

# TOWARD IMPLEMENTING A FORMAL APPROACH TO AUTOMATE THEMATIC ACCURACY CHECKING FOR DIGITAL CARTOGRAPHIC DATASETS

Barbara Bicking and Kate Beard  
National Center for Geographic Information and Analysis  
and Department of Surveying Engineering  
University of Maine, Boardman Hall, Orono, ME 04469-5711  
207/581-2187 and 2147; Fax 207/581-2206  
bickingb@osprey.umesve.maine.edu and beard@grouse.umesve.maine.edu

## ABSTRACT

The automation of the map data conversion process is one of the key issues in GIS database construction. The accuracy of the resulting digital cartographic datasets—used to provide base reference maps in the GIS application domain—is directly related to the extent of automation of this process. Currently, the map data conversion process is only partially automated. Automation is limited to the capture and verification of map geometry and topology. At present no conversion software or GIS provides a comparable mechanism for thematic data. Theme attribute coding and verifying remains a manual process. The chosen approach argues that the lack of formal definitions of the content of cartographic data is a fundamental impediment to the automation of the theme attribution and verification. This paper reports on work in progress on implementing the conceptual model developed to capture the map content by way of symbols and symbol relationships. Algebraic specifications of these objects facilitate automated map data conversion, and the assessment and verification of their accuracy and consistency at the time of their capture. A symbol-based cartographic knowledge-base and the formal specifications form the basis of a simple prototype implementation which will demonstrate the automated accuracy checking as an integral part of the data conversion process.

## INTRODUCTION

Many geographic information systems (GIS) currently in use have a cartographic subsystem used to provide base reference maps in the application domain (Ramirez and Lee 1991). The accuracy of such a cartographic database is therefore critical to the GIS database construction. It is also critical to the effective use of the data in analysis and becomes especially important when the analysis results are used in decision-making (Beard et al. 1991). In this context the automation of the map data conversion process is a key issue. It is currently limited to the capture and verification of map geometry and topology. No comparable automated mechanism is available for thematic map data. No computer-based tool exists to perform a comprehensive assessment of the fitness for use of cartographic datasets. The research reported on in this paper works toward that goal.

Research on the accuracy of spatial databases abound in the GIS literature and focuses primarily on positional accuracy, topological consistency, and quantitative thematic accuracy. Considerably less research has been done on factual or qualitative thematic accuracy, although correct thematic data are equally important for a useful database (Brusegard and Menger 1989; Veregin 1989). Thematic data give meaning to spatial objects. They distinguish spatial objects represented by the same geometric type.

And thematic data are essential when querying the database. In the application environment the majority of GIS users thus are primarily concerned with the factual accuracy of thematic data (Millsom 1991). The magnitude of thematic errors, though very important, is a lesser concern. One major source of factual thematic errors is the keyboard entry of thematic data. Attempts to automate this process have proven to be no trivial matter. No research reports exist on the automation of the theme attribution and verification. No software is commercially available to carry out thematic accuracy and consistency checking.

We have developed a conceptual model for automated capture and accuracy and consistency checking of thematic map data. The model uses the methods of algebraic specification to formally define the component domains of map symbols, including the thematic data at the class level, and symbol relationships in terms of their behavior (Bicking 1994; Bicking and Beard 1994). This paper reports on the implementation work in progress and its overall pursuit: the development of a thematically accurate and consistent cartographic knowledge-base rich in detail with an object-based accuracy and consistency checking mechanism.

The paper is organized as follows. The next section discusses why map data conversion by way of symbols and symbol relationships is more suitable for increased automation than the conventional method. The third section outlines the implementation strategy and includes examples of algebraic specifications. Section Four concludes with comments on the status of the implementation and future work.

## **TWO MAP DATA CONVERSION METHODS**

The automation of the map data conversion process is one of the key issues in GIS database construction. The accuracy of the resulting digital cartographic datasets—used to provide base reference maps in the GIS application domain—is directly related to the extent of automation of this process.

### **Current Approach To Map Data Conversion**

The standard map data conversion currently in use is basically a four-step process. In Step One the map geometry and topology are captured. In Step Two geometry and topology are verified in a post-conversion procedure (e.g. the BUILD and CLEAN commands in Arc/Info). Both steps are automated based on formal definitions of the domains are implemented in the digital environment. In Step Three the thematic data are added through keyboard entry and are stored in relational tables. This is a manual process due to the lack of formal specifications of the map theme domain. Equally, the post-conversion accuracy verification of the thematic data is carried out through visual inspection of proof-plots. Both processes require high concentration; they are tedious and prone to error.

