Section 9

Database, Standards, and Modelling
Base de données, standards, et modélisation

The Integrated Data Model for the Spatiotemporal Phenomena
Morishige Ota

Semi-Automatic Change Detection for Updating of Large Scale Spatial Databases
Roshanak Darvishzadeh - travel award winner

Spatio-Temporal Formalization:
Incorporating the Concept of Change in Cartographic Modeling
George Panopoulos, Marinos Kavouras

Archiving the Cartographic Record: The Canadian Experience
Betty Kidd

Building Aerial Photograph Referencing System and Providing Service on the Net
Noboru Kobori

On the automatic retrieval of updates in geographic databases based on geographic data matching tools
Thierry Badard

A fully integrated Geospatial Data Warehouse
François Létourneau

Interactive, User-Oriented, Error Modelling Designed to Support Rural Spatial Analysis and Information Production
David Fraser

Active Object Techniques for Production of Multiple Map and Geodata Products from a Spatial Database
P.G. Hardy

Application of Method of Object-Oriented Analysis for Multi-Detailed Representation of Cartographic Data
Michael O. Govorov, Aleksey G. Khorev, Elena L. Kasjianova
Multi-scale representation of raster drainage network based on model
Yaolin Liu and Martien Molenaar

Production and “Intelligence” of Raster Maps
C. Armenakis, E. Siekierska, F. Hegyi and P. Pilon

Enhancing a Database Management System for GIS with Fuzzy Set Methodologies
Emmanuel Stefanakis and Timos Sellis

A Model-based Approach to Geological Map Representation
Boyan Brodaric

Maintaining Parallel Map-Datasets
Peter Højholt

A Case Study of Propagating Updates between Cartographic Data Sets
Lars Harrie and Anna-Karin Hellström

Managing Québec’s Cadastral Database
Michel Morneau, Daniel Roberge

Implications cartographiques de la rénovation cadastrale au Québec
Yaïves Ferland

Object-Oriented Modeling Approach on Geo-Databases and Cartographic Products
Leena Salo-Merta

Image data compression methods and the assessment of map complexity
David Fairbairn

Modelling Objects for Screen Map Design
Mats Dunkars, Anna Bergman

A Spline Function is Applied in Map-Projection Transformation
Yinxiang Yuan
The Integrated Data Model for the Spatiotemporal Phenomena

Morishige Ota
Kokusai Kogyo Co., Ltd., Tokyo, Japan
mshige@kkc.co.jp

Abstract
Geographic Information Systems have the mechanism to store geographic features systematically in the database. To realise this mechanism, many spatial data models have been developed and applied in the history of GIS. But, most of them are “spatial” and they are independent from “time”. It is still difficult to describe the time related phenomena like “motion”, “deformation”, “substitution”, “fission” and “fusion”. This paper presents the geometric and topological spatiotemporal data models to describe these phenomena.

Introduction
This paper presents the integrated data models to describe spatiotemporal features which are extended the 2-D vector spatial and 1-D temporal models. Many spatial data models have been proposed and adopted into commercial GISs. Most of them store the temporal characteristic in the database as a thematic attribute. They do not have a capability to manage spatiotemporal features despite that the real world phenomenon appears, disappears, substitutes as time passes.

Recently, the demands for the Spatiotemporal GIS is increasing. And, many attempts and experiments have been done. Armstrong (1988) proposed the Snapshot Model; the heap of layers of which include the features on the same time slice. Langran and Chrisman (1988) discussed the Space-Time Composite Model; each object in the model has own time stamp. These approaches have significant data redundancy, and it is difficult to separate an individual geographic feature from the database. The space-time composite approach requires re-construction of thematic and temporal attribute tables whenever operations involve any changes in spatial objects. Thus, geographic entities tend to be decomposed into fragments of spatial objects [Yuan, 1996]. Worboys (1992) proposed the Object-oriented Data Model. In this model, the feature is represented by the cylindrical solid figure. They do not have disadvantages like the Snapshot Model or the Space-Time Composite Model, but it can not describe dynamic phenomena and does not have the topological structure which most of GISs adopt as the spatial data model. Pigot and Hazelton (1992) discussed the theoretical Dynamic 4-D model. But this idea is theoretical and more complex than other data models and it is difficult to implement [Peuquet, 1994].

This paper aim to provide (1) the 2-D spatial data models as the basis of the spatiotemporal data models. (2) the 1-D temporal data models and (3) the spatiotemporal data models which are the extension from 2-D spatial and 1-D temporal models. The next section provides the 2-D geometric and topological data models. The 3rd section proposes the spatiotemporal data models, which are obtained by the extension from the spatial and temporal data models. The last section summarises the results of this paper and indicates the future directions and extensions.
Basic Data Models

This section investigates the fundamental spatial and temporal models as the basis of the spatiotemporal data models.

2-D spatial data model

There has been many spatial data models already proposed and applied in the society [Peuquet, 1984; Egenhofer and Herring, 1991; FIPS PUB 173, 1992]. 2-D Spatial Data Model represents an spatial attribute of the feature as a spatial geometric or topological primitive in the 2-D Euclidean space. The fundamental spatial geometric primitives are a point, a curve and a surface. A point represents the 0 dimensional position (x, y). A curve represents the 1 dimensional object. It is a simple list of points in practice. “Simple” in this case, means that there is no self-intersection or self-tangency. A surface is the 2 dimensional object, consists of one outer boundary and may have inner boundaries. These boundaries are closed curves called a ring in this paper. Inside of a ring is disjoint from others. The rotation direction of the ring is assumed as counterclockwise in this paper. The 2-D topological primitives are also imagined in 2-D Euclidean space. They represent a topological characteristic of the feature. The 2-D topological primitives are a node, an edge and a face. They can be defined by their boundaries and/or co-boundaries. A node is the 0 dimensional primitive and bounds an edge. An edge is the 1 dimensional primitive, bounds a left and/or right face and is bounded by start and end nodes. A face is the 2 dimensional primitive, bounded by a set of edges. A face may have holes as well as a surface. A hole of face is called a loop in this paper. A loop is a topological characteristic of a ring. It is a simple closed edge, has no boundary, disjoint from others.

The 2-D spatial model can be seen from Table 1. There are two types of parent primitives called Spatial geometry and Spatial topology. Both of them have 4 sub-primitive types defined above.

Table 1. 2-D spatial model

<table>
<thead>
<tr>
<th>Parent primitives</th>
<th>Primitives</th>
<th>Definitions</th>
<th>Constrains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial geometry</td>
<td>Spatial point</td>
<td>sp = (x, y)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial curve</td>
<td>sc = &lt;sp&gt;, A number of sp &gt; 1, sc is simple</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial ring</td>
<td>sr = sc,</td>
<td>sp0 = spn-1, n &gt; 2, counter-clockwise</td>
</tr>
<tr>
<td></td>
<td>Spatial surface</td>
<td>ss = (outer_sr, {inner_sr})</td>
<td>sr is disjoint from others</td>
</tr>
<tr>
<td>Spatial topology</td>
<td>Spatial node</td>
<td>sn = {(sign, se)}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial edge</td>
<td>se = (start_sn, end_sn, left_sf, right_sf)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial loop</td>
<td>sl = (left_sf, right_sf)</td>
<td>counter-clockwise</td>
</tr>
<tr>
<td></td>
<td>Spatial face</td>
<td>sf = (&lt;(sign, se)&gt;, {(sign, sl)})</td>
<td>The outer boundary is counter-clockwise, sl is disjoint from others</td>
</tr>
</tbody>
</table>
(a, b, ..., k): a tuple of a, b, ..., k
{a, b, ..., k}: a set of a, b, ..., k
<a>: an ordered set of a, \( i=0..n-1 \)
sign: the (rotation) direction of a primitive (+ or -)

A topological primitive may have one to one relationship between equivalent geometric primitive to realise it geometrically in the Euclidean space.

**1-D Temporal Model and causality**

Entities in the real world appear, change and disappear in the certain period of time. Basically, the temporal characteristic of a feature can be described by the existing “period”. The characteristic of a feature being shorter than the granularity of time measure can be called “instant”. As time is 1 dimensional space, the 0 dimensional point (instant) and the 1 dimensional line (period) are defined as the geometric characteristics [Goralwalla, 1998]. Instant is represented by a point of time \( t \). Period is an interval of time \((st, et)\), \( st \) is the start instant and \( et \) is the end instant. Time is a metric space, but the nature of time we usually experience is anisotropic. It moves toward one direction \((st < et)\).

As long as time is the 1 dimensional space, the topological model also can be defined. As can be seen from Table 2, primitives are a temporal node and a temporal edge. A temporal node bounds a temporal edge. The temporal edge is bounded by the start and the end temporal nodes. It has a direction from start to end. The temporal topology relationships may form a directed acyclic graph.

**Table 2. 1-D temporal data model**

<table>
<thead>
<tr>
<th><strong>Parent primitives</strong></th>
<th><strong>Primitives</strong></th>
<th><strong>Definitions</strong></th>
<th><strong>Constraints</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal geometry</td>
<td>Instant</td>
<td>( it = (t) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Period</td>
<td>( pd = (st, et) )</td>
<td>( st &lt; et )</td>
</tr>
<tr>
<td>Temporal topology</td>
<td>Temporal node</td>
<td>( tn = ({previous_te}, {next_te}) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temporal edge</td>
<td>( te = (start_tn, end_tn) )</td>
<td></td>
</tr>
</tbody>
</table>

For example, when the owner of a parcel registered on 1960-02-10 plots it to two parcels on 1998-08-20, one can describe the temporal geometry and topology for these parcels (See Figure 1).
Figure 1. Temporal geometry and topology shown in the parcel fission.

As time is the 1-D directed topological space, the temporal relationships between predecessors and successors are 1 : 1, 1 : n, n : 1, or m : n. The relation 1 : 1 means substitution, 1 : n means fission, n : 1 means fusion, and m : n means the composite of these. The composite relationships may be substitution & fission, substitution & fusion, fusion & fission, or substitution & fusion & fission. Figure 1 shows the example of “fission” [Ota, 1996].

These temporal relationships are the relationships between cause and effect. A cause is an event represented by a temporal node to make new entities as effects. The new entities are represented by temporal edges. Thus, the temporal topological relationship is called as the causality in this paper. The causality constructs the temporal acyclic graph.

**Spatiotemporal Data Models (2-D space + 1-D time)**

Entities in the real world usually move and deform as time passes. When you wish to describe dynamic objects in the database, you need the Spatiotemporal Data Model, because the temporal characteristics of a feature affect its spatial characteristics.
**Spatiotemporal Geometric Model**

As long as time is conjectured to keep the order in the real world, it is independent from space. It can not be treated as one dimension having same characteristics as other dimensions in the n-dimensional space. This paper defines the spatiotemporal data model as a model in 2-D space + 1-D time.

A motion and a deformation are the inherent phenomena in the space-time. A motion can be described by a locus of the representation point of the feature $f$. The locus of $f$ is represented by the following function. Where, $p(t)$ is a function to show the position of representation point at time $t$. It moves as time increases from $s_t$ to $e_t$.

$$\text{motion}(f) = p(t), \ s_t < t < e_t$$

The deformation is defined as a deferent motion of the feature $f$. Let $q(t)$ be a position of one relative position from p at time $t$. The deformation of $f$ is represented by the following function.

$$\text{deformation}(f) = q(t), \ s_t < t < e_t$$

Thus the motion with deformation is defined by,

$$\text{motionWithDeformation}(f) = (p(t) + q(t)), \ s_t < t < e_t$$

Where, $r(t)$ is a composite of motion and deformation.

Mutations of the geographic features involve in the changes of topological types. For example, the growth of a vegetation colony from one seed may be described as the deformation from a point to a surface, and then from a surface to a set of surfaces. These topological deformations are a spatiotemporal sequence of topological types in the feature. In this paper, a motion with deformation not to change the topological type is called a simple motion. A motion with deformation mentioned above is a series of simple motions. It is called a complex motion. A simple motion is a periodical geometric primitive.

$$\text{simpleMotion}(f) = (s_P, \ sr(t), \ e_P), \ s_t < t < e_t$$

$s_P$ and $e_P$ are boundaries of the simple motion. The boundary is the instantaneous geometric primitive. $sr(t)$ is a dynamic function of a motion with deformation not to change it’s topological type from $s_t$ to $e_t$. It is open at $s_t$ and $e_t$, and converges continuously to $s_P$ and $e_P$. When $sr(t)$ is a motion of the point, it is called periodical point, and $s_P$ and $e_P$ must be an instant point. When $sr(t)$ is a motion of the curve, it is called a periodical curve, and $s_P$ and $e_P$ is the instant point or the instant curve. In case of a simple motion of surface, it is called a periodical surface, and it’s boundary may be the instant point, the instant curve or the instant surface. In practice, $sr(t)$ can be implemented as the time series of same type instant primitives. In this case, the adjacent primitive must be, at least, surjective toward the time direction.

