Environmental Systems Research Institute

GIS Trends

Map Data Standardization *A Methodology for Integrating Thematic Cartographic Data Before Automation*

by Jack Dangermond, ESRI President and Eric Harnden, Director of Aerial Information Systems

Users of GISs commonly analyze and compare map overlays of geographic features such as vegetation, soils, landforms, geology, and slope.

Automated techniques for combining these overlays have opened up extensive opportunities for analysis of the relationships between and among these spatial parameters. This is beginning to help land use planners and scientists of all types in systematic assessment and interpretation of these relationships. Unfortunately, for a given study area, maps of these features rarely exist at compatible map scales and levels of detail, if the maps exist at all. Nor is there a guarantee that existing maps were created using a consistent basemap and imagery source. Yet as GISs continue to grow in popularity, demand for unified series of thematic maps to make meaningful analyses will also grow.

At ESRI, we have frequently found that the quality of available geographic data for inclusion in such databases is typically poor. The problems with these data include incomplete map coverage, inadequate map classifications (with respect to the projected decisions which needed to be made), and inconsistencies in the maps (e.g., in map unit resolutions, line crenulation, scale, accuracy, dates of data collection, sampling method, and classification system).

Map data standardization (MDS) is a systematic compilation of thematic overlays referenced to a common basemap and imagery source. The MDS process was developed out of the need for consistency among data layers and to reduce automation costs by decreasing the number of input maps.

MDS involves a number of techniques, including

- Project design
- Inventory preparation
- Thematic mapping
- Map integration
- Editing
- Map automation

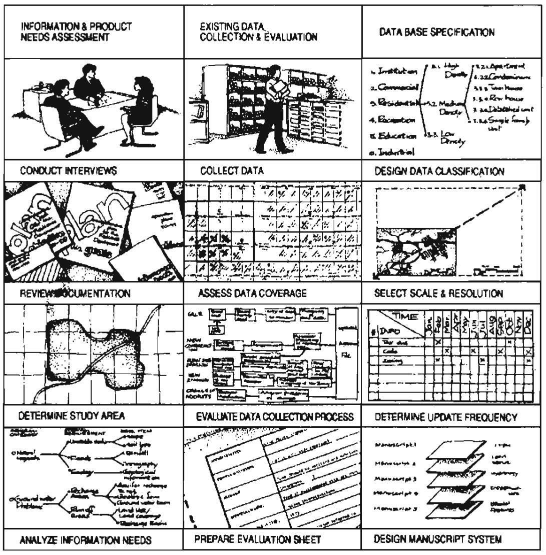
This article will focus chiefly on the first five of these techniques.

PROJECT DESIGN

A number of aspects of project design were discussed in a recent *ARC News* article by Don Chambers ("Overview of GIS Database Design," *ARC News*, Spring 1989, pages 17–21). This article concentrated on the design of the database as it appears to the GIS user. Some elements of project design are concerned more particularly with the data that go into that database.

Data Needs Assessment

Data needs assessment involves the development of a clear definition of the specific uses for the inven-



tory information, usually through meetings and interviews with the users.

During these meetings, it is valuable to review the types of information already identified as useful; to make breakdowns of the various levels of information currently used; to become familiar with the methods of handling, storing, and retrieving information, user libraries, and catalogs; and to get an overall understanding of the geographic and thematic areas for which users have data needs.

Arrangements are made to collect any data that users can contribute to the inventory. These data could be in the form of maps, reports, bibliographies, air photographs, satellite imagery, and so forth.

Interviews, although useful, are typically not enough. It is necessary to review various documents outlining functions, responsibilities, and mandates to analyze the exact data requirements.

To assist in this process, it is useful to go through an exercise of breaking down the specific analysis requirements and related data needs that are necessary to support specific functions. For example, slope and soil maps may be necessary for interpretation of soil erosion. Soil erosion evaluation may be one type of analysis necessary for water quality assessment in an organization responsible for general planning and management of the environment. A hierarchical structuring of specific data needs related to general responsibilities can be effectively represented in table, matrix, or related graphic form.

The data needs assessment process is greatly assisted by clearly documenting needs and having users review, discuss, and creatively participate in the final definition.