The USGS's National Mapping Division developed and uses the Attribute Verification Package (AVP), an automated post-conversion quality control tool, to perform rudimentary checks for correct general purpose attributes of Digital Line Graph data elements (USGS-NMD 1990). The program is limited in scope and effectiveness. Extensive manual and visual thematic accuracy checks remain necessary. A fundamental drawback of the AVP is its lack of a formal base.

### **Formal Symbol-Based Approach To Map Data Conversion**

Maps are powerful communication tools. Their power lies in using symbols to portray real world objects and their thematic and locational relationships to each other. Symbols do encode all relevant information about these objects, which is also deemed sufficiently relevant information for the base reference maps needed in the GIS

application domain. So, rather than decompose the map content into geometric and thematic information, we capture it simultaneously by way of symbols and symbol relationships. The automation of this process is based on formal specifications of all the information encoded in a symbol, namely in its geometric (type, locational and topological), representational (visual variables), thematic, and relational components. Note that most map symbols encode only object class level information at the nominal scale of measurement, although each symbol clearly represents an unique object instance. The inclusion of symbol relationships is of particular benefit: they add richness of detail, more complete relational data, and improve the accuracy of the knowledge-base.

The second core element of our approach is the construction of an object-oriented cartographic knowledge-base. It is structured such that the standardized and finite set of symbols of the 1:24,000 USGS Topographic map series—the selected source document—are stored as objects with their behavior encapsulated in their definition. The development of an object-oriented cartographic knowledge-base has the added benefit of allowing the user to produce consistent, standardized, digital reference map, much like the source document, when combined with a map design knowledge-based system like the one described by Steiner et al. (1989) or Zhan (1991).

The map conversion is then a single-step process: The symbol is captured and immediately verified as an occurrence of the knowledge-base and as accurate and consistent with its definition. The accuracy and consistency checking is part of each symbol's definition. Binary symbol relationships are checked in like fashion at the time of their capture and verified against a set of consistency matrices, based on the theme class of each symbol. The symbol-based method leads to a greater degree of automation of map data conversion and thus to increased accuracy and consistency of the resulting cartographic datasets.

## IMPLEMENTATION STRATEGY

The task this research addresses is the automated capture and verification of the thematic accuracy and consistency of cartographic datasets. Checks this research will be able to perform are: 'Check the theme accuracy of the specified symbols' 'Check the relationships of symbol x and symbol y based on their theme class', or 'Check if the symbols crossing at point (x, y) have the correct thematic attributes'. The implementation is done in the Arc/INFO environment to be able to test the formalism with cartographic data.

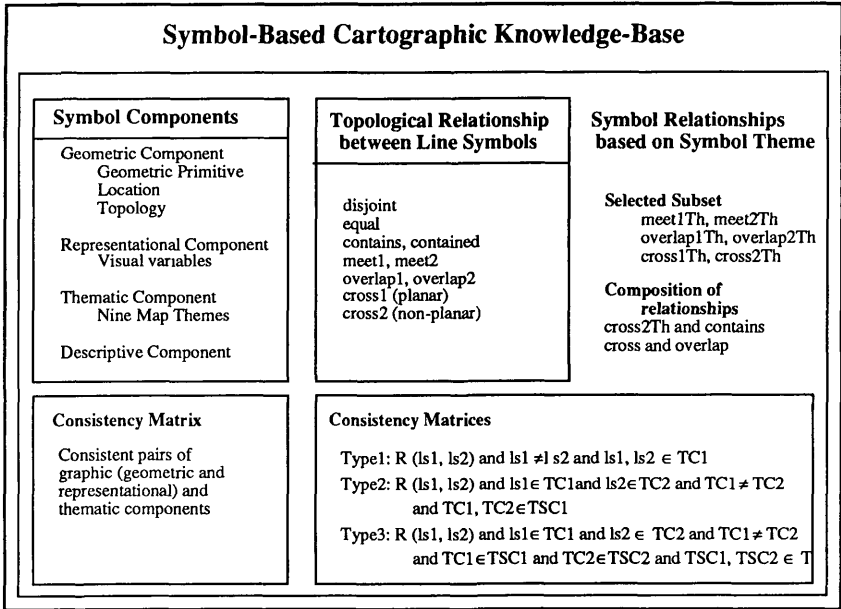
### Knowledge-Base Development

Schemata and formal specifications are used to precisely describe the task and the model properties and behavior. The schema in Figure 1 shows the system and its individual modules. (For schemata for the individual modules and their detailed description the reader is referred to Bicking (1994)). The knowledge-base construction began with defining tables in the INFO module into which all consistency matrices were imported. Unique identifier provide the needed links between the tables and the graphic data.