### Table 3. Spatiotemporal geometry model

<table>
<thead>
<tr>
<th>Parent primitives</th>
<th>Primitives</th>
<th>Definitions</th>
<th>Constrains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instant geometry (ig)</td>
<td>Instant point</td>
<td>$ip = (it, p)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instant curve</td>
<td>$ic = (it, c)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instant surface</td>
<td>$is = (it, s)$</td>
<td></td>
</tr>
<tr>
<td>Periodical geometry (pg)</td>
<td>Periodical point</td>
<td>$pp = (\text{start}<em>\text{ig}, \langle ip\rangle, \text{end}</em>\text{ig})$</td>
<td>$ig = ip$</td>
</tr>
<tr>
<td></td>
<td>Periodical curve</td>
<td>$pc = (\text{start}<em>\text{ig}, \langle ic\rangle, \text{end}</em>\text{ig})$</td>
<td>$ig = ip$ or $ic$</td>
</tr>
<tr>
<td></td>
<td>Periodical surface</td>
<td>$ps = (\text{start}<em>\text{ig}, \langle is\rangle, \text{end}</em>\text{ig})$,</td>
<td>$ig = ip$, $ic$ or $is$</td>
</tr>
</tbody>
</table>
The spatiotemporal geometric model can be obtained by the of spatial and temporal data models. The parent primitives in this case are Instant geometry and Periodical geometry. Instant geometry (ig) has 3 derived primitives. They are the instant point (ip), the instant curve (ic), and the instant surface (is). There is no definition of the instant ring. Because a ring appears in a surface only. Periodical geometry (pg) also has 3 derived primitives. They are the periodical point (pp), the periodical curve (pc), and the periodical surface (ps). They are representations of simple motions.

**Spatiotemporal Topological Model**

The spatiotemporal topology model represents topological characteristics of spatiotemporal features. It consists of 2 parent primitives and 8 derived primitives.

One of the parent primitives is the spatiotemporal node (stn); a spatial topological primitives having a characteristic of the temporal node. The other is the spatiotemporal edge (ste); a spatial topological primitives having a characteristic of the temporal edge. The spatiotemporal node has 4 derived primitives. They are STN node (stnn), STN edge (stne), STN loop (stnl), and STN face (stnf). The spatiotemporal edge also has 4 derived primitives. They are STE node (sten), STE edge (stee), STE loop (stel), and STE face (stef). The spatiotemporal topology model can describe the topological characteristics of the simple motion and it’s complex.

**Table 4. Spatiotemporal Topology Model**

<table>
<thead>
<tr>
<th>Parent primitives</th>
<th>Primitives</th>
<th>Definitions</th>
<th>Constrains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatiotemporal node(stn)</td>
<td>STN node</td>
<td>stnn = ((sign, ste), (sign, stne)), ste = sten, stee or stef</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STN edge</td>
<td>stne = ((sign, ste), start_stnn, end_stnn, left_stnf, right_stnf), ste = stee or stef</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STN loop</td>
<td>stnl = ([(sign, stel)], left_stnf, right_stnf)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STN face</td>
<td>stnf = ([(sign, stef)], &lt;(sign, stne)&gt;, (sign, stnl))</td>
<td></td>
</tr>
<tr>
<td>Spatiotemporal edge(ste)</td>
<td>STE node</td>
<td>sten = (start_stnn, end_stnn, {&lt;sign, stee)})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STE edge</td>
<td>stee = (&lt;sign, start_stn&gt;, &lt;sign, end_stn&gt;, &lt;start_sten&gt;, &lt;end_sten&gt;, left_stef, right_stef), stn = stnn or stne</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STE loop</td>
<td>stel = (start_stnl, end_stnl, left_stef, right_stef)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STE face</td>
<td>stef = (start_stn, end_stn, &lt;sign, stee&gt;), (sign, stel), stn = stnn, stne or stnf</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 shows the growth of a vegetation colony as stnf. This stnf has a stnf at the start time, a stnf at the end time as outer boundaries in temporal direction; an ordered set of stee as an outer boundary in spatial direction; and one stel as an inner boundary.
Conclusion

This paper discussed about the spatiotemporal data models integrated from fundamental 2-D spatial and 1-D temporal models. At first the 2-D spatial and 1-D temporal data models were presented. The 2-D spatial model consists of geometrical and topological primitives and includes a ring and a loop. The 1-D temporal model also consists of the geometrical and topological primitives. Then, the spatiotemporal models were presented as the integration of the fundamental spatial and temporal models.

This paper aimed to construct the implementable data models for the practical usage. The father discussions are required to improve the implementation:

1. how to describe the spatiotemporal relationships between primitives.
2. how to portray the spatiotemporal dynamic data effectively on the screen or the paper;
3. how to identity the entity in the real world which characteristics dynamically change.

Acknowledgement

Special thanks go to Masao Iri for his help and suggestions.
References

Semi-Automatic Change Detection for Updating of Large Scale Spatial Databases

Roshanak Darvishzadeh - travel award winner
International Institute for Aerospace Survey and Earth Sciences (ITC)
P. O. Box 6, 7500 AA Enschede, The Netherlands
Email: darvish@itc.nl

Abstract

The problem, which is addressed by this study is the lack of up-to-date and correct information about the development and changes of unplanned areas in Dar Es Salaam to guide urban planners, urban decision makers, and managers.

In Dar Es Salaam new expansions due to development of unplanned area and urbanisation causes maps to become out-dated rapidly. For development and planning, if large scale maps are to be of any value they must be as up-to-date as possible; that is they must also be changed in accordance with changes in the areas. Therefore, there is a need to revise old maps using a fast, efficient, and effective method. One solution to this problem can be using digital aerial photographs when revising the old maps. This can be faster, cheaper, and more convenient. Approaches used for change detection are: semi-automatic change detection using on-screen digitizing and digital classification and segmentation methods.

On-screen digitizing of changes shows that (as the case study covered only one aerial photograph) this method is fast, convenient, and (depend to the expert and application) complete.

The result of segmentation and classification for finding the changes of unplanned areas in Dar Es Salaam, shows 73% accuracy when the result of on-screen digitizing is used as a reference map for evaluation and by using existing knowledge such as roads and object geometry contained in the GIS database. Classification and segmentation can be used to derive information from an aerial photograph. The nature, variety and usefulness of such an information would have to be considered according to the project at hand, i.e., the structure of the area, the type of available aerial photograph, and the type of information to be extracted.

For evaluation of the segmentation and classification method with on-screen digitizing, several criteria have been considered: accuracy, completeness, time efficiency. Although the result of on-screen digitizing had a better accuracy and completeness, it should be kept in mind that the knowledge and intelligence of an expert has been applied. In this case study, the time used for on-screen digitizing was less than for classification and segmentation but this may not be the case in a production environment. In the case of dealing with a bigger databases it is obvious that on-screen digitizing may need more time. For classification and segmentation, once the method is developed, is much faster. However, a compromise may be made between the cost of achieving high levels of accuracy and the need for rapid change detection. Classification and segmentation may give a better result where the buildings are larger than those found in the Lake Magomeni area. Also a better result can be achieved with a higher resolution aerial photograph which has a good contrast.
1. Introduction

In Dar Es Salaam new expansion due to development of unplanned area and urbanisation causes maps to become out-dated rapidly. Therefore, there is a need to revise old maps using a fast, efficient and effective method. To revise an old map, it is necessary to detect the changed object in the area. An old map, and old and new aerial photographs are important basic materials for change detection. There have been a number of processes that use analogue aerial photographs for change detection. An example is plotting the changes with a mechanical stereo-plotting machine. But these processes are time-consuming, laborious, and, consequently not economical. Finding a more appropriate solution to this problem has become a genuine challenge. However, because of the possibility that now exists to scan analogue aerial photographs, it is possible to utilise them in a more efficient way.

2. Problem Statement

The problem addressed by this study is the lack of up-to-date and correct information about the development and changes of unplanned areas in Dar Es Salaam to guide urban planners, urban decision makers and managers. Estimation by the City Council of Dar Es Salaam [UNDP, 1992], indicates that approximately 70 percent of the population live in unplanned settlements. These settlements are vast in extent and their population grows so fast that traditional data collection, processing and analysis techniques cannot provide planners and decision makers with a correct view of them. The time interval between the production of the first map with the last map of a series may be too long. As a result a lot of new changes may have occurred again. Traditional map production processes are inadequate for this situation.

One solution to this problem is the use of digital aerial photographs when revising the old maps. This is faster, cheaper, and more convenient. This study has shown how digital aerial photographs can be used to maintain the spatial database and update maps and the relevant information for urban planning.

3. Method

In this section I will discuss the method applied for change detection for updating of a large-scale spatial database in three steps as follows:

3.1 Quality Evaluation of the Existing Digital Spatial Database

Before extraction of new data from the new aerial photograph(1994), it is worthwhile to compare the existing 1992 aerial photograph with the digital spatial database which has been produced from that photo. The reason for doing this is that the 1992 digital data will be used as reference data and the quality and accuracy of the source data should be ascertained before using it for further processing. In this study this job was done by overlaying the aerial photograph on the digital data and comparing some of their details using visual interpretation [Darvishzadeh, 1997].

3.2 Comparison of Old Map and New Photo (Visual Change Detection)

The methodology that has been used in the experiment is illustrated in Figure 1. In this approach, after scanning of the new photo, georeferencing, enhancement and other necessary corrections were done. Then the geocorrected photograph was overlaid with the old digital data and changes detected using on-screen digitizing.
In this case the detected changes were incorporated in the existing digital data. The following sections describe the procedure and the final results.

### 3.2.1 Scanning of Photo

An oblique color aerial photograph of 1994, which was taken with a 35 millimeter camera, was scanned as one band for conversion to digital format. As the scale of the photograph was almost 1/2500, a resolution of 150 dpi (dot per inch) was selected, which gives one pixel for every 40 centimeters on the ground.

### 3.2.2 Georeferencing

After conversion of the image to the ILWIS program, the next step is georeferencing in order to bring the image to the coordinate system of the 1992 digital database. This is very important, as for registration of the image with the digital data they have to originate in the same reference. In the georeference module of ILWIS, 14 well distributed control points in the image and the corresponding control points in the digital data were selected. A projective transformation was used with a sigma of 2.38 pixels (0.95 meter), which is acceptable regarding to the scale of the digital data and photograph.

![Visual change detection approach diagram](image)

**Figure 1:** An overview of the visual change detection approach

### 3.2.3 Overlaying & Change Detection

After georeferencing, the aerial photograph and selection of a part of roof layer (which falls in the same area of photo) from the digital data, they were overlaid together. As a result the buildings which were somehow a change recognized.

In ILWIS, having the aerial photograph as a backdrop image, with overlaying the buildings roofs, the changes were on-screen digitized in a new layer and the digital data has been updated (Figure 2).

Table 1 shows the result of change detection and the number of houses which is detected.
Table 1. Comparison of old and new data in terms of roof changes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Houses</td>
<td>1039</td>
<td>1108</td>
<td>+69</td>
<td>6.6%</td>
</tr>
<tr>
<td>Roof Area (m²)</td>
<td>76000</td>
<td>80126</td>
<td>+4126</td>
<td>5.4%</td>
</tr>
<tr>
<td>Housing Density (building/ha)</td>
<td>35.8</td>
<td>38.2</td>
<td>+2.4</td>
<td>6.7%</td>
</tr>
</tbody>
</table>

3.3 Alternative Method (Analytical Change Detection)

In this approach, after scanning of the new photo, the same procedure for making the geo-corrected image as in section 3.2 was followed. In the later stage, after rectified photo registration, classification and segmentation were used to extract roof changes, using existing knowledge contained in the GIS database. The final result (detected change) will be compared with the result of on-screen digitized data. Figure 3 illustrates this approach.

Figure 2. The changed and unchanged roofs in the area

Figure 3: An overview of the analytical approach
3.3.1 Scanning Photo & Georeferencing

A similar procedure was used as that explained in sections 3.2.1 and 3.2.2 with the exception that the aerial photograph was scanned and processed multi-spectrally. A resolution of 150 dpi was chosen to have the same accuracy in the source data as in section 3.2.

3.3.2 (Image) Test Area

For the next steps in this approach, three subsets were selected from the georeferenced images. The subset images consist of 520 lines and 500 columns, almost covering the middle part of the aerial photograph (an area of 41600 m²). Classification and segmentation was done only for this part. Figure 4 shows the color composite of the subsets for the tested area.

3.3.3 Selection of Test Area from the Digital Database

From the updated (1994) layer of the digital data (updated roofs of buildings) that was detected in section 3.2, the test area was selected (reference map).

The boundary of one of the subsets was identified and only the roofs which would fall in that area were selected. This data will serve two purposes:

1. Knowledge acquisition from existing GIS (e.g., roads) for the knowledge-based classification (improvement)
2. Evaluation of the classification and segmentation results for extraction of roofs.

3.3.4 Pixel Classification

The pixel classification was performed by a standard maximum-likelihood classification using the three bands of the scanned aerial photos. The classification was guided by the updated roof map. The training data which was applied for supervised classification comprises the following 17 classes: roof1, roof2, ..., shadow, and trees. The mean vector and covariance matrix of the seventeen classes were determined from a total of 150 to 300 samples per class. Figure 5 shows the result of the classification.

![Figure 4. Study area.](image1)

![Figure 5. Classification result](image2)

After classification, all the roof classes were combined into one class (roof) and all the shadows and trees into another class (green). There also exist another class, which consists of unclassified pixels. Since there was no information about this class, it is named others (see Figure 6).
This classified map was subjected to accuracy assessment. The accuracy assessment was done by means of a confusion matrix and the overall accuracy (expressed as the percentage of correctly classified pixels) was computed. The overall accuracy of the classification result was assessed by a cross tabulation of the rasterized updated roof map (reference) and the classified map (see Table 2).

However, classification of remotely sensed imagery is effected by e.g. isolated and mixed pixels and spectral confusion of land cover types [Abkar, 1994]. Therefore, in the majority of cases, classification based solely on spectral observation is not sufficiently accurate for extraction of the roofs, especially in this area where reflectance of the road and bare soils is very similar to the roofs. Consequently, no samples were taken for roads and bare soils. This means that a lot of roads and bare soils will be classified as roofs. For the roads this is not a problem, because we know where the roads are from the digital database. For the bare soils the situation is different; we have no prior information about them. This problem can not be resolved by considering bare soils in the classification. If so, the opposite case will happen that means many of roofs will be classified as bare soils.

