A Prototype Decision Guide and Audit Log for Preparation of Spatial Databases

Mathew Krogulecki
Data Base Graphics, P.O. Box 6723, East Lansing, MI 48826

ABSTRACT: A manual, forms-based system is described which potentially can facilitate documentation of data in the context of database production for some GIS applications. The system documents the data transformation from the point of identification to the point of integration in analysis. The forms-based system consists of five components: (1) map selection, (2) test plan, (3) preprocessing, (4) digitizing, and (5) edit. In addition to documentation, it provides a decision guide for data selection and the means to establish an audit trail for maps and data through the production system.

INSTITUTIONS AND AGENCIES utilize geographic information systems (GIS) for decision making purposes. If modeling efforts are carried out using geographic data to provide information about resource availability or suitability, the data going into these systems warrant scrutiny. As these data are substantially transformed as a result of manipulations designed to render the data suitable for use in GIS systems, the transformations are best recorded also.

The ability of a user to evaluate the appropriateness of a data source can be enhanced by documentation and an associated guide. The information provider can enhance the use of the product by documenting the process used to create it. It is in this context that a prototype decision guide and audit log of steps in the preparation of digital spatial databases has been developed.

GOALS AND ASSUMPTIONS

The decision guide and audit log should be educational and practical. To the extent that it can assist the user in understanding the framework and production steps involved in developing a spatial database, it can be educational. To the extent that it can be used as a "production diary," it can be practical.

There are two assumptions that underly the design and application of the forms: (a) more can be communicated through the use of a partial graphic format, i.e., paper forms, and (b) the forms can provide documentation about decisions made during production.

DESCRIPTION

The main stages of the system being modeled were adapted from Marble (1984). The stages are (1) map selection, (2) test plan, (3) preprocessing, (4) digitizing, and (5) editing.

The selection stage involves an inventory of existing data, perhaps compared to specifications developed with some use of existing sources. Specifications refers to some identifiable characteristics of map data that are required to portray a particular theme or need. These can also be described as cartographic and thematic content.

The test plan stage involves identification of points in the production process where data accuracy might be audited to insure or relate some minimum standard. The cartographic content and thematic content can be monitored as they are transformed in the production process.

The preprocessing stage details what transformations or enhancements the map undergoes before it is actually manually entered into the computer. The digitizing stage is the actual entry of the data into the computer. Finally, the editing stage represents the implementation of the quality assurance procedures that are undertaken after the major data transformation that is brought about by digitizing.

The form for map selection is shown in Figure 1. It provides the opportunity for data inventory and comparision and is based on a comparison table from Hudson (1987). The Table helps to answer questions about the cartographic and thematic content of the maps. Moreover, it also helps form the questions needed to address the requirements of the data bases.

The first column is provided to list special criteria to be used in comparison with the actual, existing map data. These specifications, based on map attributes listed down the left side of the form, may define the characteristics of the best map data for some theme of interest. The sources of map data for the theme listed can then be documented and compared with the specification, and can facilitate selection. The form provides documentation of available data sources and a structure for comparison of data types. Analysis of forms allows decisions about data suitability and they can be reviewed.

The second form, Figure 2, is meant to be used with the first form. It details the potential or selected data source by reference, and lists all attributes or codes describing it's thematic content. The map is given a fuller description through a listing of it's categories.

Once data have been selected for automation, the process becomes one of defining what quality control measures or tests should occur during the subsequent stages of production. These tests allow a user to form some idea of the data quality association with a digital data product. The general form of a quality report is outlined by Moellering and the National Committee for Digital Cartographic Data Standards (1987) and the Journal of American Congress On Surveying And Mapping (1988). As described by these sources, the lineage of the digital data could be constructed from the information provided on the forms, with tests for accuracy described in the test plan form (Figure 3). Table 1 describes some potential elements of the form.

The preprocessing form (Figure 4) records the preparation of map data for digitizing. Major reformatting of data includes redrafting and photomechanical reproductions, as well as data transfers.

Coordinates and georeferencing are scrutinized during preprocessing to insure that all maps can be registered to the same coordinate system and range of coordinates. This is specified in the data inventory or laid out in the test plan. Tests of positional accuracy would occur at this point. Photomechanical or optical processes may also be used to "warp" a map into alignment with the base, or data may be registered to the base map by means of some optical transfer process. Whatever the treatment, the details will be recorded on the sheet.

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 56, No. 4, April 1990, pp. 505–509.

0099-1112/90/5604–505$02.25/0

Data Inventory and Specification

MAP SELECTION

Project ______ Preparer ______

Date to be completed ______ Date of completion ______

Theme ______

REFERENCE

TITLE OF DATA

REFERENCE

DESCRIPTION

CONTACT ADDRESS / PHONE

THEMATIC CONTENT

CHARACTERISTICS

DATA TYPE ______
NUMBER ______
RANGE ______

LIMITATIONS FOR USE

LISTING

GEOCODE	NAME	DESCRIPTION/RULES FOR IDENTIFICATION

FIG. 1. Map Selection – data inventory and specification. Adapted from Hudson (1987).