Collection and Evaluation of the Existing Database

Data collection involves the acquisition of geographic information sources such as maps, books, reports, imagery, aerial photography, and related documents supplied directly by the user (typically during the data needs assessment interviews), or acquired from other institutions. This is accomplished by searching bibliographic indexes, public institutions, government agencies, and private companies for information to fill the categories of data identified during the data needs assessment.

Each data item collected is reviewed and evaluated to establish its usefulness and appropriateness for inclusion in the GIS.

Data are first grouped into general data need categories (e.g., natural resources, administrative district, and infrastructure). Specialists in each field rate each item as important, useful, marginal, or of limited use. In addition, notes are recorded regarding the format, scale, map projection, date, media, classification categories, area of coverage, and any useful considerations on how the data items match or conflict with each other.

Those data items of major value to the mapping effort for the GIS are identified for in-depth consideration during the classification and map design steps.

After a thorough review of the data has been completed and the results compared to the data needs of the user groups, data categories for which reliable coverage does not exist across the study area are identified. Separate data gathering or mapping projects are then designed to ensure that these categories are not omitted from the database.

Database Design

The design of the database depends in part on both the information needs identified and the assessment of the existing information base.

The data in the GIS database consist of spatial information and attribute information, linked together. To obtain maximum utility from the GIS, it is necessary to develop carefully both the structure of the classification systems used for the attributes and the style and format of the spatial component.

Development of Classifications

To meet each of the data needs requirements, general categories of information and detailed classifications must be developed. The data collected are carefully reviewed and consideration is given to the level of detail needed for each category. The classification schemes are structured hierarchically to allow aggregation of classes at different levels for use at a variety of map output scales or tabular summary levels.

The classifications are typically developed toward making them as similar as possible to the classification that the user groups are familiar with. Where new types of data are prepared, the classifications should be based on current use within the professional writings for each given discipline.

All classifications developed for the GIS need to be expressed as numeric codes to facilitate computerized handling of the data.

Classifications represent qualitative, quantitative, or descriptive groups of individual data occurrences in a systematic order.

In some instances, classifications are developed that operate at two levels. The first level identifies the mapping unit by type, for example, a geologic formation. This is a descriptive class. The second level is an expansion on the descriptive class intended to provide quantitative and qualitative values, for example, measurements or ratings given to the geologic formation, such as age, rock type, stability, strength, and so forth.

The first-level classifications are associated directly with the map units by a sequential code list. The second-level values are associated with the corresponding first-level classes by an expansion code matrix. This matrix lists the first-level descriptive codes followed by their second-level qualitative or quantitative values expressed as numeric codes.

This two-level approach to classifying and coding the data can minimize the amount of space required to store the information in the computer and can also make the subsequent map overlay modeling efforts more efficient.

The numeric codes developed to represent each classification can be used as values by the computer to generate tabular listings, draw maps, or produce analytical models. Therefore, it is important to keep all of the classifications in a logical order; start with the smallest and end with the largest; start with low and end with high; and so forth.

Data classes should match map units in levels of specificity (i.e., the most detailed classification levels match with the most detailed mapping unit). Unnecessary data classes that are more detailed than the ability to identify and map units should not be allowed.

All of the data codes are structured to be entered into the computer as numbers. Thought is given to the output potential of creating a variety of shades, colors, or symbols using various computer devices. Limitations imposed by those devices can be reduced by incorporating the appropriate structure into the classification.

Design of Manuscript Maps

A manuscript map is a map sheet originally prepared as an input document for automation. The manuscript sheet is typically related to a specific basemap size and scale. In addition to the classification structure, the design must include rules regarding minimum polygon size resolutions and line crenulations. Finally, a determination must be made regarding which variables will be placed on which manuscripts.

The use of an integrated approach to mapping presents some special map design problems as well as some distinct advantages for resource inventories. Unlike manually prepared maps that rely on a variety of colors, shades, line widths, and symbols to portray the information they contain, maps designed for computer database input show data simply as points, lines, or polygons (areas). A numeric code provides the descriptive values for each map unit. Symbol recognition, color separation, clarity of shading, and line width are not factors considered when designing computer data input maps. Spacing of lines and points, line crenulation, and minimum polygon size are important, as these are problem areas for the automation process.