Table 2: Confusion matrix when reference map was crossed with classified roofs

<table>
<thead>
<tr>
<th>Reference map</th>
<th>Classification results</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>roof</td>
<td>others</td>
</tr>
<tr>
<td>roof</td>
<td>84646</td>
<td>33973</td>
</tr>
<tr>
<td>others</td>
<td>58523</td>
<td>82858</td>
</tr>
</tbody>
</table>

Overall accuracy  = 64.42 %

1st-stage of classification improvement: As a result, existing knowledge from the digital database (e.g., roads) were used to improve the result of the classification. Roads were extracted from the digital data, rasterised and combined with the classified map and the roof map (reference). The overall accuracy of this result is shown in Table 3.

Table 3. Confusion matrix when updated roofs (reference map) with roads was crossed with classified roofs plus roads

<table>
<thead>
<tr>
<th>Reference map</th>
<th>Classification results</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>roof</td>
<td>road</td>
</tr>
<tr>
<td>roof</td>
<td>84646</td>
<td>0</td>
</tr>
<tr>
<td>road</td>
<td>0</td>
<td>30411</td>
</tr>
<tr>
<td>others</td>
<td>43438</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall accuracy  = 70.23 %

2nd-stage of classification improvement: The segments in the final classified map (1-stage) that are unclassified with a size of less than 5 pixels, were detected (Figure 7) and removed (converted to roof class). In Table 4 the overall accuracy of this result after this stage of improvement is shown.
The selection of the threshold (5 pixels) was based on two criteria:
1. The hypothesis that there are no roofs smaller than this threshold (0.8 m²).
2. Choosing a larger threshold will affect opposite in the accuracy result. In fact, first only the segments which are unclassified and their number of pixels is one were removed. Then those which their number of pixels is two has been removed (added to roof class). The same was done for three and four pixels. When this number changes to five the accuracy will suddenly come down instead of going up.

Table 4. Confusion matrix when updated roofs (reference map) including roads was crossed with the improved classified result

<table>
<thead>
<tr>
<th>Reference map</th>
<th>Classification results</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>roof</td>
<td>road</td>
</tr>
<tr>
<td>Roof</td>
<td>85270</td>
<td>0</td>
</tr>
<tr>
<td>Road</td>
<td>0</td>
<td>30411</td>
</tr>
<tr>
<td>Others</td>
<td>43736</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall accuracy  =  70.35 %

The explanation for this is that, although there are some segments on top of the roofs which are unclassified and have five pixels or even more, there also exist some misclassified roofs that have the same situation. By adding to the threshold the area of the misclassified roofs also will increase. So the accuracy, instead of going up, will come down.

The final classified map after the second stage classification is displayed in Figure 8.
3.3.5 Quadtree Segmentation

Human image vision generally tends to divide an image into homogeneous areas first, and will characterize those areas more carefully later [Gorte 1995]. Applying this approach to digital image analysis leads to segmentation, which divides the image into segments that correspond to objects in the terrain. The success of segmentation depends on the availability of: high resolution imagery in such a way that the relevant objects are represented by a significant number of pixels, powerful hardware, and an efficient implementation regarding the size of the remote sensing images [Gorte 1995].

The segmentation program uses a multi band image (in this case 3 bands) as input and gives one segmentation as output. Moreover, information, like object locations, sizes and perimeters can be retrieved. The process has a slight tendency to create segments of regular shapes, according to quadrants. This effect could be completely removed by making the process perform a few iterations, and by increasing the threshold values (readers are referred to technical report, experimental quadtree software, Gorte 1995). In this study the threshold was chosen by trial and error. The result of this segmentation is illustrated in Figure 9.

This result contains many small segments (noise) and mixed pixels. To remove all these the segments which their number of pixel is less than ten have become zero (see Figure 10), meaning that, all the objects with an area less than 1.6 m² will be deleted. Although by doing this some information may be lost, but those information is of no interest in this study as it is here supposed that a roof has at least an area of 2 m².

Further improvement of this result was achieved by the segment-based classification described as follows:

3.3.6 Segment-Based Classification

The aim of segment-based classification is to determine the class (label) of a segment (polygon), of which the geometry is contained in the segmented map using the above segmentation program. Therefore, the pixels within the polygon are identified from the classification result and the class of the polygon is determined from these pixels.

The following steps were taken to arrive from a pixel-based to segment-based classification:

First the georeferenced image is classified using per-pixel classification resulting in a label per-pixel. The segmentation result has been Area Numbered to effect distinct area numbering; connected raster elements with the same value belong to the same segment. The output map from Area Numbering, was superimposed (crossed) with the final classified map in order to get the statistics of pixels within each segment. Subsequently, a frequency Table was established to determine the label of each segment. Then the most occurring (predominant) class label for each segment was calculated (using the Aggregation function) and assigned to the segment (including the unclassified labels).
The classification accuracy was assessed by comparing the output of the segment-based classification with the updated roof map (reference), by calculating confusion matrices and the overall accuracy. This result is presented in Table 5.

Table 5: Confusion matrix when reference map (updated roofs) including roads were crossed with improved segmentation including roads

<table>
<thead>
<tr>
<th>Reference map</th>
<th>Classification results</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>roof</td>
<td>road</td>
</tr>
<tr>
<td>Roof</td>
<td>87939</td>
<td>0</td>
</tr>
<tr>
<td>Road</td>
<td>0</td>
<td>30411</td>
</tr>
<tr>
<td>Others</td>
<td>38654</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall accuracy = 73.33 %

4. Change Analysis

The segment-based classification result was used to analyze the remaining differences between the old roof map (1992) and the roofs as determined with the segment-based classification (as it serves as the best result). The implementation of the segmented-based classification method resulted in a final map which is presented in Figure 11. This result was used to derive the changes. For this reason, the final result of image segmentation (Figure 11) was subtracted from the old (1992 roofs) data (see Figure 12). In the ideal case, the remaining parts of the subtraction should be only the changes. The result of changes found in the on-screen digitizing were overlaid with that result (see Figure 12). Visual analysis of this result shows the following types of error:

1. Existing error in the boundaries of roofs. As seen in Figure 12, there are a large number of small areas (Slivers) indicating errors along the boundaries of roofs. This incorrect classification can be largely explained by spectral confusion and incorrect referencing of roof segment geometry.

2. Existing thematic error in the old roof map (1992 digital data) which served as ground truth. Good examples of this type of error are the detected changes shown in Figure 12.

3. Error due to spectral similarity of roofs with e.g. roads, soils and the surrounding environment. The remaining unclassified and misclassified segments can be explained by this type of error. This type of error, can be seen from the visual interpretation of Figure 12 (large areas). This is an inherent problem of spectral classification and segmentation.
5. Conclusions

The aim of this study was to compare the changes derived from on-screen digitizing with extracted changes by means of classification and segmentation. Based on this study several conclusions may be derived as follow:

5.1 Visual Change Detection

- Scanned high resolution aerial photographs could play an important role in the various fields for spatial information gathering and map updating. They could provide the necessary data to create information for urban planners and decision makers.
- The on-screen digitizing approach which has acquired some popularity in the GIS environment, enables overlaying several layers of raster and vector simultaneously. On-screen digitizing or heads up digitizing, seeks to exploit both the speed of document scanning, which quickly delivers a digital record of a map, and the intelligence of a manual digitizing operator who can interpret entities to be stored in a database.

Monitoring the Changes

It has been observed that in the period of 1992 till 1994 densification has mostly occurred in the area. Referring to Table 1 it can be observed that there still exist some potential spaces for housing.

From the result of the on-screen digitising it can be realised that mostly small houses were added to the area. In this case area, usually new roofing materials have different reflectance, so new roofs were easy to detect. There is not any spatial pattern distinguished in the area for the new roofs. Their pattern continued to be haphazard. To have the real consumption about this spatial pattern a larger area should be considered as a case.

5.2 Knowledge Based Classification and Segmentation

First the per pixel classification was applied to the data. This result was improved using existing knowledge, such as roads contained in the GIS. Later on the result of the classification was further improved to derive a reliable class label, mainly for roofs, by using spatial context information that is given by object geometry through Quadtree based image segmentation (73% when the result of on-screen digitizing is used as the reference map for evaluation). Although errors due to incorrect referencing of roof segment geometry and errors due to spectral similarity of roofs with e.g. roads, soils and the surrounding environment, which is an inherent
problem of spectral classification and segmentation, still exist. The result of segmentation and classification could be improved by adding extra knowledge, for example adding some information about the shapes of the roofs or their textures (usually new buildings because of the new materials have different reflectance). Also this result can be improved by the accurate spatial registration of the image, which is possible through making an orthophoto.

5.3 The Choice between On-Screen Digitizing and Knowledge Based Segmentation Method

Although the result of on-screen digitizing had a better accuracy and completeness, it should be kept in mind that the knowledge and intelligence of an expert has been applied.

In this case study, the time used for on-screen digitizing was less than for classification and segmentation but this may not be the case in a production environment. In the case of dealing with a bigger databases it is obvious that on-screen digitizing may need more time. For classification and segmentation, once the method is developed, is much faster.

However, a compromise may be made between the cost of achieving high levels of accuracy and the need for rapid change detection.

Classification and segmentation may give a better result where the buildings are larger than those found in this photo area. Also a better result can be achieved with a higher resolution aerial photograph which has a good contrast.

References & Bibliography


Spatio-Temporal Formalization: Incorporating the Concept of Change in Cartographic Modeling

George Panopoulos
Cartography Laboratory, Faculty of Rural and Surveying Engineering
National Technical University of Athens
9, Heron Polytexneiou St., 15780 Zographou Campus, Athens – Greece
Tel: +(3+01) 772-2642 Fax: +(3+01) 772-2670 gpani@survey.ntua.gr

Marinos Kavouras
Cartography Laboratory, Faculty of Rural and Surveying Engineering
National Technical University of Athens
9, Heron Polytexneiou St., 15780 Zographou Campus, Athens – Greece
Tel: +(3+01) 772-2731 Fax: +(3+01) 772-2670 mkav@survey.ntua.gr

Abstract
The most recent attempts to develop a coherent spatial theory exhibit the fact that geospatial change is one of the most interesting topics in their research agenda. One of the greatest difficulties in temporal modeling lies in combining the 2D and 3D models with Time. This paper proposes an approach to formalize common concepts between space and time. There is an effort to introduce notions, which traditionally belong to static spatial schemes, to a model able to deal with temporal aspects. Such notions include generalization (not anymore regarded only as a visualization problem between different scales or levels of detail, but as a problem of temporal resolution), temporal topology, spatio-temporal relationships and multiple representations over the course of time. Emphasis is given on defining and utilizing the concept of change, in order to avoid the use of simple temporal snapshots. Change between time instances is expressed through higher level processes, catalogued into the temporal domain of each object. The attributes of this domain include the past and present state of each object (static, under change, ceased) and the changes themselves, linked to each other. The aim is to formalize these concepts into a unified model, able to tackle not only spatial aspects of phenomena, but their change as well. Such a formal model may well serve as a standardization schema, acting as a specification for the conceptual modeling of spatio-temporal data. Finally, from the wide variety of applications requiring advanced spatio-temporal modeling, cadastre applications were selected for the evaluation of the proposed methodology. Data for such applications are subject to constant change, making their representation a challenging task.

1. Introduction
The importance of representing the temporal aspects of spatial information has been widely recognized. We live in a dynamic, constantly changing world. While Spatial Information Systems are widely used to capture geometric and thematic information of our surroundings, they fail however to include the temporal aspects which may exist in a dynamic environment. Yet, needs for temporal applications in spatial information systems
rise everyday. The element of time is of great importance in GIS, mainly in socio-economic applications, cadastre, planning and monitoring of dynamic geoscientific phenomena. Still, the modeling of temporal change is no simple issue. Although extensive research has been conducted worldwide, it has met with partial success and has made slow progress. The main difficulties lie in the complexity of grasping the concepts of time and change and then applying these concepts to existing spatial models. Current efforts in modeling include a wide variety of spatio-temporal models, as the simple model of frequently taken data snapshots, the space-time composite data model, time stamping, the history graph model, the three-domain model, moving objects model and many object oriented models [Abraham and Roddick, 1996; Langran, 1992]. However, most of these models have not been implemented in commercial GIS nor have they been utilized to face specific spatio-temporal applications.

The first aim of this paper is to propose ideas on a spatio-temporal model, which incorporates the temporal aspects of objects, in order to enable the representation of any transition they may undergo during the course of time. The model focuses on large-scale representations and more specific, cadastre applications. Emphasis is put on conceptual issues related to geospatial data (including representation, spatial modeling, integration, communication and interaction), having object-oriented capabilities in mind. Thus the model would be at a generic level, formalized in such a way as to act as specification for the modeling of applications of this kind. The proposed ideas do not refer to a new model, one made «from scratch», but rather to an enhancement and an integration of approaches found in various other spatio-temporal models into a more specific and coherent way of coping with spatial change. The second aim of the paper is to examine the introduction of common spatial notions, as generalization, relationships and multiple representations, into the proposed model. The model should be able to represent the temporal aspects of these traditionally spatial concepts. Adding to these notions the element of time and describing the way they are affected by the constant changes of the objects is quite a daunting task. We seek a suitable model to cope with this kind of representation.

Bearing all this in mind, the present paper is structured as follows: Section 2 follows with a presentation of existing spatio-temporal models that include certain approaches, which are utilized later. Section 3 concerns our approach to spatio-temporal modeling. There the basic principles are defined and the chosen representations are described. Using a large-scale (cadastre) example of entities, we conduct a case study of the model in Section 4. Section 5 addresses the representation of spatio-temporal relationships and temporal generalization through the model. Finally, Section 6 concludes with a summary about the capabilities of this approach and outlines any further research issues.