### Examples of Algebraic Specifications for the Symbol-based Data Capture and Accuracy Checking

We start with a selected subset of symbols for the prototype implementation from the map theme *Transportation—roads, railroads, and linear hydrographic objects*—and include a subset of topological relationships between them—*meet, overlap*, and the planar and non-planar *cross*.

Figure 1: Schema of the Cartographic Knowledge-base with Symbol-Based Thematic Accuracy Checking



Algebraic specifications are written for each module and its components and are combined as needed. Only those properties are specified which are essential to satisfy the task, thus avoiding overspecification (Horebeek and Lewi 1989; Gutttag and Horning 1978; Liskov and Zilles 1975). Note that the specifications—given below in abbreviation—use the construct operations *create* and *assign* in simple and modified form and the observe operations *get* and *consistent* to obtain information about the specified object and its accuracy and consistency. Each component is considered a set. In the context of this implementation each set must satisfy certain constraint conditions, some of which are upwardly cumulative (Beard 1991). These will be revised and removed—if needed—as the knowledge-base grows.

- Condition 1: All sets must be non-empty.*
- Condition 2: All facts about sets, i.e., geometry, topology, graphics, map theme must be explicitly stored in the cartographic database or be inferable from the stored data.*
- Condition 3: A line must not intersect itself, i.e. all interior points have unique values.*
- Condition 4: A line must not close to form a circle, i.e. its start node and its end node are unique also.*
- Condition 5: A line can only have one start and one end node. From this follows that rotaries, meanders, and bifurcations are excluded by definition.*
- Condition 6: All topological relationships are binary relationships.*
- Condition 7: Topological relationships must be explicitly stored in the cartographic database or be inferable from the stored relationships.*

Specification 1 describes the behavior of the sort *symbolType* and imports the sort *geometricPrimitive*, defined elsewhere. It allows to construct a basic graphic symbol object of one of the three primitive geometric types: point, line, and area.

### Specification 1: The Symbol Type

**SORT** symbolType USES geometricPrimitive, Boolean  
**OPERATIONS**  
create:  $\rightarrow$  symbolType  
assignGeomPr: symbolType x geometricPrimitive  $\rightarrow$  symbolType  
getGeomPr: symbolType  $\rightarrow$  geometricPrimitive

The specifications 2 through 4 are derived from the generic specifications for *symbolClass*, *mapDomain*, and *symbol*, respectively. Specification 2 describes the behavior of the cartographic rendering of a line symbol. To this end it USES the sorts symbolType, lTyp, lTex, lSiz, lCol, and lSha—the individual visual variables which have the most expressive power for line symbols. Note that *lTyp* is not a visual variable as defined by Bertin (1983). It is introduced to differentiate between a single line and two parallel lines, called *casing* in cartographic terminology.

### Specification 2: The Line Symbol Class

**SORT** lineSymbolClass USES symbolType, lTyp, lTex, lSiz, lCol, lSha, INCLUDES visVar with lTex, lSiz, lCol, lSha for visVarVal  
**OPERATIONS**  
createLsc: symbolType x lTyp x lTex x lSiz x lCol x lSha  $\rightarrow$  lineSymbolClass  
getVisVarVal: lineSymbolClass x visualVariable  $\rightarrow$  visualVariableValue