Because the computer has perfect logic and total recall, it is possible to pack a great deal of information onto a single map. This ability to pack information is limited only by the computer's programmed ability to discern discrete points and, from a more practical point of view, the ability of the human cartographer to prepare the maps for automation.

It is desirable to put information from similar categories onto one manuscript when designing input maps to take full advantage of shared data locations or boundaries. It is also easier for data users to comprehend data if correlated information is grouped together. Finally, the integration forces the resolution of inconsistent data classes.

Some typical integrated map manuscripts that might be designed

Continued on page 17

Environmental Systems Research Institute

GIS Trends

Standardization

Continued from page 16

for a regional inventory are listed in Table 1.

Manuscript	Data Form	Number of Attributes
Integrated Terrain Unit	Polygons	10-20
Surface Hydrography	Polygons and Lines	10-20
Special Natural Features	Points and Lines	5-10
Infrastructure, Settlements, and Cultural Features	Points and Lines	10-20
Political and Administrative Units	Polygons	20-30
Biological Resources	Polygons and Lines	15-30

Table 1

Beyond the actual design of the database files, it is necessary at this time to specify the exact data sources, imagery, levels of fieldwork, and basemaps, as well as mapping methods that will be used in the actual creation of the manuscripts.

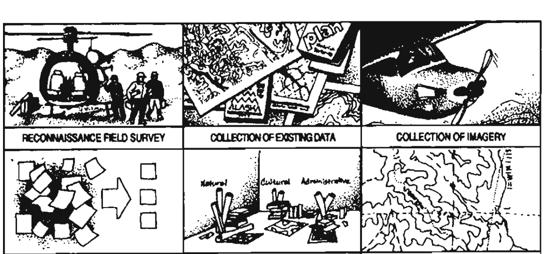
INVENTORY PREPARATION

Before the actual inventory mapping is begun, a number of basic, preparatory steps need to be completed, if the final product is to meet ongoing database needs.

These steps include the following:

- Conducting a reconnaissance field survey
- Collecting existing data and source mapping
- Selecting the best data
- Grouping and cataloging of existing data
- Basemaps preparation

Some of these tasks may be completed during project design efforts



Inventory preparation techniques.

Reconnaissance Field Survey

This task is intended to familiarize investigators with the actual geography of the study area, thereby establishing a sound basis for later work. This effort is often done when basemaps, aerial imagery, and collateral data are being collected for the project or database. This involves a field visit and examination of the area to be mapped by a team of resource specialists. For large areas, a small airplane is usually employed for a full overflight of the study area. A series of oblique aerial photographs are taken for later use and assistance in the interpretation of various types of aerial imagery. Portions of the study area accessible by road may also be investigated on the ground. Direct observation, recording of environmental conditions, review of aerial photograph signatures, and selective ground sampling may also be conducted on these ground visits. As a result, investigators are left with a "sense of the landscape."

Collecting Existing Data and Source Mapping

Data sources are collected based on the results of the database design. The lack of a usable data source for a specified component is also identified at this point.

Basemaps and imagery are acquired when the mapping scale for the inventory has been established. Basemaps with topographic, hydrographic, and cultural features, as well as a spatial reference grid such as latitude and longitude, are reproduced on stable-based Mylar. The most recent, cloud-free aerial photography or satellite imagery is acquired, with the parameters again being determined by database design. For data types lacking adequate sources, plans are made to prepare thematic maps using photo interpretation or field surveys.

Some of the data materials may need to be specially ordered or acquired, including computer information such as digital terrain tapes, recent imagery, or basemaps that may require special permission for use.

Selecting the Best Data

The reliability of the delineations and classifications of each data source is evaluated. This effort is especially important where multiple map coverages for the same attribute exist, and must be sorted through to determine the most reliable and consistent information to be used in the mapping. For example, a geology map prepared at a large scale generally inspires more confidence than one prepared at a small scale.

Grouping and Cataloging of Existing Data

Usually, an automated card catalog is set up at the time data are collected; this ensures that the data are organized properly and can be searched and extracted in a rapid manner. Each piece of collateral data is also labeled so that nothing is lost during the inventory process.