2. Existing Efforts concerning the Modeling of Space and Time

“Time is not something objective. It is not an entity, nor a random event, nor a relationship, but a subjective condition, created due to the nature of the human mind.” E. Cant

Spatial change is intrinsically related to time. Yet, the concept of Time is quite incomprehensible, thus turning any effort of modeling change into a difficult task. While it is easy for people to comprehend their surrounding space, time is a different matter altogether. It is simple to understand where each object lies and model this fact. However, we do not have full understanding of the time period this fact took place, but we perceive it by its effects. We use abstract definitions of time periods, like “past,” “present” and “future,” without exactly knowing when this period is. All we know about the course of entities through time is what happened to them in the “past,” what information we collect for them in “present” and we can estimate their behavior in the “future”. [Coveney and Highfield, 1990]. Most present spatio-temporal models consider time as linear, non-reversible. Langran defines Cartographic Time as a fourth dimension, following the Newtonian conception of a line (or rather a series of events) that does not interact with space [Langran, 1992]. To this, we will add the fact that this linear succession of events has a one direction, it “moves forward” much like an arrow.
Having this view of time in mind, we take a brief overview of a number of existing models. There are many proposed models for spatio-temporal entities documented in the international literature [Abraham and Roddick, 1996; Pavlopoulos and Theodoulidis, 1998; Langran, 1992]. In this section, only a number of these approaches are reviewed, as a basis for the ideas proposed in the next section.

Apart from the simple snapshot model and the space-time composite data model, one of the most basic approaches to the capture of change is **Time-stamping**. In this model every object has a tag with a pair of time stamps: one for the time of creation and one for the time of cessation. Though the model faces certain difficulties when tracing the history of a single object (as every version of this object is recorded separately), there is however an object-oriented version of the time stamping approach [Ramachandran et al., 1994]. In such an approach, data are collected for past, current and future states of each object. Still, the model actually administers change only in terms of its effects and not as explicit information (no information on “what happened” or “why it happened”).

Another quite interesting approach is the one expressed in the **History Graph** model. In terms of this model, three different types of temporal behavior are identified for real-world entities: 1) continuously changing objects, 2) basically static objects change by long events and 3) always static objects changed by sudden events. Cadastre map objects are typical examples of the last behavior. The whole model is based on the simple idea that an object may either be in a static state, a changing state or in a ceased state (see Figure 1). Static states of objects are called «object versions», while the changing states are called «transitions». Two time stamps, describing the interval of time in which the state of an object is valid, identify each object version. Each transition is an entity that relates object versions with its successors or predecessors, its period also described by two time stamps. Both versions and transitions can be characterized by a time interval. Object that change suddenly would then be described by transitions with zero duration (i.e. events), while continuously changing objects would be described by versions of zero duration.

![Figure 1. Generic behavior of temporal objects, according to the History Graph Model](image)

Yuan proposed another interesting view of spatio-temporal modeling by defining the **Three-Domain Model**. Her model represents semantics, space and time separately and provides links between them to describe geographical processes and phenomena [Yuan, 1994]. The semantic domain holds uniquely identifiable objects that correspond to human concepts independent of their spatial and temporal location. This is in contrast to other models where, for example, a land owner is represented as an attribute of a land parcel. In the three domain model, the land owner is a semantic entity that is linked to a land parcel (spatial object), with changes to the parcel associated with dates (temporal objects), and possible other land parcels involved in the transformation. Loss of ownership is implemented by linking another semantic entity to the land parcel together with the temporal object representing the date of change (sale).
Some researchers have tried to model spatio-temporal databases using the concept of *Moving Objects*. This approach allows an entirely general treatment of time-changing geometries, whether they change in discrete steps or continuously. This continuous change is represented with moving objects. If an object’s position in space is relevant, then the «moving point» is the basic abstraction; if also, the extent is of interest, then the «moving region» abstraction captures moving as well as growing or shrinking regions. Moving points, lines and regions are viewed as three-dimensional (2D space plus time) or higher-dimensional entities, whose structure and behavior is captured by modeling them as abstract data types.

Finally, based on the object-oriented paradigm, there exist *object-oriented models*, which include objects, classes, encapsulation, inheritance and polymorphism. Worboys was the first one to emerge object-orientation in spatio-temporal modeling. He introduces the concept of spatio-temporal object, as a duality, a unified object with both spatial and bitemporal extents [Worboys, 1994]. Claramunt and Thériault define an object with three attribute sets: the temporal domain, the spatial domain and the thematic domain [Claramunt and Thériault, 1996]. They use this model to formalize a number of spatio-temporal «processes». These processes are actually the kinds of change such an object can undergo. This formalization of the semantics of changes is quite thorough and extremely useful.

### 3. Model Ontology and Basic Principles

This work aim at combining some of the modeling ideas presented in the previous section into an approach to suit spatio-temporal modeling of cadastre features. The basic principles of such a model will be presented hereafter. We define an entity as the abstraction of a real-world feature, while «objects» are the database representations of entities. Langran [Langran, 1992] describes features (entities in our case) as something that exists for ever, since the beginning of time, allowing it to be born, changed, ceased and reincarnated. The course of an entity is described through its object versions in the system. The objects that consist an entity may change, cease to exist or new ones may be born. These are actually the various states of the same entity. The object versions may change suddenly (time instances) or between the versions may interject a time period (Figure 2) [Claramunt et al., 1998]. One of the first questions that arise is whether a changed object is actually a new object version of the same entity or rather a new object belonging to a different entity. Since the proposed model is presented at conceptual level, such a matter could be solved by the strict definitions of each entity in the model’s data dictionary. As the model is application-specific (large scale-cadastre entities), it can contain *formalized rules* for each entity, which indicate the kind of changes or what degree of change actually turns an entity into a ceased one and a new one is born (or re-born).

**Figure 2.** Successive Object versions of the same Entity
One of the foremost principles would be that the model uses the object-oriented approach to define the data structures. Object-oriented models provide a natural method of describing entities [Worboys, 1994], avoid data fragmentation and enable useful capabilities for managing time. In the model a number of entities are defined. Each entity acts like a class that contains a set of Objects. All these objects share attribute sets common to their entity. These attribute sets are distinguished into three separate domains [Claramunt and Thériault, 1996]:

- The Thematic domain, where belong the thematic attributes of the object.
- The Spatial domain, where the geometric representation and location of each object are described.
- The Temporal domain, which represents the entity’s temporal data.

The temporal domain is the most important for the modeling of change. Into this domain are concentrated all the aspects of the temporal object. First of all, using the history-graph model idea, here is recorded whether this object is under change, it is static or has ceased to exist. This information can be divided to “past,” “present” and “future” states. By Time-Stamping, the history of each version of the object is recorded and can be traced. The time recorded at the time stamps is “Valid” time, that is, the actual time the changes take place. Transaction time (also called database time) is also catalogued, in order to trace when the information was actually updated. If the transition between two successive object states is instantaneous, then the event took place in a time instant. If the object stays at changing state for a time period, then it results that it is under continuous change.

Apart for recording the present state of each object (static, changing and ceased), in the temporal domain is recorded the kind of change that actually became the event during which the object was altered. The changes that can transpire will all be predicted by the model (see Figure 3). Each entity will be associated with certain changes that it can undergo. These rules, part of each entity class definition, can act as integrity constraints when temporal information is concerned. The prescribed changes, recorded into the temporal domain of each object, will be linked with the respective changes that affect other objects. For instance, in the cadastre model, an event in the life of the entity “Parcel A” (split) is linked to the changes that objects “Parcel B” and “Parcel C” sustain (birth). This way, not only the history of each object may be traced, but also the events that lead to any transition in its life could be confirmed.

Finally, the spatial domain is related to the temporal domain of each entity. The spatial characteristics of each object change over the course of time, so the abstract types of «moving objects» are used to represent them. As an object shifts through various versions over its life, its geometry and location may be altered as well. By utilizing the moving forms, the history of its representations and the relations between its representation and the representations of other objects become evident. For instance, lets assume that one of the spatial representations of an entity is a point. Then an object representing this entity would have the form of a moving point. A moving point would be abstracted as a line over a time period. Some points of this line would specify the position of the object when it was static, the continuous parts of the line would describe the object when it was constantly changing, where any breaks in the line would respond to sudden movement of the object to a new location.
Figure 3. Some prescribed changes of geometric form [Claramunt and Thériault, 1996]

4. A Case Study of the Model

The aforementioned ideas were applied into a real-world cadastre paradigm, in order to study the behavior of the various objects that emerged. The case is quite abstract, with only a few cadastre entities, but demonstrates many of the temporal requirements such an information system may have. The land area is imaginary, consisting of a crossroads and a number of parcels and buildings. Figure 4 shows the spatial and ownership variation in the course of some decades in this century.

The chronology of events is as follows (granularity 1 year):

- **Year 1966**: Information is recorded for two streets, a land parcel owned by Mr. Jones (Parcel 1), a land parcel with unknown owner (Parcel 2) and a building into Parcel 1, also owned by Mr. Jones (Building 1.1)
- **Year 1973**: Mr. Jones sells a part of his parcel, so a new parcel is created, owned by Mr. Martin (Parcel 3), while Parcel 1 changes in size.
- **Year 1981**: Mr. Jones mortgages his parcel and building, Mr. Martin builds a building into Parcel 3 (Building 3.1) and parcel 2 is bought by Mr. Smith.
- **Year 1997**: Mr. Jones sells his parcel and building to Mr. Martin. Parcel 3 joins with parcel 1 and extends over it. The building into the previous parcel 1 is the same structure but as it changes owner and parcel, is now considered as a new building (Building 3.2)
In terms of modeling this paradigm, three (3) main entity classes are defined: the Land Parcel entity, the Building entity and the Owner Entity. In a fully developed data dictionary more large-scale entities would be included, as roads and parcel boundaries. Yet, the selected three are considered enough for this study. Also, one must note the lack of spatial characteristics for the entity “Owner”. Still, an owner is considered a real-world feature and thematic and temporal information is also recorded for this entity. Finally, it is a matter of question whether there should be another entity called Ownership, or it should be used as a relationship. Though the model would be more coherent by regarding ownership and its types as single entities, for the needs of this study, we regard it as a relation between Owners and Parcels/Buildings. Each land parcel, building and owner constitute instances of their respective entities. Below we describe some of those objects that represent them, taking into account the predicates of the model. Special attention is paid to the temporal domain of each object’s characteristics. Present valid time is considered to be 1999.

**Owner Entity**

Instance: *Mr. Jones*

*Spatial domain attributes:* None

*Thematic domain attributes:* \{Id\}, \{LastName\} = “Jones”, \{FirstName\}, \{Address\} …

*Temporal Domain:*

- \{Past\},1 = \{begin=1966, end=1973, ownership = owns, object ={P1, B1.1}, static\}
- \{Past\},2 = \{begin=1973, end=1973, ownership = owns, object ={P1, B1.1}, change=ownership_change{sell to Mr. Martin}, change=motion_change{P1 size change}, change=life_change{P3 birth}\}
- \{Past\},3 = \{begin=1973, end=1981, ownership = owns, object ={P1, B1.1}, static\}
- \{Past\},4 = \{begin=1981, end=1981, ownership = mortgage, object ={P1, B1.1}, change = ownership_change{mortgaged to Mr. Martin}\}
- \{Past\},5 = \{begin=1981, end=1997, ownership = mortgage, object ={P1, B1.1}, static\}
- \{Past\},6 = \{begin=1997, end=1997, ownership = none, change=ownership_change{sell to Mr. Martin, change=life_change{P1, B1.1 death}, change=motion_change{P3 form change}, change=life_change{B3.2 birth}}\}

*Present* = \{begin=1997, ownership = none, static\}
Instance: Mr. Martin

Spatial domain attributes: None

Thematic domain attributes: {Id}, {LastName}="Martin", {FirstName}, {Address}…

Temporal domain:
{Past},1=[begin=1973, end=1973, ownership = owns, object ={P3}, change=ownership_change{buys from Mr. Jones}, change=motion_change[{P1 size change}, change=life_change{P3 birth}]]
{Past},2=[begin=1973, end=1981, ownership = owns, object ={P3}, static]
{Past},3=[begin=1981, end=1981, ownership = mortgage, object ={P1, B1.1}, ownership = owns, object ={P3, B3.1}, change=ownership_change{mortgaged from Mr. Jones}, change=ownership_change{builds B3.1}, change=life_change{B3.1 birth}]
{Past},4=[begin=1981, end=1997, ownership = mortgage, object ={P1, B1.1}, ownership = owns, object ={P3, B3.1}, static]
{Past},5=[begin=1997, end=1997, ownership = owns, object ={P3, B3.1, B3.2}, change=ownership_change {buys from Mr. Jones}, change=life_change{P3, B1.1 death}, change=motion_change{P3 form change}, change=life_change{B3.2 birth}]
{Present}=[begin=1997, ownership = owns, object ={P3, B3.1, B3.2}, static]

Land Parcel Entity

Instance: Parcel 1

Spatial domain attributes: {Location}, {Geometric Representation} = Moving Region

Thematic domain attributes: {Id}, {LandUse}, {Area}, {OwnedBy}, {IncludedBuildings}…

Temporal domain:
{Past},1=[begin=1966, end=1973, ownedby = Jones, includedbuildings ={B1.1}, static]
{Past},2=[begin=1973, end=1973, ownedby = Jones, includedbuildings ={B1.1}, change=motion_change{P1 size change, add new form to moving region}, change=life_change{P3 birth}]
{Past},3=[begin=1973, end=1981, ownedby = Jones, includedbuildings ={B1.1}, static]
{Past},4=[begin=1981, end=1997, ownedby = {Jones, mortgaged}, change= ownership_change {mortgaged to Mr. Martin}]
{Past},5=[begin=1997, end=1997, ownedby = {Jones, mortgaged}, includedbuildings ={B1.1}, static]
{Past},6=[begin=1997, end=1997, ownedby = nobody, change=ownership_change{sell to Mr. Martin}, change=life_change{P1, B1.1 death}, change=motion_change{P3 form change}, change=life_change{B3.2 birth}]
{Present}=[begin=1997, ceased]

