DIGITAL DATA CONVERSION

I. PURPOSE

The purpose of this document is to provide guidelines covering the digital conversion of existing manually drawn data into GIS format. The goal is to obtain the best possible digital map representing an original analog map. As such, it is assumed that a suitable source map exists. While the ultimate quality of the database is dependent on the quality of the source materials and the care used in automation, the procedures for original mapping are not discussed. It must be remembered that precise automation procedures will not improve an inferior base map:

Garbage In - Garbage Out

Digital database standards have been proposed to provide:

1. Agencies and vendors contracted to automate spatial information with guidelines for digital data to be input into the VGIS with respect to format, quality control and data documentation.

2. Users of the database with information on the technical specification of existing data to aid in determining the viability of those data for specific applications.

Standards are necessary to assure the quality of GIS data, ensure compatibility with existing data in terms of format and attribute coding, and facilitate the most efficient organization, manipulation and analysis of these data.

II. SOURCE MAP MEDIA

- Stable-base mylar

Whenever possible, the most scale-stable media should be used in the automation process. Most preferable would be the original mylars upon which the map data were drafted. If originals are not available for direct encoding, then positive contact prints (i.e., mylar original to mylar copy) would be desirable. The following represents, from most to lease desirable, the media to be used as the source material for automation:

1. Mylar original
2. Mylar contact reproduction from mylar original
3. Non-stable base paper from mylar original
4. Non-stable base paper

III. GEOGRAPHIC REGISTRATION

- *Use VGIS tics for Vermont orthophoto base maps*
- *Use VGIS tics for USGS topographical maps*

A minimum of four registration points (tics) are required for each automated map. Tics are generally located in the four corners of the manuscript map. Whenever possible, Vermont's orthophotos are the VGIS base map for small area efforts. A set of tics is available for 1:5000 orthophotos and USGS 7.5 maps.

In cases where new tic marks are to be established, they should never be created through digitization. New tic locations should be established through keyboard entry of exact geographic coordinates.

IV. DIGITAL TOLERANCES

- *RMS not to exceed 0.005*
- *Fuzzy tolerance equivalent to 0.002 inches at the scale of the manuscript map*
- *Dangle distance approximately 0.10 inches*

There are several digital tolerances which affect the accuracy and resolution of a digital map that can be explicitly defined during the map automation process. The Fuzzy Tolerance defines the minimum distance separating arc coordinates. The Dangle Distance defines the minimum allowable length for a dangling arc. The Root Mean Square Error Tolerance (RMS) defines the error incurred when predefined tics are used to register a map on the digitizer.

The RMS error is a very important tolerance to consider because it indicates, in part, the accuracy with which the digitizing technician captures the locations of the tic marks for a map in the X and Y directions. In order to preserve the spatial integrity of map features during the automation process, it is extremely important to keep the RMS error as low as possible when a map is registered on the digitizing table. The recommended maximum RMS is 0.005. VGIS requires that records of the RMS be kept during the automation process and that these records be delivered along with the digital data.

V. DATA CAPTURE

- *Use a minimum number of coordinates to define a line*
- *Eliminate unnecessary pseudo nodes*

In setting the standards for quality control, it is important to consider the resources for storing state-wide digital data. In order to facilitate the most efficient use of computer storage, it is necessary to encode only the minimum number of vertices needed to capture the essence of a cartographic feature within the 0.010 inch accuracy limit is discussed below in the section on Cartographic Accuracy.

A pseudo node separates two arcs from one another. These should only used when an attribute change occurs along an arc. Unnecessary pseudo nodes increase the number of arcs and the size of the attribute tables, and degrade the quality of plotted line symbols.

VI. COINCIDENT FEATURES

- *Do not double digitize map features*

Coincident features are those which are in common between two or more data layers. For example, if a political boundary is formed by a river bank, the river bank and political boundary are coincident features.

Coincident features must be digitized only once. Regardless of the care taken in digitizing, slight differences are inevitable if a feature is digitized more than once. These differences lead to problems with small "slivers" if the layers are topologically joined.

Coincident features, digitized once, can be placed into master template coverages from which they can be retrieved when needed for incorporation into another layer, or they can be taken directly from the original coverage. In the above example, if the river was digitized first, the arcs necessary for the political boundary can be selected and placed in the new coverage.

VII. ATTRIBUTE CODING SCHEMA

- *Use VGIS approved coding schemes*

All attributes should be coded according to the coding standards adopted by VGIS. These are detailed in "Attribute Definitions and Codes", *VGIS Handbook*, Part 3, Guidelines, Section D.

Typically, the adopted coding schemes reflect those commonly in use by organizations responsible for collecting and maintaining the particular type of data. For instance, coding of a soils database will follow the conventions of the USDA Soil Conservation Service. Nationally recognized formats also exist for coding of data layers such as wetlands and land use.