Basemap Preparation

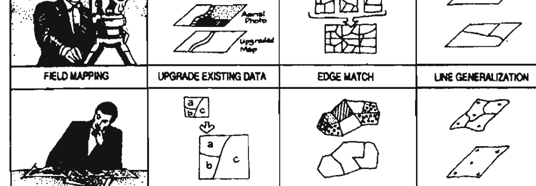
An important basis for integrated resource mapping is a well-prepared set of basemaps. In the United States, the most practical system of basemaps is provided by the U.S. Geological Survey (USGS). Topographic quadrangle maps are, essentially, equal area and equal distance maps; provide a number of locational reference points; and can be organized into a consistent module sequence convenient for mapping and automation. These maps range in scale from 1:24,000 to 1:250,000 and can be transferred from paper prints or acquired in fixed stable-based Mylar film from the USGS. The topographic contours shown on these basemaps are extremely useful for registering map data and photo interpretation of information relative to the exact locations on the earth.

THEMATIC MAPPING

Before creation of the final manuscript maps, a set of primary thematic overlays (termed *"themes"* for the remainder of this article) are prepared. These sheets consist of a map for each inventory variable or group of compatible variables, and are drawn to the scale and map projection of the topographic basemaps and referenced to a common imagery source. Four different methods are used to prepare the compilation sheets:

- Direct transfer
- Image interpretation
- Photograph revision
- Field mapping

They are described below.



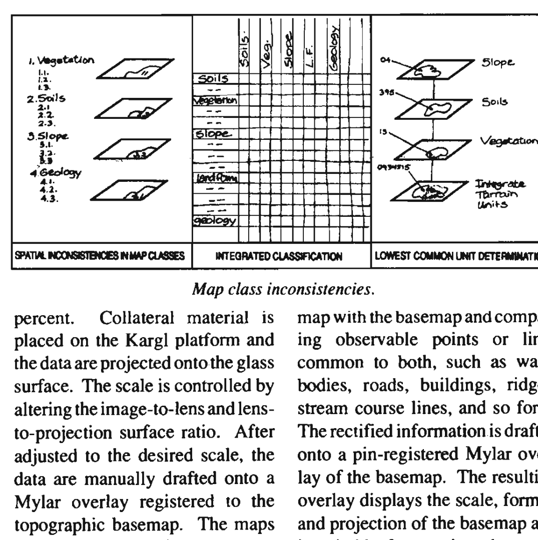
Direct Transfer

Collateral information sources that do not require reinterpretation are assembled by map module. For those items that are at the correct scale, the data are simply transferred directly to a Mylar overlay of the basemap with careful rectification of the data to the new basemap.

If the data are at a scale other than the mapping scale, rescaling precedes the rectification and transfer steps. Rescaling may be performed manually or photographically.

Manual Rescaling

Data items that display simple legends or contain relatively gross mapping units are usually rescaled using an optical pantograph (Keuffel and Esser Kargl). This is a type of reflecting projector with a rated distortion of less than 0.01 percent. Collateral material is placed on the Kargl platform and the data are projected onto the glass surface. The scale is controlled by altering the image-to-lens and lens-to-projection surface ratio. After adjusted to the desired scale, the data are manually drafted onto a Mylar overlay registered to the topographic basemap. The maps drawn on the Kargl instrument are then carefully edited to ensure that all information has been transferred correctly.



Map class inconsistencies.

Photographic Rescaling

When data items contain complex or detailed information, manual rescaling is too time-consuming and the risk of coding error through transposition or omission is increased. In these instances, it is more efficient to rescale the collateral data photographically. A precision-copy camera with a high-resolution lens is then used to produce paper positives. The proper scale is assured by using a method called triangulation in which three points of known location are marked on the collateral data and then matched against the basemap on the camera's viewing glass. An attempt is made to match all three points as closely as possible. This process achieves the most satisfactory match of map scales in both the horizontal and vertical directions.

Rectification

Because the various collateral items are often originally prepared using a variety of cartographic techniques and mapping formats, it is necessary to rectify each data item to the topographic basemaps. This rectification ensures that each data item is shown in its proper location and configuration according to the new projection. This is accomplished by carefully reregistering the data map with the basemap and comparing observable points or lines common to both, such as water bodies, roads, buildings, ridges, stream course lines, and so forth. The rectified information is drafted onto a pin-registered Mylar overlay of the basemap. The resulting overlay displays the scale, format, and projection of the basemap and is suitable for use in subsequent data integration steps. Common maps requiring rectification are soil and vegetation maps originally drafted on nonplanimetric photos.