Building Entity

Instance: Building 3.1

Spatial domain attributes: {Location}, {Geometric Representation} = Moving Region

Thematic domain attributes: {Id}, {Use}, {Area}, {OwnedBy}, {NumberOfFloors}…

Temporal domain:
{Past},1=[begin=1981, end=1981, ownedby = Martin, change=ownership_change{built from Mr. Martin}, change=life_change{B3.1 birth, add form to moving region}]
{Present}=[begin=1981, ownedby = Martin, static]
5. Ideas on Spatio-Temporal Relations and Generalization

Large-scale map objects modeled accordingly to the approach presented in this work may be suitable to express concepts found commonly in spatial schemas. Two of these concepts are generalization of entities over the course of time and the spatio-temporal relations of entities. The fact that the model utilizes the object-oriented approach facilitates generalization at a conceptual level, as the model allows hierarchies of entity classes and multiple representations [see Panopoulos and Kavouras, 1997; Kokla and Kavouras, 1999]. By expanding this logic, one can include formalized rules over the representation of each entity class over the course of time. Temporal generalization regards the representation of each object (or its changes) over different time scales. In a large-scale cadastre application, these time scales could be defined into the model itself. They would actually be in different granularities (e.g. 1 month, 1 year and 1 decade). The changes of each object could be represented accordingly to its class definition into the conceptual model. For instance, let us assume two different time scales (1 year and 1 decade) for the example of parcel 1 in the presented case study. According to rules recorded for the Parcel entity, one could count the time periods during which Parcel 1 was in static state and in changing state; and then represent these changes for the two scales. For one-year granularity, all the altered states of the parcel would be shown: a basically static object that ceased to exist sometime in the last decade. As for the visual representation of the objects through time, their moving abstract types would be generalized themselves. For the same parcel of our example, its generalized form over the course of time would be its second form (after the year 1973). Actually, this would be the generalized form of its moving region for time intervals of one decade.

In the context of relationships, both spatial (topological) and temporal relations of objects have been defined (e.g. some temporal relations include “one event before another”, “both events happening simultaneously”, “both events ending together” etc.). The combination of these two discrete formalizations into spatio-temporal relations is a work already under way [Claramunt et al., 1998]. The proposed model would prescribe all the kinds of possible spatio-temporal relations among cadastre entities. Again, its object-oriented capabilities and the recording of the successive states of each object would facilitate the “translation” of these relations between the various object versions. The changes of each object’s state are linked, so that they can be traced (usually an event results to the transition of more than one entity) and therefore related objects can be found. It is quite important for the model to include, in a formalized state, all the relations of its moving objects. If these relationships are defined, then the spatio-temporal relations among objects will be represented.

6. Conclusions

In the context of modeling the temporal aspects of cartographic entities, this paper presents an approach regarding large-scale applications, emphasizing on cadastre entities. The proposed model regards cartographic time as a linear directional arrow, not interacting with space. It suggests the definition, at a conceptual level, of real-world entities, which exist forever and can be born, altered, ceased and reincarnated. Objects are the representations of entities, and a set of objects can, at any given time instant, consist an entity. Each object has three sets of attributes: spatial, thematic and temporal. The temporal domain is linked with the two others. In the temporal domain are recorded the past and present states of each object. These states can be only three: “static”, “under change” and “ceased”. The periods of each object’s state are recorded through time stamps. Also, all the changes that affect the object are recorded in its temporal domain. If a change is related to a change altering another object, then these changes are linked, in order to fully describe the event. The conceptual definition of each entity prescribes which changes it may undergo, acting as an integrity constraint. The spatial representations of these “temporal objects” are expressed through the utilization of abstract types known as
moving objects. These types represent all the altered forms of each object through its versions. The model is further explained by a cadastre paradigm, used as a case study.

The proposed ideas do by no means form an exhaustive conceptual model. Yet, they define the nature of such a model, able to cope with spatio-temporal change in cadastre maps. It is within the logic of this model to include strict formalizations of all the relevant entities, their domains, their attributes and their relations (as the model specifically refers to cadastre applications, such formalization can be accomplished). If such a model was implemented, in a way similar with the case study, then it would gather all the necessary advantages to act as formal specification for cadastre database modeling. A lot of design considerations remain to be taken into account, in the context of entity definition and relationship formalization, as well as the introduction of concepts like temporal generalization. The ultimate goal of such research would be the development of a scaleless geo/spatio-temporal database containing information with the necessary intelligence to support a variety of complex socio-economic and environmental applications.

References


Coveney, P. and Highfield, R. (1990), The Arrow of Time, W. H. Allen

Kokla, M., and Kavouras, M. (1999), Spatial Concept Lattices: An Integration Method in Model Generalization, to be presented in Third Workshop on Progress in Automated Map Generalization, 12-14 August 1999, Ottawa, Canada

Langran, G. (1992), Time in Geographic Information Systems, Taylor & Francis


Ramachandran, B., MacLoad, F., and Dowers, S. (1994), Modeling Temporal Changes in a GIS using an Object Oriented Approach, in Advances in GIS Research, Sixth International Symposium on Spatial Data Handling (SDH’94)

Worboys, M. F. (1994), Unifying the Spatial and Temporal Components of Geographical Information, in Advances in GIS Research, Sixth International Symposium on Spatial Data Handling (SDH’94)

Archiving the Cartographic Record: The Canadian Experience

Betty Kidd
National Archives of Canada
bkidd@archives.ca

In the National Archives of Canada, maps of national significance from both the private and public sectors and from all time periods are a major component of the nation’s archival heritage. The “total archives” concept—named in the 1970s but in practice for a century before that time—explains the Canadian experience, not only at the federal level but also at the provincial and territorial levels. This situation varies from that in many of the other large westernized nations where the cartographic record is often divided between libraries and archives—the libraries acquiring private sector mapping and the archives, records of the state. The practical Canadian archives model, however, was copied by many newer nations in the twentieth century.

From the time that the Dominion Archives (later Public Archives and now National Archives) was founded in 1872, and especially since 1907 when a separate administrative unit—initially the “Map Room”—was established for cartographic records, specialized archival staff have ensured the acquisition of, organization and description of, preservation of and access to the current cartographic products in all forms—including geomatic records—as well as of the older maps, plans, charts, globes, atlases, etc. which document the history of Canada. An important component has always been the active participation of cartographic archivists and map librarians in the national and international cartographic communities where a shared subject expertise ensured communication and specialized research use of archival holdings and which contributed to collection growth and excellency.

In the last decade and half, however, economic constraints have caused many changes in the Canadian archival community, focusing the use of resources on the records of the corporate sponsor (usually traditional textual records) and resulting in an imbalance in public/private sector activities. In the same period, movements towards functional organization structures and standardization of processes across all archival media have ignored the physical and intellectual needs of specific media holdings (present and future), sacrificing specialized media (including cartographic) expertise, reducing the involvement of media archivists in subject communities and endangering the record of the future.

The paper will examine the past and current Canadian experience and will project necessary action in the future to ensure that the archival cartographic/geomatic record in Canada is not diminished in quality nor excellence.

During a 1936 House of Commons debate, the Prime Minister of Canada, W. L. Mackenzie King made the disparaging remark that “if some countries have too much history, we have too much geography.” Certainly as a nation stretched across the northern part of North America, Canada is a vast nation—5,514 km. from east to west, 4,634 km. from north to south, and 9,970,610 sq. km. in area. Thus, it is not surprising that, as a nation, Canada has been dependent on surveyors and cartographers to gain knowledge of these enormous distances and immense spaces, the knowledge necessary to enable settlement, development and national security.
Cartographic Archives and Map Libraries:

The products of the explorers, the surveyors and the map-makers in the past, the present and the future are what is referred to in this paper as cartographic records. In the 1989 National Archives and Records Administration of the United States (NARA) guide to cartographic records, these are defined as “Graphic representations drawn to scale of selected physical and cultural features of the surface of the earth and other planets. Included are maps, charts (hydrographic/nautical, weather, and aeronautical), photomaps, orthophotomaps, atlases, cartograms, globes, relief models, and related records, such as field survey notes, map history case files, and finding aids. Also included are digital cartographic records, such as geographic information system records...” [NARA, B-2]. Earlier it had been noted that “Cartographic archives may be in the form of either unique manuscript and annotated printed records or multiple copies involving such reproduction methods as engraving, etching, lithography and photography. It is a mistake to assume that archives contain only unique manuscript documents” [Ehrenberg, 1975, 3]. In the National Archives of Canada (NA) guide to these records, it is noted: “Records pertaining to maps and geomatic systems fall into three categories: the data-gathering stage (surveys), the data processing stage, and the production and copying stage.”[Cardinal, 1999, 4].

Since the differentiation between the mandates of the two types of institutions which acquire cartographic materials—that is, map libraries and cartographic archives—is not always understood, a brief explanation is necessary. Map libraries—in Canada, most often located in universities—normally emphasize the reference use of the maps acquired and in their holdings. Cartographic archives—most governmental—are more concerned with the preservation of the records since their mandates are not only to serve current users but also, potential users many generations into the future. Both types of institutions may hold copies of the same maps, but it is the archival copies which are expected to be available in the future. At the April 1978 seminar on cartographic archives hosted by the Public Archives of Canada, Yves Tessier, then map librarian at Laval University, stated that “co-operation must be assured between the two types of map repositories in the areas in which we are interdependent and complement each other. To map libraries, map archives are a must! We need a place we can rely on when exhaustive research is needed...We need a place we can rely on when we want to find documents as they were produced, be it a few years or centuries ago. Co-operation excludes competition ....we are bound to a joint venture—for the sake of users and posterity.”[Tessier, 12-13]. In her 1993 paper “Cartographic Archives and Map Libraries: Two Sides of the Same Coin?” presented at a joint session of the ACMLA and the Association of Canadian Archivists, Elizabeth Hamilton noted that: “the Association [ACML] was formed to provide a forum for mutual concerns relating to the stepchildren of both libraries and archives. Rather than do turf battles, there was recognition at the outset that we could do far better by recognizing the commonality of our interests and the important contributions which could be made through our differences in environment, philosophy and approaches to the material.” [Hamilton, 2].

“Total Archives” - A Canadian Concept:

To understand the development of archives in Canada and thus, to understand the Canadian experience with cartographic archives, one must understand the “total archives” concept. Simply, “total archives” means one institution acquiring records in all media from both the public and private sectors—a system characterized by economies of scale and ease of research and access. The concept is based on the history of the Dominion Archives which, in effect, quickly evolved such a broad collecting mandate following its creation in 1872—the first national cultural institution in a fledgling nation, only five years of age. In many nations where other cultural institutions, such as state libraries, were founded prior to or in the same period as state archives, the archives were often limited to the acquisition of the records of the state while libraries, museums, etc. acquired private papers and other media records. However, in Canada, a National Library was not founded until 1952.
Although this approach was pioneered at the NA, it was followed by virtually every provincial, municipal, university, and church archives in the country, and by many non-European countries around the world. The term itself was not introduced until the 1970s—by Wilfred I. Smith and Hugh Taylor. This was the period in which “total archives” reached its highest point and the support and “generous amount of autonomy [provided] to ... media divisions” [Kidd, 1981, 474] witnessed a surge in activity and growth in cartographic archives in Canada. Dr. Smith’s vision of “total archives” and his support of media archives is evident in a 1973 statement that “There has been a tendency to neglect other and equally valid records. Examples of such records are maps, plans, and architectural drawings.... It would give me a great deal of personal satisfaction if, one day, the custodians of ... specialized materials who are called curators or other such names could be called simply archivists.” [Smith, 1974, 6]. By the late 1970s, the designation in the NA had broadened to other than those working with textual records.

**Cartographic Archives in the National Archives of Canada:**

**The First Century:**

After the founding of the Dominion Archives in 1872, some of the first archival records to be accessioned were maps and plans left behind by the British authorities leaving Canada. By 1873, Douglas Brymner, the head of the new archives, was referring to the necessity of making tracings of maps in England and France; in 1881, the first purchase of maps occurred; and by 1883, there was a collection of 426 maps and plans. In 1907, Hensley Reed Holmden was appointed to head the newly created “Map Room” (numbering 4,285 items), his charge until his retirement in 1924. He quickly organized the scattered cartographic holdings, prepared the first published catalogue of the holdings—*Catalogue of the Maps, Plans and Charts in the Map Room of the Dominion Archives*, 1912—and made cartographic records accessible for research use, even overseeing the tracing of maps for researchers. Holmden’s successor, Norman Fee headed the Map Division for the next 22 years (until 1946), a period characterized by steady growth in holdings (except during the war years), classification and cataloguing of the holdings according to an in-house system, the introduction of a multifarious storage system which grouped maps by size and researchers’ access to photostatic copying.

