and Data Processing) that initially developed the CTS. DBaseIII commands with syntax similar to those which conservation staff previously used now initiate DOS batch files and pcARC macros. Conservation staff have also learned how to modify templates so new applications can be developed.

Most of the geoprocessing is done through pcARC "Simple Macro Language" (SML) programs that manage pcARCEDIT and pcARCPLOT sessions. For example, to plot the highly erodible soils of a single farm tract, a dBaseIII program inserts the number of the tract in a SML macro and then executes it. It has also been necessary to develop some DOS batch files that directly call pcARC and in-house executable programs for certain functions. For example, this was necessary for a program to link in an inverse distance weighted affine transformation, and for another to calculate the town, range, section, and quarter section of polygon centroids. This latter program is used as a cross check between the location of a farm tract and fields in the geographic coverage and the Public Land Survey System description of their location as recorded in the CTS

Because both dBaseIII and pcARC programs occupy large portions of available memory in the operating system, it has been less troublesome to not have them both residing in core memory at once. Parameters are passed by intermediate files or global (system environment) variables. Likewise, the internal storage of data in INFO or TABLES (pcARC) is a binary format, which is converted to an ASCII format before it can be passed to BaseIII and vice-versa.

It was possible to "batch" some portions of the analyses. For example, determinations of highly erodible lands for the federal farm bill required the overlay of soils and farm field boundaries. This was done for an entire township at a time. Because it requires 4 to 6 hours on an IBM-AT, command files were used to control the entire process. The command file ran the overlay, passed data to dBaseIII, performed the analyses, and generated reports after the conservation staff has gone home. The resulting overlay product could then be used interactively, such as showing farmers the location of highly erodible soils in relation to farm fields on CRT displays.

Perhaps the most interesting problem was reconfiguring farm fields within a tract, in essence using the combined system for interactive farm conservation planning. Figures 3 and 4 are pen-

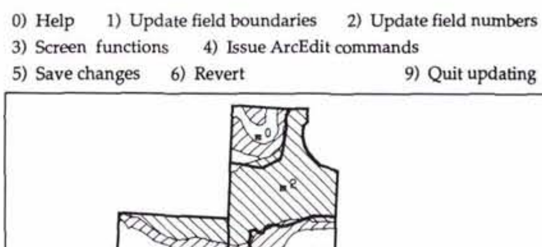

FIG. 3. Farm fields, hydric soils, and highly erodible soils for farm number 7295.

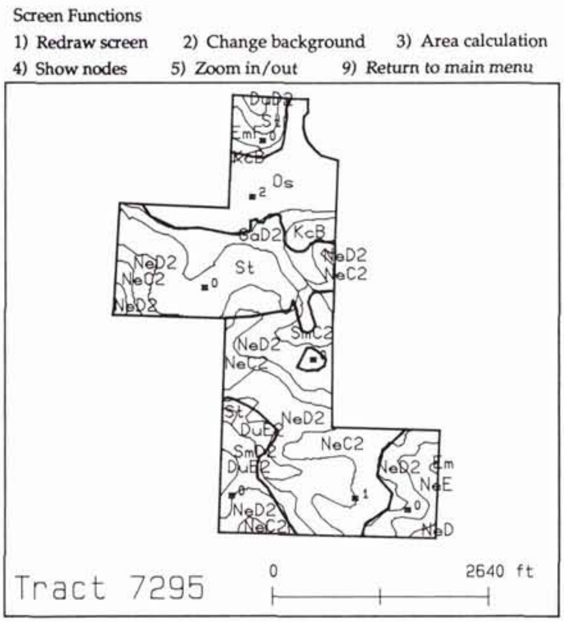

FIG. 4. Farm fields and soil mapping units for farm number 7295. The fields have been reconfigured to reduce tillage on highly erodible soils.

plotter versions of color graphics terminal displays. With the combined system, conservationists use similar displays as farm plan worksheets. Figure 3 shows current farm fields, highly erodible soils, and hydric soils on Tract 7295 in western Dane County. At the top of the map is one of several menus for manipulating the field boundaries and changing background information. The area and identification of every soil type within each fields, can be passed to the Cooperator Tracking System for analysis using dBaseIII programs. Information about cropping and conservation management for each field resides in the CTS. Field and farm area weighted erosion rates are calculated with modified version of the Universal Soil Loss Equation.

After calculating erosion rates under current conditions, the conservationist can re-digitize field boundaries. In re-configuring fields, the goals are to eliminate highly erodible land and to incorporate lands with lower erosion potential into crop production. Changes can also be made in the cropping and management practices. These "what if" scenarios can be done with the farmer standing by to provide feedback on potential changes. After changes are made in conservation practices, cropping, and/or field boundaries, the data are again overlain and soil losses calculated. If projected erosion rates are acceptable, the changes can be archived as a farm plan, and later incorporated in the field coverage when actual changes on the ground are confirmed. A plot can be quickly generated for a farmer's use, showing the reconfigured fields and the location of highly erodible soils.