Reclassification

Frequently there is a difference in the map classification between an original source map gathered during the search for existing data and the final classification that is desired on the thematic map to be included in the database. When this type of difference occurs, it is necessary to reclassify the map by collapsing or subdividing the spatial units expressed on the map sheet. In the case of collapsing the classes, a simple visual method is used for dropping out line separations that are not appropriate to the new classification. Where refinement is desired, units are divided according to other maps and data identifying where these new lines should break down a map unit.

Continued on page 18

Environmental Systems Research Institute

GIS Trends

Standardization

Continued from page 17

Generalization

As scales and unit resolutions of mapping change, the original line work delineating boundaries and linear features varies in levels of abstraction. When this occurs, it is necessary to generalize line work cartographically. This is usually done using previously specified standards and keys to ensure consistency as the graphics are modified.

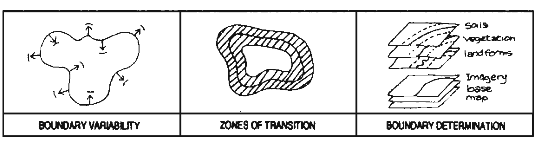
Edgematching

In cases in which study areas contain multiple map sheets, an important thematic mapping function involves edgematching of lines crossing over adjacent map sheets. This function is performed by abutting joining sheets and editing and modifying line work and attribute codes using imagery or other source maps.

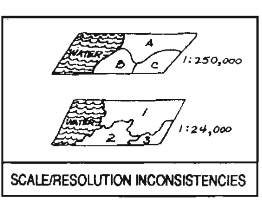
Image Interpretation

Based on the initial evaluation of the source data, data deficiencies are identified. In some instances the deficiencies are remedied through the use of a remote sensing technique called image interpretation, more commonly referred to as photo interpretation.

Image interpretation begins with the development of an interpretive key that associates visually recognizable units of color, density, texture, pattern, and the environmental setting or situation (e.g., landform, elevation, drainage pattern, and relation of one area to another), with the elements of the data classification system.



Resolution of boundary transition.



Scale/resolution inconsistencies.

In cases in which stereo photography is available, stereoscopic interpretation is used to provide three-dimensional character to the key. This key provides information to the interpreters, allowing consistent judgments and decisions in the interpretations of the map variables. The interpreters contribute to the key those factors that worked best for them individually, while maintaining consistency with keys used by other interpreters. Each interpreter keeps a set of working decision rules that are reviewed and updated by all interpreters.

Interpretation can rely on a variety of remote sensor data types including LANDSAT Thematic Mapper or MSS, SPOT, radar images, high- and low-altitude color infrared aerial photography, high- and low-altitude black and white photography, and so forth.

Registration of the images to the basemaps must address parallax problems. Although the images are often scaled to the basemaps, a perfect match is frequently not attained. To resolve this, the interpreter must register one area of the image at a time, constantly reregistering the image to the basemap as mapping progresses. Accurate registration requires alignment of linear or point features on the photo enlargement with those same features on the basemap. Such features include roads, stream courses, valleys, mountain ridges, and peaks.

Photograph Revision

In instances in which the initial data evaluation shows a data source to be of good quality, but out-of-date, subject to frequent change, or cartographically generalized, it is possible to upgrade the data through the use of image interpretation. This technique does not result in the creation of totally new data, but, rather, of improved and more detailed thematic map sheets for use in compiling the resource inventory.

The process involves a visual comparison of the data source to the project imagery. A transparent Mylar copy of the original data source is attached to the topographic basemap and laid on top of the imagery; then the observed changes are traced onto the data source overlay. As with image interpretation, registration is maintained by shifting the image to match linear or point features.

Field Mapping

Where there is an absence of available thematic data, it is usually necessary to conduct in-field mapping of the data coverage to be inventoried. This in-field mapping can be at a variety of intensities, and is usually done using some form of sampling method (e.g., forest plots, soil pits, and borings). Subsequently, these samples are interpolated or extrapolated using the various image interpretation and related techniques described above.