The years following the Second World War witnessed a cartographic explosion in Canada, and a corresponding increase in the cartographic holdings of the Public Archives. The two chiefs in this period were A.J.H. (Jack) Richardson (1946-1954) and T.E. (Ted) Layng (1955-1973). The phenomenal growth in holdings—from 20,400 in 1945 to some 500,000 in the early 1970s was the result of the expansion of the mandate to include Canadian current maps, foreign maps and architectural records. From 1949, collecting current maps as they were produced was accepted as the most efficient and economical method of acquisition but not until the mid-1960s when the division assumed responsibility for the Canadian contribution to the *Bibliographie cartographique internationale* did this become a systematic acquisition program. In 1965, approximately 150,000 foreign maps (subsequently processed and reduced to 115,000 maps) were transferred from the Department of Mines and Technical Surveys. The division subsequently received foreign maps on a regular basis as part of a tripartite arrangement (the other partners being the Department of National Defence and the Department of Energy, Mines and Resources) which established exchange agreements with other nations. In 1970, the department and the Royal Architectural Institute of Canada agreed to establish a National Architectural Archives program. Actually, there had been an architectural component since the department’s founding, a common occurrence in archives where cartographic and architectural records are administratively combined. In this same period, a new area classification system (1950) and a new cataloguing system (1969) were introduced, several key catalogues were compiled and published—namely, Ted Layng’s *Sixteenth Century Maps Relating to Canada* (1957-8), still an unsurpassed standard reference for this time period and *County Atlases of Canada: A Descriptive Catalogue*.
and outreach to the communities culminated in the founding of the Association of Canadian Map Libraries in 1967. A 1973 report prepared for the Council of Library Resources in the United States noted of the National Map Collection (the divisional name in use since 1968) that, “in terms of item control, physical care, and reference service, it is a model for others to emulate.” [Ehrenberg, 1973, 5].

**The Growth Years (to mid-1980s):**

Former Assistant National Archivist, Michael Swift, noted in 1995 that “When I first worked for the Archives in the mid-1960s, the institution had a staff of about 100 people and the annual budget was about $1 million. Twenty-five years later, the National Archives had a staff of 800 people and an annual budget of $60 million.” [Swift, 3]. Certainly this growth was experienced not only at the NA, but also in other archives and in the Canadian economy in general.

In this period, five distinctive “collecting” areas were identified: 1) history of cartography; 2) current cartography of Canada; 3) current cartography of other parts of the world; 4) the federal record—cartographic and architectural; and 6) the private architectural record. A sixth area—the acquisition of cartographic books, periodicals, etc. by the Public Archives Library—was also noted as essential to the division’s operations. A major divisional reorganization in 1976 allowed, for the first time, a systematic approach for the acquisition of government cartographic and architectural records and a concentrated effort to acquire originals of significant early maps, globes and other cartographic forms from the private sector through donations and purchases. The 1970s saw the holdings double to approximately 1,000,000 items. Mapping was changing as automation of mapping processes, CAD systems and GISs were being increasingly used. From the mid-1970s, the division was aware of the need for a pro-active approach to ensure that these records would be preserved. Initially, discussions took place with the Machine Readable Archives Division (MRA) asking that an active acquisition program for these records be instituted; in 1983, David Brown of MRA (who would later transfer to the cartographic area) prepared a report entitled “Various Archival Implications Associated with Geographic Base Files”. In the same year, a paper at the Auto-Carto Six conference held in Hull noted that “The greatest challenge to be faced by the cartographic archivist in the last decades of the 20th century will be the acquisition, control and servicing of computer-based cartographic systems” (Cardinal and Kidd, 2) and in which the necessity to include the archival requirements at the planning stage of any project was stressed. When MRA was abolished, the cartographic division began to lobby for the necessary resources.

The division’s card catalogue, built from the 1950s was published in 16 volumes by G.K. Hall & Company in 1976, after an intensive clean-up period by divisional staff. The cataloguing standards staff worked closely with the ACML with the objective of establishing a national union catalogue and national cataloguing standards. International developments quickly overtook these initiatives and divisional staff and ACML members were instrumental in the preparation and publication of the International Standard of Bibliographic Standard (Cartographic Materials), the second edition of the Anglo-American Cataloguing Rules and in 1982, Cartographic Materials: A Manual of Interpretation for AACR 2. After feasibility studies of the DOBIS, UTLAS and MINISIS systems, the cataloguing in the division was automated, utilizing the UTLAS (now AG Canada). This system was used in addition to the on-going preparation of finding aids, especially for government records and private architectural fonds.

In 1977, the 105mm microfilming program was introduced in the division—originally envisioned as a conservation measure, for the last 22 years, its usefulness for research and reference has been vital. In 1974-75, for the first time in the department’s history, a map conservation specialist was hired—finally, maps were treated as more than “wallpaper” when requiring conservation treatment. In the division, many measures were introduced to ensure the proper handling of the maps and plans by staff and by researchers, storage methods were reviewed, vertical storage gradually eliminated and acid-free containers introduced.

In 1978, the NMC hosted the first seminar on cartographic archives for those responsible for maps at provincial and territorial archives. The opportunity to discuss common problems and to plan future programs was deemed highly successful. In response to a questionnaire sent to the provincial and territorial archivists after the second seminar in 1980 and in anticipation of a third, respondents noted that the seminars had contributed to the development of their staff members as archivists, that the opportunity to discuss their work with colleagues from other provincial archives resulted in archivists who “came to see the problems we have in common to most archives”, that the map holdings became better organized, handled and described, and that the contextual and informational content of the holdings were better appreciated. The third seminar was held in 1983 and a fourth in 1988. An informal “Cartographic Archivists’ Newsletter” continued communication throughout this period.

In 1980, the National Librarian, Guy Sylvestre, published his report entitled *The Future of the National Library of Canada* in which one recommendation was the transfer of the National Map Collection to the National Library. The Dominion Archivist responded: “Actually, I thought this matter had been settled. Early in the 1970s, a major part of the Canadian cartographic community—producers, custodians and users—strongly opposed just such a proposal. The Hon. Hugh Faulkner, Secretary of State at that time, assured us and the cartographic community in 1974 that the National Map Collection would remain part of the Public Archives. ...We oppose the proposal because of the fact that a large part of the Collection’s holdings are public records which cannot be alienated from the Public Archives. What the report in effect would bring about is a splitting of the Collection which would result in two competing collections as well as the duplication of many functions. Canada would lose what is now envied by other countries—one, comprehensive, national collection of documents of a cartographic nature.” [Smith, 1980, 2].

The respect for differences which characterized “total archives” was facing challenges by the end of this period of growth. Although the mainline archival principles were established based on experience with textual records, the specialized media archival areas developed their principles and procedures based on the needs of the records and often in cooperation with professionals from other cultural institutions (e.g., libraries and museums) and with the producers of the records themselves (e.g., film archivists with film makers and cartographic archivists with cartographers). In fact, specialized media archivists were often more active in specialized societies and associations for the curators of such records or in subject associations than in the mainline archival associations. The differences in philosophy between textual and specialized media archives and the resultant misunderstandings were demonstrated in an article by Terry Cook entitled ‘The Tyranny of the Medium: A Comment on “Total Archives”’ in *Archivaria*. Cook noted that “Indeed, it might be declared that there are two principles [total archives and provenance] warring in the bosom of a single profession.” (Cook, 141). He continued “Quite simply, the internal divisions of archival institutions along media lines has created a de facto fragmentation of the archival whole, as defined by the principle of provenance... This is evident in the daily operations of not only the Public Archives of Canada (PAC), but also most other national archives and of many provincial and municipal archives of this country” (Cook, 142-3). What Cook ignored were the philosophies and tradi-
tions that had evolved for each of the specialized media, including cartographic records, and even the opinions of classic archival “fathers” such as Schellenberg, as pointed out in Andrew Birrell’s article “The tyranny of tradition” in the next issue of Archivaria. As far as Schellenberg was concerned, it did not matter if archival materials were physically separated as long as the principle of provenance was observed in the medium and even here, he was casual. Of maps, he stated “large accumulations of maps can be handled more easily if kept by provenance than if classified by area,” but he also said “the principle of original order may be applied with considerable latitude to cartographic records” (Schellenberg, 312-3).

The Difficult Years (mid-1980s and 1990s):

Although the years since the mid-1970s had seen economic problems in Canada and little growth in archives, it has been approximately the last decade and a half which have been the more difficult years for government bureaucracies in Canada and for the archival institutions which form part of these bureaucracies. Economic constraints and government downsizing necessitated review of on-going archival programs and resulted in reduction, consolidation or elimination of certain programs and a smaller work force. At the same time, the archival profession was developing and changing philosophies and ideas were being introduced, some of which were radically different than in previous years. Although the impact of many of these professional changes are positive, added to the other changes—economic and organizational—which have characterized these years, the positive aspects have often been overshadowed by the weariness caused by unceasing change.

The nature of archives means that despite the constant change, the day to day work continued and there have been numerous accomplishments worthy of note in the cartographic archives area. In addition to the annual acquisitions of note, these include:

a) the establishment of an archival geomatics program following the allocation of one person year commencing 1988-89 “for the appraisal and acquisition of machine-readable cartographic data” (Senior Management Committee decision) and the reassignment of a second cartographic staff member to the program. The successes with the records of the Canada Land Data System, including Canada Land Inventory and the Canada Land Use Monitoring Program records and with those of the Canadian Ice Centre as well as with many other smaller accessions have demonstrated the effectiveness of this program. Additional resources are required to ensure continuing success.

b) the automation of control work, including the use of the UTLAS system, the purchase of an IBM System 36 mini-computer in 1986 (for a number of years, used for accessioning, the microfilm register, finding aids, etc.), and the recent implementation of MIKAN, an integrated model for accessioning, arrangement and description, as the departmental control system. Since the MIKAN system does not as yet support the item level description required for cartographic and other archival records, other systems continue to be utilized. As well, a system has been developed for the transfer of cartographic finding aids to a standard automated format.

c) the development of descriptive standards (revision of AACR2, Rules for Archival Description - RAD). For RAD, coordinated by the Bureau of Canadian Archivists, NA staff members have been active participants in the development of the chapters for specific media records (for the cartographic chapter, Velma Parker was a key participant).

d) better environmental conditions for storage—but the distance factor and the access difficulties have yet to be resolved.

e) the closer working relationship of archival staff from all media in acquisition, disposition and control work.

f) the progress in identification and scheduling of government cartographic records. From 1985-1990, divisional surveys of the National Capital Region (pilot project), the Pacific Region, the Toronto area and the Atlantic Region were carried out. The introduction of the departmental Multi-Year Disposition Plan, a “macroappraisal” approach, in which VSA became a participant included such benefits as improved con-
textual knowledge. However, many of the cartographic records received in recent years have actually continued to be “direct transfers” and the volume of cartographic records received has significantly declined.

g) the progress on Carto-Canadiana. The division participated in the 1988-89 ACMLA feasibility study, then published several microfiche editions and was extremely pleased with the successful negotiations with the National Library to include these records in the new NL CD-ROM product. Similarly, progress has been made with the National Union Catalogue in that cartographic records are now included in the NUC housed in the National Library.

h) the success of the 105mm microfilming program (also known as the Large Microfilming Format (LMF) program. In 1986, a state of the art Opti-Copy camera was purchased; this provided a 30% improvement in resolution and unprecedented excellency in colour. A concern is the lack of capital funding for eventual replacement or upgrade.

i) the successful negotiations with the British Library for the loan of the Goad fire insurance plans for copying purposes.

j) continuing, although reduced, activity in professional communities (including involvement of NA cartographic staff in the organization of ICA’99). Restrictions on foreign travel and reduced operational budgets combined to reduce active staff involvement in many professional associations.

k) the passage of the National Archives of Canada Act in 1987, in which maps and other specialized media records were for the first time named in Canadian legislation as archival records—the previous act in 1912 had not specified the types of records. However, legal deposit for maps was not included, much to the disappointment of the community—and of the division—which had lobbied for its inclusion in the new act.

The latter half of the 1980s and the 1990s has witnessed massive cuts in the federal public service. In 1986, the arrival of a new Dominion Archivist, Dr. Jean-Pierre Wallot, unfortunately coincided with the introduction of such cuts. One of his first tasks, as part of the government decision to reduce the public service by 15,000 positions, was to reduce the staff of the Public Archives by 8.8% (72 person-years). The National Map Collection was directed to cut 3 person-years by “phasing out the Foreign Map Unit and other cartographic and architectural activity” (although that unit no longer existed). The major cuts were still to come. The first phase of Program Review from 1995-96 to 1997-98 witnessed a 16% reduction in the National Archives’ budget ($8.8 million) and Phase II in 1998-99, a further 3.5% ($1.5 million). The impact on most areas of the NA has been significant. For the cartographic area, the losses were both in operational dollars for acquisition, control, administration, etc and in person years. The person year reductions were staff members who volunteered for early retirement or departures—no one was forced to leave—but unfortunately the exodus of these key personnel has left major gaps in both ongoing operations and knowledge of the holdings. For example, the three persons who left in the last year had a cumulative experience in archives of more than a century. With no growth allowed at this time, it is doubtful that this expertise will ever be duplicated.

The reduction in person-years and the elimination of a number of work units resulted in organizational studies and changes. The new organizational structures introduced were characterized by functionalism, consolidation of seemingly like activities, “standardization” of processes and emphasis on corporate records. For some, such ideas represented increased efficiencies but for specialized media archives where differences had been celebrated and these differences respected, if not always understood, by senior managers, these changed directions introduced a difficult period. Certainly, the status and identity of the cartographic program—and that of other specialized media—was significantly devalued. Divisions were reduced to section status or less in the following reorganizations. The December 1986 departmental reorganization left the division relatively intact, but nevertheless did effect operations—responsibility for the 105mm microfilm program was to be shared with the Conservation Branch, one person-year was transferred to a new central Researcher and Reference Services Division, and reductions in library services impacted on the cartographic library. An unwelcome result of the 1987 act was the interpretation that “national” should no longer be used in divisional and program names—thus the name “National Map Collection”, a name requested by the community in the 1960s and in use since
In 1991, the separate administrative status of the cartographic area—gained in 1907—was lost when the Cartographic and Architectural Archives Division (the name which replaced the NMC) was administratively combined with the Moving Image and Sound Archives Division and the new division was named the Cartographic and Audio-Visual Archives Division (CAVA). Two years later, a major departmental reorganization combined components of CAVA with components of the Documentary Art and Photography Division (itself the result of previous consolidations) and the Visual and Sound Archives Division (VSA) was established. The 1993 functional reorganization stripped public service, custodial and shared microfilming responsibilities from the new division, handed these to new or enlarged divisions elsewhere in the organization and left VSA with a restricted mandate for acquisition, intellectual control and specialized public service. As well, continuing involvement in their communities, especially at the international level, became difficult. Within the department, the sword of a public/private organization—which would eliminate organization by media—favoured by many in the 1986 and 1993 organizational changes and again in a branch reorganization in 1997 continues to hang over the specialized media.