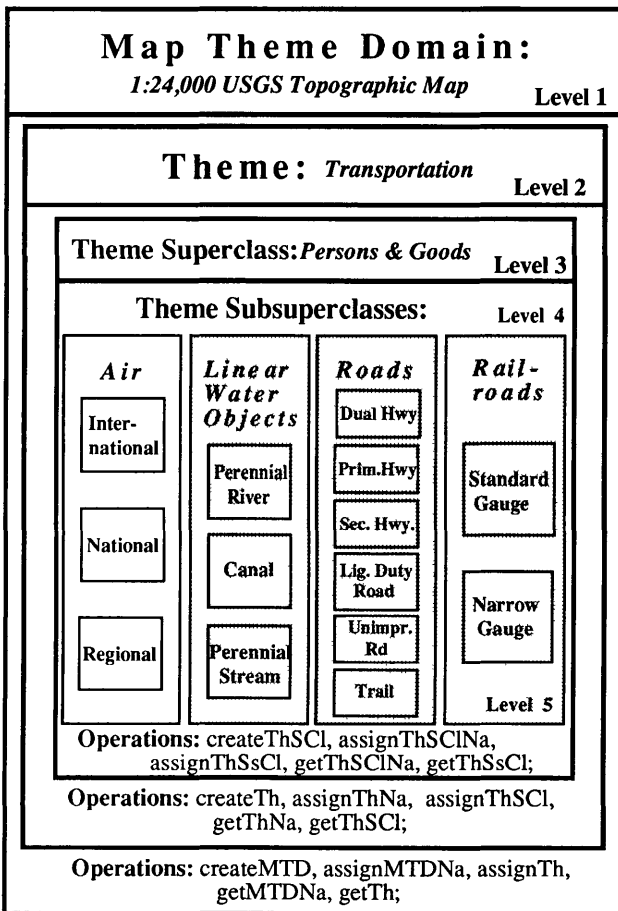
The map theme domain (MTD) of cartographic symbols is hierarchically structured, comprising a finite set of themes (T) followed by a sequence of theme superclasses (TSC) and subclass (TSsC) and theme classes (TC). Figure 2 shows the specific case of the 1:24,000 USGS Topographic map series and a subset of the theme *Transportation*. The specifications for the *mapThemeDomain*, the *theme*, and the *themeClass* are parallel in structure. Specification 3 is representative for them.

### Specification 3: The Theme Class

**SORT** themeClass USES themeClassName, themeClassMember  
**OPERATIONS**  
createThCl:  $\rightarrow$  themeClass  
assignThClNa: themeClass x themeClassName  $\rightarrow$  themeClass  
assignThClMem: themeClass x themeClassMember  $\rightarrow$  themeClass  
getThClNa: themeClass  $\rightarrow$  themeClassName  
getThClMem: themeClass  $\rightarrow$  themeClassMember

The linking of the representation and content domains is achieved with the object *symbol*. The symbol is composed of a *lineSymbolClass* component—it comprises the geometric and the graphic domains from Specification 1 and 2—and the *themeClass* component from Specification 3. Its accuracy and consistency is checked against a consistency matrix composed of accurate and consistent pairs of *lineSymbolClass* and *themeClass*.

Figure 2: Schema of the Map Theme Domain and *Transportation* Theme



Specification 4: The Symbol

**SORT** symbol USES lineSymbolClass, themeClass, matrix, Boolean

**OPERATIONS**

createSym: lineSymbolClass x themeClass —> symbol

getLsc: symbol —> lineSymbolClass

getThCl: symbol —> themeClass

consistent: symbol x matrix —> Boolean

Specification 5 describes in generic terms the behavior of the thematically constrained symbol relationships. It imports the sort *relationshipName* and the sort *symbol* from the previous specification to create a symbol relationship.

## Specification 5: The Symbol Relationship

<b>SORT</b>	symbolRelationship USES symbolRelationshipName, symbol, matrix, Boolean
<b>OPERATIONS</b>	
createSymRel:	symRelName x symbol x symbol —> symbolRelationship
getSymRelNa:	symbolRelationship —> symRelName
getFirstSym:	symbolRelationship —> symbol
getSecSym:	symbolRelationship —> symbol
consistent:	symbolRelationship x matrix —> Boolean

All specifications require slight syntactic adaptations to facilitate their compilation in the Macintosh version of *Gofer*, a functional programming language (Jones 1994). This work is currently in progress. Their integration into the Arc/INFO environment will follow this implementation step.

### CONCLUDING REMARKS

The research presented here emphasizes the importance of factually accurate and consistent thematic data in cartographic and GIS databases. It takes an object-oriented approach to data conversion and verification with the accuracy and consistency checking as part of the object, i.e. symbol and symbol relationship, definition.

Based on the insights obtained from Mark and Xia's (1994) implementation of the 9-intersection model for spatial relationships developed by Egenhofer and Franzosa (1991), we will implement the formalism for line-line relationships in Arc/INFO (Egenhofer 1993).