MAP INTEGRATION

The next major technical procedure involved in a standard inventory project is the integration of multiple "themes" onto a single sheet. This process is most commonly used in the integration of thematic natural resource polygon maps such as soils, vegetation, slope, landform, and geology. This technique has been called integrated terrain unit mapping or ITUM; we now term it Map Data Standardization, or MDS. It involves the systematic integration onto a single manuscript of geographic parameters that were initially mapped on several sheets.

Standardized mapping (MDS) vastly reduces the cost, management, and analysis problems typically encountered in automation and subsequent use of data for GIS applications.

Table 2 summarizes some of the basic principles that underlie the rationale for this standardized approach to generating manuscripts for use with GISs. This table presents these principles in relation to the problems encountered during map overlaying when using a strictly parametric approach. Also presented are the ways in which the standardized approach solves these problems.

The first problem resolved through the use of MDS is a basic cartographic one and is important from the standpoint of information, interpretation, and display. Among map variables, many have boundaries and attributes that are naturally interrelated and coincident (principle one in Table 1). With parametric mapping, these variables are mapped independently by different professionals with different objectives, different map accuracies, and at different times; therefore, the overlay of this information almost invariably results in classification inconsistencies and geometric sliver errors.
Sliver errors are small errors created by the overlapping of coincident boundaries.

A second associated principle presented in Table 1 is that polygon boundaries reflect gradational changes and are not an absolute fixed line. Therefore, the line represented on a map is actually an abstracted line representing a zone of transition. As a result, when parametric maps are overlaid, the truly coincident lines almost never exactly overlay one another. The integration process resolves this type of error by using a common basemap and common imagery to assist in the definition of the boundaries among and between the interrelated variables contained in the typical GIS database. The integration process effectively reduces the time required to computerize and overlay a large number of different maps. The manual integration results in considerable in the savings of labor and computer time to automate and manipulate the data. Moreover, the storage and analysis of the data on the computer are both more efficient.

Integration of Themes

Simply compositing each of the previously prepared "themes" onto a single map sheet would result in the creation of numerous splinter polygons that become meaningless and confusing and lead to substantial data management problems.

The mechanical process of integration involves the manual merging of data shown on the individual themes. This is done by comparison to images, topographic basemaps, and each other to eliminate meaningless and confusing splinters as the manuscript is drawn. Finally, the attribute values from the original themes are associated with the spatial units (polygons) on the standardized map through the use of sequential coding lists.

Principle	Common Situations	Problems In Map Overlay	IPU Solutions
Classifications of landscape are interrelated	Parametric mapping rarely recognizes interrelationship of land attributes	Inconsistent - Map classes - Sliver errors	- Reorganize classification system - Spacially derive lowest common integrated unit
Polygon boundaries relect gradational change	Inconsistent gradtional boundary delineations between parameters	- Sliver errors	Common boundary determination through multistage integration
Multiple scales of data can be accurate but inconsistent when integrated to common scale	Line crenulations and unit resolutions vary	- Sliver errors - Resolutions	Remap all thematic data to common resolution and scale
Areal information changes over time	Maps are out of date and inconsistent with respect to time	- Map classes - Data reliability	Update information overlays to common date

Table 2

MDS was developed to capture and coordinate the various environmental components of a natural resource database, but has now been expanded to include almost all types of mapped data. This type of mapping is a method of compressing a number of environmental factors from a variety of data sources onto a single map. The map displays homogeneous units (i.e., units that have the same general characteristics distributed throughout). This procedure overcomes a number of basic problems.

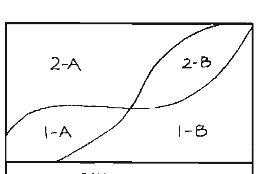


Figure 7. Sliver errors.

Using MDS, all information prepared for automation is subject to a second round of verification, updating, and improvement using remote sensing. During this process, data discrepancies resulting from multiple sources at multiple scales are often detected and corrected, either through remote sensing interpretation procedures or by referring the questionable areas to an agency or organization responsible for initial interpretation, verification, and correction. In general, the result of this process is that the data incorp ated into a database are superior in quality to the data input to such a process.