In cases where a coding scheme has not been adopted, it is crucial that whatever system is used be fully documented. Also, when devising a new scheme, it is important to allow for as many different foreseeable uses of the data as possible. For example, in a land use coding scheme, it is preferable to have individual codes for different types of agricultural use (e.g., cropland, pasture, barnyard) than just a single code specifying agricultural use. The individual codes may be aggregated, but the single code cannot be broken into its components,

thereby limiting its use.

VIII. CARTO-GRAPHIC ACCURACY

- *Digitize map features to within the equivalent of 0.01 inch of the original manuscript, at the scale of the manuscript.*

Digital map quality warrants special consideration. It is very important that the quality of the source maps be preserved as carefully as possible throughout the automation process. Careful, consistent and systematic digitizing, plus thorough verification are essential.

It is recommended that stringent guidelines be adopted for point, line and polygon accuracy and they should exceed National Map Accuracy Standards (NMAS). Recommended VGIS standards for feature accuracy are:

- *90% of the planimetric features on the digital map must be within 1/100 inch (0.010) of the centerline of that feature on the manuscript map when plotted at the original scale.*
- *100% of all features must be within 0.020 inch.*

The 0.010 inch interval is equivalent to a standard 0.010 plotter pen width. When a proof hardcopy plot of the digital map is overlayed on the original base map, discrepancies will be seen as an open space between the plotted feature and the original manuscript.

IX. SPATIAL TOPOLOGY

- *No overshoots*
- *No undershoots*
- *No slivers*
- *No open polygons*
- *No unlabeled polygons*
- *No unresolved line segment intersections*

Data submitted to the VGIS shall be topologically clean and free of errors. The ARC/INFO commands CLEAN and BUILD are used to create topology, while the LABELERRORS and NODEERRORS can be used to detect certain topology problems.

X. EDGE MATCHING

- *Adjacent coverages must be edge matched*
- *Map features are not to extend beyond coverage boundaries*

Line segments (arcs) which intersect the boundaries of a coverage must be accurately edge matched with the corresponding line segment in the adjacent coverages. Computer edge matching techniques ensure an exact match. In lieu of an exact match, arcs must be matched to within 0.010 inch, centerline to centerline.

Arcs must not extend beyond (overshoot) or fall short of (undershoot) the coverage boundary.

XI. DOCUMENTATION

- *All coverages shall be documented*

The producing agency is responsible for providing a standard list of descriptive documentation for each digital map submitted to VGIS. Standard documentation methods are described in "Data Layer Documentation", *VGIS Handbook*, Part 2, Standards, Section D. The reasons for including documentation with each map are as follows:

- Documentation provides a history of each map.
- Documentation provides users with the necessary information to determine the utility of a map for a specific purpose.

XII. DELIVERY OF DIGITAL DATA

Several data formats and media for delivery of data include:

***Data Formats*:**
- *ARC/INFO Coverage ("host" or pcARC/INFO format)*
- *ARC/INFO IMPORT/EXPORT interchange file format*

Media:
- *3.5 inch floppy diskettes*
- *CD-ROM*
- *100MB ZIP disks*
- *4mm DAT tape*
- *8mm tape*

Digital spatial information shall be delivered to VGIS in a format agreed upon in advance by the suppliers and receivers of data.

If PC ARC/INFO is used in automation, small data sets can be delivered on diskettes by first converting to the EXPORT interchange format. Large data sets should be transferred to computer compatible tape, CD-ROM, or 100MB ZIP disks using the specifications described in the previous paragraph.

Other data and media formats are described in "Formats for Data Distribution," *VGIS Handbook*, Part 4, Section B.

XIII. PROOF PLOTS

- *Mylar proof plots*
- *Pen width - 0.010 inch*

Acceptable methods for producing proof plots should be agreed upon by the data supplier and receiver. Preferably, for each digital map submitted, a mylar proof plot of that map must be delivered along with

the corresponding source maps. The proof plot will be of the same scale as the manuscript. All lines on the proof shall use a line width of 0.010 inch or less and be drawn with liquid ink pens.

XIV. LOG FILES

The ARC/INFO LOG file or comparable file describing the steps used in the automation of each map shall be included in the delivery of each digital map.

XV. PRODUCTION HISTORY AND RMS TRACKING

Because the ARC/INFO LOG file does not record the history of the RMS error during automation, a separate written record must be kept of the RMS. The details for recording the production history and RMS tracking are provided in "Data Layer Documentation", *VGIS Handbook*, Part 2, Standards, Section D.

XVI. ACCURACY ASSESSMENT

- *90% of all features within 0.010 inch*
- *100% of all features within 0.020 inch*
- *Topology complete and accurate*
- *100 % of all attributes coded correctly*

It is the responsibility of the producing agency to verify that the original data have been encoded within the accuracy limits set by VGIS. Ninety percent of the cartographic features on a map must be digitized within 0.010 inch measured from the center line or center point of a feature. 100% of all cartographic features must be digitized within 0.020 inch.

The topology of all maps must be free from overshoots, undershoots, slivers and open polygons.