Data Inventory and Specification

MAP SELECTION

Project ______ Preparer ______

Date to be completed ______ Date of completion ______

Theme ______

ATTRIBUTES	*SPECIFICATION*	*SOURCE1*	*SOURCE2*	*SOURCE3*
COVERAGE study area min/max x,y status				
BASE MAPS coordinate system map projection geodetic control				
DATA SOURCE				
MAPPING SCALE				
MINIMUM TYPE SIZE				
NO. OF CATEGORIES				
PRODUCTION DATE time span update frequency				
DATA TYPE density				
DATA BASE SIZE paper maps digital files				
STANDARD PRODUCTS material # sheets				
ADDITIONAL LAYERS				
PREVIOUS USES				

FIG. 2. Map Selection – data inventory, thematic content.

DEFINE QUALITY CONTROL MEASURES

TEST PLAN

Project ____________________ Preparer ____________________

Date to be completed ____________________ Date of completion ____________________

Map ____________________ Sheet/Panel ____________________

Digital file ____________________

Theme ____________________

ACCEPTANCE TESTS

STAGE	TEST FOR	TEST	ACCEPTANCE
PREPROCESSING			
DIGITIZE			
EDIT			

FIG. 3. Test Plan – define quality control measures.

Preparation for Digitizing

PREPROCESSING

Project ____________________ Preparer ____________________

Date to be completed ____________________ Date of completion ____________________

Theme ____________________

Map ____________________ Sheet/Panel ____________________

REFORMAT/REPRODUCTION

SPECIFICATION
material ______
thickness ______
pen size ______

SPECIAL INSTRUCTIONS

ACCEPTANCE TESTS
Photo-mechanical ______
base material
number of sheets
special references

Data transfer ______
photos to base
coordinate transfers/georeferencing

Redraft ______
mylar separates of lines, attributes

ANNOTATION

GEOREFERENCING
Subdivide map-coordinate min/max

Identify setup coordinates

POINT NO.	AXIS	LAT/LONG INTERSECT	EQUIVLENT COORDINATE	HIGHLIGHT ON MAP
1	X	______	______	[ ]
1	Y	______	______	
2	X	______	______	[ ]
2	Y	______	______	
3	X	______	______	[ ]
3	Y	______	______	
4	X	______	______	[ ]
4	Y	______	______	
5	X	______	______	[ ]
5	Y	______	______	
6	X	______	______	[ ]
6	Y	______	______	

POINTS, LINES, AREAS
Annotation

Highlighting

FIG. 4. Preprocessing – preparation for digitizing.

Create A Digital File

DIGITIZE

Project ____________ Preparer ____________

Date to be completed ____________ Date of completion ____________

Theme ____________

Map ____________ Sheet/Panel ____________

Filename ____________

SETUP

PROCEDURE
method: ____________
number of points: ____________

TOLERANCES
deviation magnitude ____________
acceptable RMS error ____________

SPECIAL INSTRUCTIONS

ACCEPTANCE TESTS

X Axis			Y Axis		
Reference	Measured	Deviation	Reference	Measured	Deviation

X-Deviation magnitude
Positive + ____
Negative - ____
Accuracy +- ____

Y-Deviation magnitude
Positive + ____
Negative - ____
Accuracy +- ____

RMS Error ____

DIGITIZE

PROCEDURE
mode ____________
snap tolerance ____________

SPECIAL INSTRUCTIONS
Lines hours ________
Attributes hours ________

ERRATA

TAKE DOWN

Map Stored/Archived [ ] Location ____________ Digital File Backed Up [ ] Location ____________

FIG. 5. Digitizing – create a raw digital file.

Quality Assurance and Reformat

EDIT

Project ____________ Preparer ____________

Date to be completed ____________ Date of completion ____________

Theme ____________

Map ____________ Sheet/Panel ____________

Filename ____________

QUALITY ASSURANCE

TOLERANCES
node snap radius ________
weeding close points ________
fuzzy tolerance ________
dangle length ________

SPECIAL INSTRUCTIONS

ACCEPTANCE TESTS

Attribute accuracy ____________
ALL FILE CODES IN MASTER GEOCODES LIST ☐

Logical consistency ____________
CLEANED DATA ☐
SOFTWARE NAME AND VERSION

TOPOLOGY BUILT ☐
SOFTWARE NAME AND VERSION

Completeness ____________
ALL CODES IN DATA DICTIONARY REPRESENTED IN FILE ☐

REFORMAT OF DATA

DATA COMPRESSION
SOFTWARE NAME AND VERSION

DATA BACKUP / STORAGE

FIG. 6. Edit – quality assurance and reformat.