Figure 4 shows an updated version of Tract 7295, depicting each soil mapping unit symbol (another display option) and slightly re-configured field boundaries. In this example, average erosion rates for the fields in Tract 7925 were reduced from twice the tolerable rate to the tolerable rate. The erosion rate went from 2.98 m.t./ha./year (8.1 tons/acre/year) to 1.54 m.t./ha./year (4.2 tons/acre/year) by switching to contour plowing and eliminating some of the most erodible lands from field 1 (lower right) and adding less erodible lands to field 2 (top). The result is 3.6 hectares (9 acres) less crop land, but it will probably yield similar gross production by using more productive land and better conservation practices.

The interactive farm planning aspect of the system has not been in operation long enough to know how farmers will react, but so far most people have been impressed. We do know that the county-wide soil erosion control plan using geographic information system techniques has been well accepted by farmers and technicians because it is comprehensive, consistent, and detailed (Ventura *et al.*, 1988).

CONCLUSIONS

The microcomputer based information system combining a topological vector-based GIS with relational data base management capabilities provided all the information for farm program management and monitoring and most of the data necessary for individual farm plan recommendations. This system was built on the principles of multipurpose land information systems, including successful sharing of digital data. Under their mandate, the Dane County Land Conservation Department assumed the role of data custodians and used transactions as a basis to maintain data currency. As other institutions adopt similar information technologies, these principles may make it possible to create fully automated systems for a wide variety of conservation and farm management applications, at decreasing cost to the citizen taxpayer.

ACKNOWLEDGMENTS

The work reported herein has been funded by the United States Department of Agriculture-Soil Conservation Service and supported in-kind by the University of Wisconsin-Madison (School of Natural Resources, College of Agricultural and Life Sciences, Institute for Environmental Studies, and College of Engineering), and Dane County (Land Conservation Department and Division of Systems and Data Processing). Particular thanks go to John Amundsen, who developed the Cooperator Tracking System, Kevin Connors, the Dane County Conservationist who has persevered through this technology transfer process, and Aman Tandias, who has worked out many of the difficult microcomputer problems.

REFERENCES

Dane County Land Conservation Committee, 1986. *Cooperator System for Erosion Control and Waste Management*. Madison, Wisconsin: Dane County Land Conservation Committee, 37 pages.

Dueker, K. J., 1985. Geographic Information Systems: Toward a Geo-Relational Structure, *Proceedings, Auto-Carto 7*, Washington, D.C., 11-14 March 1985. pp.

Gurda, R. F., B. J. Niemann, S. J. Ventura, J. Amundson, and H. Braunschweig, 1988. Developing Data Management Models for Multi-Agency Land Information Systems, *1988 ACSM-ASPRS Annual Convention Technical Papers*, Volume 2, pp. 39–46.

Morehouse, S., 1985. ARC/INFO: A Geo-Relational Model for Spatial Information, *Proceedings, Auto-Carto 7*, Washington, D.C., 11-14 March 1985. pp. 388–397.

Niemann, B. J., J. G. Sullivan, S. J. Ventura, N. R. Chrisman, A. P. Vonderohe, D. F. Mezera, and D. D. Moyer, 1987. Results of the Dane County Land Records Project, *Photogrammetric Engineering and Remote Sensing*, 53(10):1371–1378.

Niemann, B. J., and others, 1988. The CONSOIL Project: Conservation of Natural Resources through Sharing of Information Layers, *GIS/LIS'88 Proceedings*, San Antonio, Texas, 30 November - 2 December 1988. volume 1, pp. 11–25.

Tandias, A., 1989. Integration of pcARC/INFO and DBaseIII, *Technical Papers, 1989 ASPRS/ACSM Annual Convention*, Baltimore, Maryland, 2-7 April 1989, Volume 4, pp. 30–39.

Ventura, S. J., 1988. *Dane County Soil Erosion Control Plan*, Madison, Wisconsin: Dane County Land Conservation Committee, 104 pages.

Ventura, S. J., N. R. Chrisman, K. Connors, R. F. Gurda, and R. W. Martin, 1988. A Land Information System for Soil Erosion Control Planning, *Journal of Soil and Water Conservation* 43(3): 230–233.

Waugh, T. C., and R. G. Healey, 1986. The GEOVIEW design: A Relational Database Approach to Geographical Data Handling, *Proceedings, Second International Symposium on Spatial Data Handling*, Seattle, Washington, 5-10 July 1986, pp. 193–212.

(Received 9 March 1989; accepted 15 May 1989; revised 10 July 1989)