Once the knowledge-base is established the data conversion for the selected subset of line symbols and symbol relationships will be carried out. The implementation of the described approach requires that the map content is captured by way of symbols, either through laser scanning with color and line pattern recognition capability, or line tracing, or digitizing. Digitizing was chosen because such a laser scanner is currently not available to us. Furthermore, by focusing on digitizing we eliminate any errors associated with this technology. These will need to be accommodated in a full-scale program development. The symbol-based data will be stored in thematic layers in Arc/INFO and each layer will be verified prior to their vertical integration into a *Transportation* layer. Since the implementation is work in progress, we expect revisions and adaptations of the what is done to-date.

### REFERENCES CITED

- Beard, M.K. 1991: *Constraints on Rule Formation*. In: Butenfield, B., McMasters, R. (eds.) 1991: *Map Generalization. Making Rules for Knowledge Representation.*, 121-135 (Chap. 7), London.
- Bicking, B. 1994: *A Formal Approach To Automate Thematic Accuracy Checking For Cartographic Data Sets*. MSc Thesis, NCGIA and Dept. of Surveying Engineering, University of Maine, Orono.
- Bicking, B., Beard, M.K. 1994: *A Formal Approach To Automate Thematic Accuracy Checking For Cartographic Data Sets*. In: GIS/LIS '94 Proceedings, Phoenix, AZ, 63-74.
- Bertin, J. 1983: *Semiology of Graphics*. Madison, WI.
- Brusegard, D., Menger, G. 1989: *Real Data and Real Problems: Dealing with Large Spatial Databases*. In: Goodchild, M.F., Gopal, S. (eds.) 1989: *The Accuracy of Spatial Databases*, London, 55-67.
- Egenhofer, M.J., Franzosa, R.D. 1991: *Point-set Topological Spatial Relations*. In: IJGIS, 5, 2, 161-174.

- Egenhofer, M.J. 1993: *Definition of Line-Line Relations for Geographic Databases*. In: Bulletin of the Technical Committee on Data Engineering (IEEE), 16, 3, 40-30.
- Gutttag, J.V., Horning, J.J. 1978: *The Algebraic Specification of Abstract Data Types*. In: Acta Informatica, 10, 1, 27-52.
- Horebeck, van I., Lewi, J. 1989: *Algebraic Specifications for Software Engineering*. Berlin, Germany.
- Jones, M.P. 1994: *MacGofer, version 2.30*. Glasgow, GB and New Haven, CT.
- Liskov, B., Zilles, S.N. 1975: *Specification Techniques for Data Abstractions*. In: IEEE Transactions on Software Engineering, SE-1, 1, 7-19.
- Mark, D.M., Xia, F.F. 1994: *Determining Spatial Relations between Lines and Regions in Arc/INFO using the 9-Intersection Model*. In: Proceedings, 14th ESRI User Conference, Palm Springs, CA.
- Millsom, R. 1991: *Accuracy Moves from Maps to GIS - Has Anything Changed?* In: Hunter, G.J. (ed.) 1991: Symposium of Spatial Database Accuracy (Proceedings). Melbourne, Victoria, Australia, 1-16.
- Ramirez, J.R., Lee, D. 1991: *Final Report on the Development of a Cartographic Model for Consistent and Efficient Map Production*. OSU Center for Mapping, Columbus, OH, (Research Report submitted to USGS-NMD).
- Steiner, D.R., Egenhofer, M.J., Frank, A.U. 1989: *An Object-oriented Carto-Graphic Output Package*. In: ASPRS/ACSM Ann. Conv., Baltimore, MD, Tech. Papers, 5, Surveying & Cartography, 104-113.
- USGS-NMD 1990: *Attribute Verification Package (AVP) - Users Manual*. US Geological Survey: National Mapping Division, Technical Instructions: Software Users Manual. UM1-19-0; (updated 09/90,03/91, 02/93).
- Veregin, H. 1989: *A Taxonomy of Error in Spatial Databases*. NCGIA-Santa Barbara, Technical Paper. 89-12.
- Zhan, F. 1991: *Structuring the Knowledge of Cartographic Symbolization - An Object-Oriented Approach*. In: ACSM/ASPRS Ann. Conv., Baltimore, MD, Tech. Papers, 6, Auto-Carto 10, 247-260.