The next problem resolved deals with cost. The cost of database computerization is a function of, first, the number of maps to be automated, and, second, of the complexity of the features of these maps, created by compressing related polygon overlays.

The basemap represents the project's "spatial reality"—the formal relationships between points, lines, and polygon features in the project area. The imagery represents the project's "thematic reality"—the actual form of the project area on the ground.

Integration begins by choosing the two most reliable and/or graphically complex themes and pin registering these simultaneously to the basemap and the imagery, forming a "sandwich" or set of four sheets. The most graphically complex themes are chosen first to minimize the number of edits required as the standardization process proceeds iteratively.

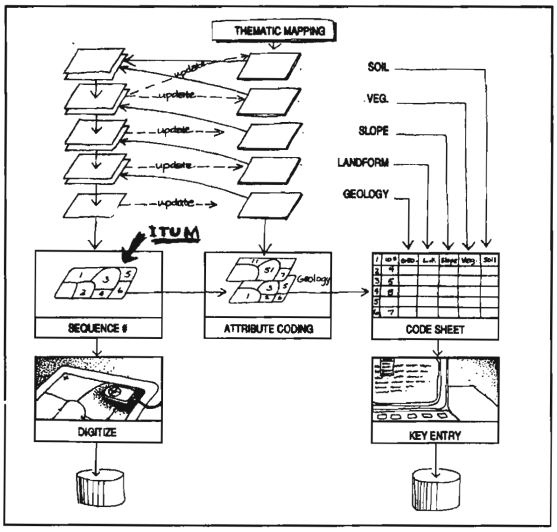
A new Mylar overlay is placed on top of this set of four sheets. On this new Mylar will be redrawn the first theme as its boundaries are changed to agree more closely with those of the second theme. The

Continued on page 19

Standardization

Continued from page 18

location and extent of the changes are established by comparison to the basemap and project imagery. This new Mylar overlay becomes a standardized form of the theme to which it refers and will be termed a "standardized theme." At the end of integration, each of the original themes has such a new, corresponding standardized theme.



Map integration.

Next. the standardized theme replaces the original theme in the sandwich, a new Mylar is placed on top of the resulting sandwich, and the second of the two themes is now standardized.

A third theme is now compared with first one and then the other of the two completed standardized themes. A third standardized theme is thus produced.

Now, one after the other, all of the original themes are standardized by comparison, pairwise, with each of the standardized themes that have been produced. As a result, every line on every standardized theme has been examined and revised as necessary. The persons performing the integration compare each theme, pairwise, with the others, register and compare them to the images and the basemap, and decide where to draw the polygon boundaries on the new attribute code sheet.

For example, a geologic unit may be identified as an old terrace deposit, and a corresponding soil type may be described as forming on old terrace deposits. The units, therefore, should have coincident boundaries. Slight adjustments to the lines drawn around the unit can be made using imagery and basemaps as guides.

This process continues until every map unit has been checked against the images, basemap, and the original theme. The original theme is then replaced by the new standardized theme in the set overlaying the basemap in the sandwich.

By the same process, a new standardized theme is drawn for each of the other themes, using those polygon boundaries already decided on from the previous integration process.

The process continues until all the themes are standardized and new standardized themes have been created for each one.

Then the attribute code sheet for each standardized theme is checked for correct transfer of information and correlation between data items. These attribute code sheets will later be used in the encoding and editing step that follows.

Integration involves making judgments about the reliability of each of the variables, the correlation between the individual variables, and how well they have been registered to the basemap. These decisions determine the configuration and location of which lines will be used to form the final standardized polygons.

When complete, the standardized themes are ready for consolidation onto the final manuscript. The manuscript is drafted by compositing all standardized themes onto a single Mylar sheet, one at a time. Each polygon formed represents a unit of the landscape having its own unique set of characteristics separating it from the adjacent areas.

Linear and point data standardization and integration follow similar procedures but usually do not require new variable code sheets. In most cases, straight transfer of theme data onto the manuscript yields no conflicts.

Numbering and Encoding

The spatial data shown on the integrated manuscript and the attribute values shown on each of the standardized themes are prepared for machine processing in a way that allows the values of each data layer to be associated with the corresponding units shown on the integrated map. This is accomplished by sequentially numbering each integrated map unit on the manuscript and then preparing a list of sequentially numbered coded attributes that match the sequentially numbered spatial units.