Associated with the organizational changes and revised mandates were physical relocations of staff, transfers of card catalogues, finding aids and reference books and files (most of which were unique) and which had been jointly used by staff and researchers throughout the years to the new public service area, moves of archival records to Renfrew (1988) and to Gatineau (1998)—admittedly better storage conditions but the distance factor meant difficulties in access, and new procedures and a computer tracking system introduced by the custodial area. The result was that archival staff no longer had readily available the tools required for their daily acquisition and description work; accustomed to easy access to original records, finding aids and reference tools and being able to have these all available to perform their functions efficiently, archival staff became increasingly frustrated. Equally frustrating was the separation from the research community with which the cartographic staff had long enjoyed interaction.

The current state of cartographic archives is discouraging when compared to that in the earlier “Growth Years” but it must be acknowledged that the last decade and half have been difficult for archives in general. There are continuing successes, some of which are noted at the start of this section. Of greatest concern is that at the present time, most of the few remaining cartographic specialists are within a decade or so of retirement, that many have responsibilities other than those related to cartography and that some of the functional areas in the department are moving or are encouraged to move from media specialization to “generalists”. Following the massive cuts experienced since the late 1980s, there have not been opportunities to hire and develop the cartographic archivists who could be the future leaders in cartographic archives. The future is uncertain.

On March 12, 1998, the Minister of Canadian Heritage, Sheila Copps, commissioned Dr. John English, a professor of history at the University of Waterloo to undertake a study on the future of the National Archives of Canada and National Library of Canada through an extensive consultation process. Although his report was expected to be released late in 1998, no announcements have yet been made. The following quotes from the ACMLA brief are representative of the reactions of the various specialized media communities and those of the Canadian Historical Association represent the reaction of a major user community. The brief from the ACMLA notes “The report focuses on the National Archives of Canada, which has been responsible for cartographic materials for over 125 years and serves as the de facto National Map Collection. However, considering recent changes at the Archives, our members are extremely concerned that not only is this ‘traditional’ status being eroded, but also there is a clear de-emphasis of cartographic materials.” Recommendations include “a National Map and Geographic Information Collection (MAGIC) be formed to regain the expertise, quality services and professional example once evident at the national level,” “appropriate subject specialists be assigned for public service responsibilities...,” “adequate resources be allocated for the description of the cartographic holdings of the National Archives...,” “investigate a legal mechanism for ensuring that the cartographic output of our country is collected and preserved...”and “the Archives undertake an active digitizing project for their cartographic material and make the images readily available via the Internet.” That of the CHA
supports “total archives” in Canada, noting “Since the end of World War Two, the NA has been committed to acquiring and preserving archival records in every medium from the private and public sectors of Canadian life... This commitment should be re-affirmed by the NA and through its leadership, by the larger Canadian archival community.” The CHA strongly objected to the centralization of researcher services in the NA, insisting that “it is absolutely essential that the archivist interact with the researcher so that the specialized knowledge of the records can be shared and utilized ... as long as the knowledge of speciality archivists is not being utilized, the rich holdings of the NA remain under-exploited.” The historians also recognized some of the special problems faced by specialized media archivists noting that “...for media-specific archivists, who work with maps, architectural plans, photographs, and other media-based records...to do their job effectively, they need to be in close contact with not only researchers, but with the records themselves; otherwise, their knowledge of the material becomes mediocre and their skills begin to disintegrate.”

**Provincial and Territorial Archives:**

The provincial and territorial archives in Canada have, in large part, shared the philosophy and practices of the National Archives with respect to specialized media records, including cartographic records. Since these archives are smaller in terms of holdings and assigned resources, there have been numerous variations. The “difficult years” described above have impacted on these archives in a variety of ways. In some, the situation has changed only slightly; in others, there have been major shifts in terms of responsibilities for the private record or in the definition of what constitutes the public record—certainly, in all, there are fewer resources dedicated to specialized media archives—the cartographic archivist of the past may still exist but he or she has assumed responsibilities for other types of records, for portfolios, etc. One long-term cartographic archivist in a provincial archives recently noted that “Because I am a shared archivist, there hasn’t been the time or energy to devote to this area”—perhaps 10% of his time is spent on cartographic records. The presentation in August will include statistical and other information about cartographic archives in these institutions.

That there were underlying tensions was evident in the mid-1980s as cartographic archivists active in the ACML began to suggest a change in the association’s name. In 1987, an appeal was made at the conference for a name change to “provide archivists with a greater sense of professional affiliation and commitment to the Association” and even more importantly, “the relevance of an archivist’s professional affiliation with the Association would be enhanced in the view of resource allocators... competition on all levels is intense.” (Hutchison, 28). The membership agreed and after 20 years as the ACML, the association became the Association of Canadian Map Libraries and Archives (ACMLA).

Certainly, there are attitudes and actions in the archival community which are alarming for those concerned about the future of cartographic archives in Canada. Two examples are the “archival appraisal scheduling policy and guidelines” prepared by the Nova Scotia Archives and Records Management which notes a list of classes of material which “are not archival and shall not be scheduled for archival appraisal”, including “(f) Non-written or non-verbal documents (e.g., graphic materials, cartographic materials, architectural and technical drawings, moving images)” and an electronic message received on March 9, 1999 from a senior Canadian archivist noting that his archives “has abandoned media as a basis for organizational structure—we are now organized along core functional lines. From a professional staff expertise, media specialization is now quite secondary as we prefer solid generalists with pre-employment graduate level education in archival studies. Ideally, media specialization should reside at the technical and para-professional levels.”
Recommendations:

What is the future of cartographic archives in Canada? In the past, the acceptance of the “total archives” concept meant that cartographic archives were supported in most Canadian archives. In general, the archivist profession does not have an untarnished reputation in its handling of maps; Cornelius Koeman, a historian of cartography, wrote in 1968 that “By discrimination against bulky and awkward maps, archivists have done great damage to their value as historical documents.” [Koeman, 78]. Canadian cartographic archivists until recently could argue that, in Canada, this had not occurred. In the future, will we be able to do so? To conclude, a few recommendations:

To cartographers:

Cartographic archivists need your support—we consider ourselves as your partners in ensuring the preservation of the cartographic record. Thus we ask that you be concerned about the preservation of your and your agency’s cartographic products—believe and insist that these have long-term value and will be of interest to researchers in the future. Do not allow your legitimate interests in the current and future uses of the products and the related financial benefits to interfere with the longer term preservation of these maps, geomatic records, etc. Remember that especially with modern technologies, if long term needs are not incorporated at the planning stage, that at the time when the system is no longer operational, the data may be lost and be irretrievable. Build up a long term friendly relationship with your jurisdictional archives—it is their mandate, not that of the map production agency, to ensure the preservation of cartographic archives.

To archives administrators:

Continue the commitment of your archives to “total archives”, the foundation of the Canadian archival system, and to the records in all media forms. Archivists with knowledge and understanding of cartography (including geomatics) are essential in your institution—hire and encourage them to develop their knowledge through study, contacts and formal education. Don’t make them feel that they are the “foster or step-children” of archives, as was felt by earlier generations”—recognize the value of the holdings for which they are responsible.

To cartographic archivists:

Your role is essential. Believe in the cartographic record and continue the fight for the record which you know is among the most valuable in your institution.

References:


Koeman, Cornelius (1968). Levels of historical evidence in early maps. *Imago Mundi*, XXII.


**Note:** The opinions expressed in this paper are those of the author and do not represent the position of the National Archives of Canada.
Building Aerial Photograph Referencing System and Providing Service on the Net

Noboru Kobori
Japan Map Center
4-9-6, Aobadai, Meguro-ku, Tokyo, 153-8522, JAPAN
Fax : +81-3-3485-5418
E-mail : kobori@jmc.or.jp

Abstract
Japan Map Center provides Aerial Photographs of Japan over 1 million sheets. We take the long time to get order, because we have to select the sheets for each users. We are going to start new project to select easy with referencing system. Thus we are faced many problem that is costs and time and so on. But we can select easy the aerial photographs. We can open this reference system on the net, must clear the right issues, how one solution using the digital watermark in graphic format.

Now
Basic Aerial Photographs (AP) of Japan has been photographed and maintained by the Geographical Survey Institute (GSI), Ministry of Construction and distributed by Japan Map Center (JMC). The number of APs add up to 1 million sheets, including archives collected before and after the Second World War. Owing to the vast number of sheets, it takes much time to specify the required AP when a customer orders. A customer requested AP is decided by specifying the required place and period. We should ask the customer under the four step;
1. Ask place and period
2. Decide flight course number and photo number
3. Finding the principal point
4. Writing the order form
The JMC asks the customers to the required period and to specify the place by finding the principal point on index maps based of the 1:50,000 topographical maps. This has been the method for distributing APs. If a customer requested a part of the sheets, we have always taken more time to decide the point that requested by a customer. Thus we takes about 15 minutes in case of contact printing of the sheet or 30 minutes in case of a part of the sheet.
Future

JMC has commenced building of an AP Referencing System on PC. For building the system, principal points were digitized on the index maps (see Figure 1). Of course, we should digitize four corner points, because each principal point calculates automatically the value of longitude and latitude. AP outline of the sheet that depending on the scale of the AP was also decided to digitize on the index map. Another work scanned the AP sheet accurately, saved graphic files by the unique number. Additional attributes such as AP number, scale, period are added (see Figure 2). This work attached principal point in the index map and AP graphic files. It takes about from 30 minutes to more 1 hour to input these data of one index map.

A customer can use this system and select the required area on the PC. Then, by selecting the photographed period, flight courses, principal points appear on the index map in display. When the customer selects the principal points near the required place, the most suitable AP is shown on display. Since selection of AP number takes place at this stage, order form for all area in the sheet AP becomes possible.

Using the new AP Referencing System, the time taken for each order for AP has been reduced. But the time is due to operate the PC that is using well or not. More effect using this referencing system is not necessary our stuff to decide the sheets by the customer every time. Furthermore, there is much significance on the fact that the customers are able to confirm their required AP on the display.

Figure 1. Digitizing principal points on index map based 1:50,000 topographical maps
Figure 2. Additional attributes such as AP number, scale, period

More future

This system functions on a single PC platform at present but it is expected that referencing and ordering can be made on the internet with the improvement of the system. We should resolve property right issues when transmitting AP pictures on the net. We are now beginning to wrestle the technical problems using the digital watermark in graphic format.
Problem

Most problem is time. It takes about 10 minutes to scan one AP sheet. Because this graphic file needs of high resolution. The graphic files have been about 100 thousand finished only 10%. We should scan the APs more quickly.

Second problem is property right issues. We have only one technique using the digital watermark in graphic format to start the transmitting AP picture on the net. We should confirm to safety the digital watermark, when download the user and cut the graphic files. If this watermark doesn’t use to transmit on the net, we should search another method.
On the automatic retrieval of updates in geographic databases based on geographic data matching tools

Thierry Badard
Institut Géographique National, Service de la Recherche/Laboratoire COGIT
2 à 4, avenue Pasteur, 94165 – Saint Mandé Cedex - France Thierry.Badard@ign.fr

Abstract

The integration in geographic databases of the updates delivered by a producer of reference data sets is still problematic. One of the main reasons is that the detection of updates in the new version of the database may be impossible to process. The purpose of this paper is to provide a generic tool for the automatic retrieval of the updates in geographic databases in order to make their integration easier. This mechanism is based on geographic data matching tools. Thus the implementation of systems of identifiers or time management is not necessary in the databases. The use of this process in the definition of new delivery modes of updates for geographic databases is also discussed.

1. Introduction

Nowadays, Geographic Information Systems (GIS) are considered to be truly analysis and decision-making tools. For that reason, a growing number of organisations invest in such systems and add specific information necessary to the tasks that they have the responsibility for. However, the implementation of such systems is difficult and relatively expensive. That is why these institutions purchase reference data sets from geographic information producers. From these they are able to develop their own information systems. Thus, these organisations are considered, by the producer, as users of reference geographic data sets.

To carry out their assignments, these institutions clearly need updates from the producer, in order to have the most faithful and realistic image of the geographic space reality. However, at the present time, the integration of these updates is still problematic. Indeed, few different delivery modes of the updates are available for geographic databases. Thus the whole up-to-date database is often delivered even if only 10% of the objects stored in the database have changed (this is a common estimation of the rate of evolution per year for geographic databases [Raynal, 1996]). This implies that users have to integrate the whole new version of the database in their information systems, which may induce significant risks of information loss or leave the database in an inconsistent state. Otherwise, they have to retrieve the updates by themselves. This task may fail if a reliable and stable system of identifiers is not implemented in the reference database.