TABLE 1. ACCEPTANCE TESTS FOR POSSIBLE INCLUSION IN TEST PLAN

Test For	Test
positional accuracy	comparison to ground; to source of higher accuracy
coordinate representation	comparison to known range
digitizer accuracy	calibration
setup accuracy	deviation magnitude;RMS
redundant points	some fraction of the maximum coordinate distance
attribute accuracy	machine check: that all file codes are in master geocodes list
	detailed graphic check,check plot
	polygon overlay
logical consistency	visual inspection on graphic data (check plot) and/or machine check for: a)line intersection, b)areas completely described, c)over/undershoots, d)no lines entered twice, e)polygons too small or too close
	machine check, topology tests: a) chains intersect at nodes, b) cycles of chains and nodes are consistent around polygons, c) inner rings embed consistently in enclosing polygon
completeness	machine check: that all codes in the data dictionary are represented in file

To maximize the use of digitizing personnel, the maps may be redrafted into separate sheets. One or more products can be made for line work and, possibly, separate sheets can be made for attributes. This is a major reformation of the data, while highlighting or annotating features on the map may be a similar exercise but not as extreme. Other features may include coordinate points on the map that will be used to establish the grid on the digitizer bed itself. Other points may aid in testing the grid as setup on the digitizer.

The digitizing form (Figure 5) includes the major components for setup, digitization, and takedown. Setup procedures use georeferencing previously established on the maps during preprocessing. Setup accuracy can be measured using assigned methods and standards from the test plan. Methods include measuring magnitudes of deviation from reference coordinates or measuring the root-mean-square error (RMS) of the setup points. Standards might be set according to limits imposed by the positional accuracy of the map. For example, the U.S. Geological Survey utilizes a horizontal accuracy of 0.85 millimetres (1/30 of an inch) on the map or approximately 20.3 metres (67 feet) on the ground.

With a coordinate base established across the digitizer surface, all features on the map can then be "digitized." Again, methods and standards can be assigned from the test plan and a record can be kept of hours to perform the task and any errata encountered during the process.

For instance, digitizing mode and a snap tolerance can be defined on the form. The mode refers to coordinate data capture. It can occur one point at a time, or in a continuous stream of coordinates. The line work can be digitized as arcs and nodes, or as an entry of coordinates without designating line intersections or nodes. A "snap" tolerance is set to determine how close a digitized point has to be to a node before it can be considered identical to that node. Errata may be defined as errors in the map or from preprocessing steps and can be noted on the sheet. The audit log also addresses functions such as takedown of the map and proper storage and archive of the map and digital data.

Figure 6 shows the edit form, where most of the digital data quality assurance is documented and reformatting is performed.

The edit stage consists of two phases, including (Marble, 1984) one within a data set and the other between data sets. The latter is described as a database level check, the former as a multilevel check. Test plans are implemented through the edit form and documented as to their completion and correction of inadequacies in the data set.

Correction of inadequacies can be accomplished using one or more of three general procedures: (1) browse/interactive edit, (2) machine check, and (3) graphic check/check plot.

Finally, topology can be built. This is the process of associating line segments with the polygons or areas that they describe. Topological consistency can then be reported by machine check according to the tests for topology mentioned in the test plan. Their completion may serve as a minimum quality statement about some aspects of the data production, and could mention software name and version number used to further document the conditions.

The forms may be altered to suit the individual application. Should any such device be employed, it should be evaluated with the following points: (1) how well do they describe the production process, (2) is the boundary of the study realistic, (3) does the user know more about the production process after using the forms, (4) how well do the forms define quality assurance as well as when to apply it, (5) do they offer an opportunity for documentation in all places it is needed, and (6) how does this model fit into existing models of spatial data base production.

CONCLUSIONS

Recording production details in a systematic fashion can provide valuable information to the user related to quality. In addition, the forms offer a guide for completing the tasks included in the process.

REFERENCES

American Congress on Surveying and Mapping, 1988. Digital Cartographic Data Quality. *The American Cartographer*. Vol. 15, No. 1, pp. 131–135.

Hudson, W. D., and M. J. Krogulecki, 1987. A Comparison of the Land Cover and Use Layer of Three Statewide Geographic Information Systems in Michigan. *Proceedings of GIS '87, Second Annual International Conference, Exhibits and Workshops on Geographic Information Systems*, San Francisco, California, 276–287.

Marble, D., J. Lauzon, and M. McGranaghan, 1984. Development of a Conceptual Model of the Manual Digitizing Process. *Proceedings of the 1984 International Symposium on Spatial Data Handling*, Vol. 1: Geogräphisches Institut, Abtielung Kartographie/EDV, Universität Zürich-Irchel, Zürich, Switzerland.

Moellering, H. (ed.), 1987. *A Draft Proposed Standard for Digital Cartographic Data*. Report #8, National Committee for Digital Cartographic Data Standards. Ohio State University, Columbus, Ohio.

(Received 9 March 1989; accepted 12 April 1989; revised 2 August 1989)