After every map unit has been numbered, an encoding form is prepared. The manuscript data are coded by associating the sequential number on the manuscript with the sequential number on the encoding form. Each data variable is assigned a specific *field* (a set of columns) on the encoding form. Each manuscript data layer is encoded by overlaying the manuscript on top of the appropriate standardized theme overlay. The variables are coded by reading the manuscript's sequential number and the corresponding data code number for that polygon. The data code is then recorded in the assigned data column for that corresponding sequential number. The coding process continues until all the data variables are recoded for all polygons, lines, and/or points on the manuscript.

EDITING

Two basic categories of data editing are performed. These include editing of the manual manuscripts and editing of the automated data. During the manuscript preparation process, quality control checking is performed at each map step. Each theme is checked for uncoded polygons, lines, and points, unclosed polygons, consistency of information, correlation between variables, basemap registration, photomorphic signature, and transposition errors. The maps are edgematched to adjacent modules for line location and code consistency.

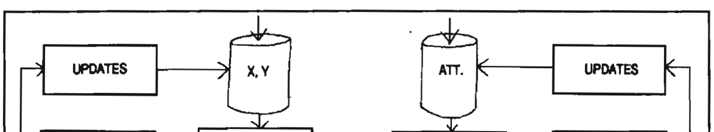
During the integration process, the standardized themes are edited for correlation between variables by overlaying them, one on another, pair-wise, and looking for areas with incompatible attribute combinations.

During the encoding process, the manuscript is edited for missing lines by overlaying the manuscript on top of each attribute code sheet. The manuscript is also edited for drafting errors and unnumbered and double numbered polygons, lines, and points. The manuscripts and encoding forms are code and line edgematched by matching manuscripts to adjacent module manuscripts and comparing corresponding polygon and line code values and line positions with each other.

The manuscripts are overlaid on the map sheets to check consistency of study area boundaries and module borders. After the final review, the standardized manuscript and the corresponding attribute code files are ready for digitizing and key entry.

Subsequent to automation, the database manuscripts and attributes are also subjected to further editing. This editing falls into two basic categories: editing of the manuscript line work and editing of the map element attributes. The original manuscript is compared with the digitized graphic output to determine graphic errors (e.g., spike errors and incorrectly digitized locations for points, lines, and polygons).

Subsequently, the computer is used to create dropline plots representing plotter displays of each thematic variable (theme) included within the attribute database. These themes might be soils, vegetation, geology, landform, slope, and so forth. These dropline plots are compared manually with the standardized themes. Errors in either attributes or graphic location are noted and corrected.



Editing.

Automated Editing

Automated techniques are used to conduct additional error checking, including code and consistency checking between and among the variables. The attribute file is edited in terms of class ranges, as well as consistency among the classes maintained within the integrated file. The class range edit consists of checking each variable against a theoretical minimum and maximum variable range that is possible for any given map element described by an attribute class. The computer simply reads the variable class within the attribute table and compares it against a theoretical class range provided in an associated table.

The second automated attribute verification involves a consistency evaluation comparing all of the classification variables and their subclasses against one another to determine those additions where improbable conditions exist, such as a wetland vegetation type falling on an arid soil. These are flag checked by the photo interpreter responsible for creation of the theme, and updated where necessary.

CONCLUSION

Most of these methods are not new, but certainly are relevant, important, and sometimes essential in the conduct of an overall inventory database methodology. We recognize that there are many new automated technologies that are making substantial improvements in the manner in which geographic data are manipulated, displayed, and communicated. However, for large surveys of natural as well as cultural data, a great deal of careful thought and extensive mechanical effort must still be invested to create meaningful and workable data to use with this technology.

It is true that the process of creating a totally standardized database as outlined here results in modification of the typical existing database provided as input for such a survey. In certain cases, it can be argued that these modifications constitute the basis for disagreement regarding accuracy and may introduce the potential for mechanical error. However, for the overwhelming majority of cases, it is ESRI's experience that the opposite is true; that a well organized and integrated database leads to consistency and information agreement and helps users focus on the real issues of planning, resource management, and environmental assessment.