The purpose of this paper is to provide a generic tool for the automatic retrieval of the updates in geographic databases in order to make their integration easier. This mechanism is based on geographic data matching tools implemented at the COGIT laboratory of IGN (the French national mapping agency). They allow the detection of the geographical entities that represent the same phenomenon in two representations of the real world. This process of detection of the updates is also a component of a wider investigation, which aims at developing a generic updating tool for multi-scale databases. Thus, the details of this mechanism are presented and illustrated by the results of a first experimentation made with the road network of the BDCARTO® (a 1:100 000 geographic database produced by the IGN which covers the French territory). The use of this process in the definition of a new delivery mode of the updates for the geographic databases is also discussed.
2. The geographic data matching

Geographic data matching is akin to the notion of conflation defined in [Saalfeld, 1988; Laurini and Thompson, 1992]. It consists in the computation of correspondence relationships between sets of geographical entities that represent the same phenomenon in two representations of the real world [Lupien and Moreland, 1987; Lemarié and Raynal, 1996; Lemarié and Bucaille, 1998]. The two representations may have very different scales, levels of abstraction and/or representations. Such a mechanism has been developed at the COGIT laboratory of IGN (the French national mapping agency) for several years. It relies on the implementation of numerous algorithms involving all geographical information levels (i.e. semantic, geometrical and topologic). It can be detailed as below (see Figure 1).

![General structure of the geographic data matching mechanism](image)

Figure 1. General structure of the geographic data matching mechanism

So, the data matching mechanism adopts a bottom-up behaviour. Data matching is first performed on the geometrical primitives. Correspondence relationships between simple objects are then deduced. Complex objects (i.e. composed of simple objects and/or of complex objects) are finally reconstructed, because they depend on the matching of the simple objects. The geometry of complex objects is indeed the union of the features describing the simple objects.

This process is mostly automatic: only the enhancement step for node matching involves an interaction with a user. This interaction is however assisted by the automated detection of conflicting correspondences. The different steps of the geographic data matching process are detailed in the next sections.

2.1. Data matching of the geometrical features

This part of the data matching process is generic and independent of the databases’ schemas. It can be divided in three main steps depending on which kind of geometrical primitive is processed: points, lines or areas. Methods involved in this step are purely geometrical and topologic. Semantic information is never used. The data matching of areas is totally independent of the others. But there are interactions between the data matching steps for points and lines.

2.1.1. The node matching step

A reference database is first chosen (the other is named the comparison database) in order to define how the process will compute the correspondence relationships (i.e. from the reference database to the comparison one). A search area for each node of the reference database is then defined by means of a Euclidean distance
threshold. The value of this threshold is generally based on the knowledge of the scales and of the root mean square error on the positions of the features stored in the databases. For each node of the reference database, all the nodes of the comparison database included in the tolerance area are searched.

In order to reduce the number of matching candidates, geometrical and topologic tests are processed. For instance, a test of geometrical equality is performed and the number of arcs connected to the nodes is checked. When node matching is finished, lines are processed. This step is detailed in the next section.

2.1.2. The line matching step

This step of the data matching process involves specific algorithms inspired from those described in [Abbas, 1994] and especially based on the Hausdorff distance (see Figure 2). This is a distance which accounts for differences not only between the positions but also between the shapes of the geometrical features.

![Figure 2. Definition of the Hausdorff distance](image_url)

Since the reference database is already defined, only one component of the Hausdorff distance is computed. For each line of the reference database, all lines of the comparison database included in the tolerance area (see Figure 2) are considered as matching candidates. As before, the definition of a threshold is necessary. This threshold is here expressed in terms of Hausdorff distance and defined by the knowledge of the scales and of the root mean square error on the positions and shapes of the geometrical features.

As in the node matching, various geometrical and topologic tests are performed in order to retrieve consistent correspondences between objects. Notably, relationships previously processed between nodes are involved in this step. For instance, each matched node is scanned and angles between all the connected arcs are checked. Moreover, an algorithm of graph exploration is implemented in order to retrieve missing links. When line matching is completed, an enhancement and checking step is triggered on the correspondence relationships. Details are presented in the next section.

2.1.3. Enhancement and checking step of the correspondence relationships

To provide n-to-m relationships (i.e. n objects of the reference database are corresponding to m objects of the comparison database), 1-to-1, 1-to-n, and n-to-1 correspondences between nodes are first grouped. This grouping allows the detection of crossroads which have been simplified or detailed from database to the other. As described in [Gabay and Doytsher, 1994], the correspondence relationships established between lines are then scanned in order to enhance the node matching by grouping the intermediary vertices (points that are missed by the node matching process) with the sets of matched lines. This allows the detection of mismatched features which could have been produced during the node matching step. Simplified or detailed crossroads are finally processed by adding the linear primitives to the corresponding sets of nodes. At the end of this enhancement and checking step, node and line correspondences are consistent. Area matching can be now performed. This step of the data matching process is detailed in the next section.
2.1.4. The area matching step

This step of the data matching process is totally independent. It includes algorithms based on the computation of surface intersections to account for the full “surfaceness” of the geographical entities. Methods based on boundaries only have indeed the drawback to be much too sensitive to shape differences, which entails that numerous valid correspondence relationships are not established. In order to filter the data sets and find all the polygons which can be matched, the inclusion function (see Figure 3) defined in [Lemarié, 1996; Bel Hadj Ali, 1997; Vauglin and Bel Hadj Ali, 1998] is computed.

\[ I(A, B) = \frac{S(A \cap B)}{\min[S(A), S(B)]} \]

**Figure 3.** Definition of the inclusion function

To complement the inclusion function, a surface distance is also calculated. This distance presented in [Lemarié, 1996; Bel Hadj Ali, 1997; Vauglin and Bel Hadj Ali, 1998] is defined below (see Figure 4):

\[ d(A, B) = 1 - \frac{S(A \cap B)}{S(A \cup B)} \]

**Figure 4.** Definition of the surface distance

As in the node and line matching steps, this process implies to define tolerance thresholds in respect with the scales and the different quality indicators of both data sets. Finally, in order to retrieve the actual n-to-m correspondences between areas, all correspondence relationships are grouped. Data matching of the geometrical features is now achieved. The correspondences between semantic objects have to be processed. Details of this step are presented in the next section.

2.2. Data matching of the simple and complex objects

This step is relatively simple. If the geometrical primitives of two simple objects have been matched, objects are matched. To avoid inconsistent object matching, constraints on schemas’ classes are necessary. For instance, without such a system of class constraints, close roads and rivers could be matched. This mechanism allows also the management of the differences in databases’ schemas, which may occur if data sets present different content specifications or levels of abstraction. As before, correspondences between complex objects are deduced from simple objects. If the components of two complex objects have been matched, complex objects are matched. This ends the data matching process, retrieval of updates can now be triggered.

3. Towards the extraction of updates

Classifying updates implies to first define the semantics of the spatio-temporal evolutions undergone by the geographical entities. This is the subject of the present section.
3.1. About semantics and logic of the spatio-temporal evolutions

Several typologies of the spatio-temporal evolutions have already been established in the literature. Some approaches are closest to the implementation of time in GIS, some are closest to the modelling of real world evolutions, but are more difficult to implement.

For instance, creation and deletion of entities are the only operations considered in [Poupart-Lavoie, 1997]. A modification of an object in the database can always be replaced by the creation of the new entity and the deletion of the old one. This voluntary minimal typology has been implemented because it allows a simple and efficient delivery of the updates, making their integration easier. Nevertheless, the actual underlying nature of the spatio-temporal evolutions is lost.

Another interesting approach is described in [Cheylan et al., 1994]. Four main types of spatial entities are defined: fixed spatial entity with time-varying attribute, modifiable spatial entity (i.e. an entity with a fixed boundary but composed of a set of entities with time-varying shape and attribute), distorted spatial entity (i.e. an entity composed of a set of entities with time-varying attribute, shape and area) and transformed spatial entity (i.e. the position and the shape of such an entity are varying with time). This typology aims at identifying the spatial evolutions and attributes’ modifications undergone by a set of geographical entities. Close to real world evolutions, the implementation of this typology in a process of update extraction seems to be difficult.

A rich and easier to be implemented typology has been proposed in [Claramunt et al., 1994]. It is composed of sets of basic processes, transformation processes and movements. Basic processes are apparition, disappearance and stability. Transformation processes are expansion, contraction, distortion. Movements are displacement and rotation. This typology provides a relevant description of the evolutions undergone by a single geographical entity. Nevertheless, this typology is too restricted to represent some modifications of particular geographical objects or groups of objects, taken as a whole. For instance, only the succession of an apparition and a disappearance could allow the description of the modification of a crossroads in a roundabout. Besides, how to represent merging, splitting or aggregation of geographical entities with it ? Moreover, attribute modifications are not taken into account.

We propose thus to extend the typology presented in [Claramunt et al., 1994], in order to better take into account all the missing modifications. The formulation of the different types of evolution is voluntarily closer to the changes encountered in geographic databases. It aims at providing users with a detailed information on the nature of the changes in order to make their integration in geographic information systems easier. This new typology can be decomposed in eight main types:

- Creation (identical to apparition).
- Deletion (identical to disappearance).
- Splitting.
- Merging.
- Aggregation.
- Geometrical modification (it includes distortion and displacement).
- Semantic modification.
- Stability.

This typology allow thus the description of the evolutions undergone by one or a set of spatial entities. Besides, attribute modifications and all kinds of geographical objects are taken into account. The implementation of this typology in a process of update extraction between two versions of a same geographic database is detailed in the next section.
3.2. The process of update extraction

In order to extract updates and to classify all modifications in the different types previously defined, it is necessary to analyse the correspondence relationships computed during the data matching step. This analysis deals with the cardinalities of the relationships established between reference and comparison databases. In the following part of this section, we assume that the reference dataset is the up-to-date version of the database and the comparison one is the old version. The purely geometrical modifications are first processed and the semantic differences are then detected in order to extract all objects that have semantically and/or geometrically changed. This analysis which is at the core of the mechanism of update extraction proceeds as detailed below:

- Unmatched objects of the reference dataset are considered as the newly created geographical entities.
- Unmatched objects of the comparison database are considered as the deleted objects.
- Objects referenced in a 1-to-n relationship (with n>1) denote a merging operation.
- Objects referenced in a n-to-1 relationship (with n>1) are involved in a splitting operation.
- Objects involved in a n-to-m relationship (with n>1 and m>1) have undergone an aggregation operation.
- Objects involved in a 1-to-1 relationship are more precisely analysed. A test of geometrical equality is first performed. If the objects have a different geometry, a geometrical modification is detected. If they have undergone a geometrical modification or not, their attributes are then checked in order to find semantic differences. If at least one of their attributes differs (different value for a common attribute, different attributes found, ...), a semantic modification is detected. So, it is only in this step that the semantic level of the geographic information stored in the databases is used.
- All remaining objects are considered as unchanged (i.e. stable).

So, this process of update extraction allows not only the retrieval of simple modifications (as creation or deletion) but also the detection of complex updates (as both semantic and geometrical changes). Objects involved in a merging, splitting or aggregation operation are not semantically analysed (i.e. their attributes are not checked). It is not necessary because such modifications are generally due to changes in the capture and/or content specifications of the database. The update is nevertheless detected and the new values of the attributes are delivered with the up-to-date version of the database.

Due to the different thresholds used during the data matching process, the actual nature of an update may not be detected. For instance, an object may be left unmatched because the threshold value was too tight. This results in the detection of a creation or a deletion, even though it is a geometrical modification that should be retrieved. A similar problem may occur with the merging, splitting or aggregation operations on the elements of a network (road, rivers, ...). For instance, due to the tolerance allowed in terms of position and shape by the Hausdorff distance (see Figure 2), two different objects may be considered as merged with another one even if they do not share the same supporting primitives. So, the choice of the different thresholds involved during the data matching process is important and has to be made judiciously if the actual nature of the updates wants to be retrieved. Several experiments carried out at the COGIT laboratory of the IGN seems to prove that the different thresholds have to be set close to the value of the root mean square error on the positions and shapes of the objects stored in the databases. The automatic valuation of these different thresholds is presently a part of a PhD research engaged in this laboratory in order to make the data matching process fully adaptive.

Nevertheless, this process of update extraction based on topologic and geometrical data matching tools is complete: all the changes are detected, only the nature of the extracted updates may not be right. It is fully automatic and dispenses with the implementation of a reliable and stable system of identifiers or time management in the databases. Contrary to [Uitermark et al., 1998] in which only the changes on the buildings may be detected, all kinds of geographical entities stored in the databases may be processed. This mechanism can thus be considered as generic.

Moreover, the updating information provided by this process is detailed and close to the modifications per-
formed in the geographic databases. It allows users to automatically retrieve all the updates in the whole up-to-date version of a database delivered by a producer. The integration of these updates in their information systems is then easier: they are able to choose the relevant updates and to control their effects especially on their own added information. Users are thus not compelled to integrate the whole new version of the database in their information systems, which may induce significant risks of information loss or leave the database in an inconsistent state.

This last point addresses the problem of the delivery of updating information for geographic databases. Indeed, few different delivery modes are now available. Only the whole up-to-date database is mainly delivered to users. The definition of a new mode dedicated to the delivery of updates is presented in the next section.

4. Towards a new delivery mode of updating information for geographic databases

If few different delivery modes are now available and used, it is mainly due to the specificity of geographic information. Indeed, if for “classical” databases, it is relatively easy to represent the changes in the value of an attribute, it is more difficult to describe the evolutions of the geometrical information stored in the geographic databases. Nevertheless, the notions of state versioning and logs could be easily adapted to the geographic information domain in order to define a new delivery mode of updates.

In state versioning, changes are delivered as couples of objects (new/old) or simply as couples of new objects/references on old objects, if such a reliable system of identifiers is implemented. With logs, no objects are delivered but only the description of the evolutions that objects referenced by their identifiers (which assumes the existence of such a stable and reliable system) have undergone.

The information structuring provided by the process of update retrieval seems to be close to the one involved with state versioning and logs. As updates are extracted from the correspondence relationships, updated objects or groups of objects are known and the “temporal succession” relationships are established between the different versions. Moreover, the type of the update is known. A detailed and minimal delivery mode of the updating information could thus be to deliver the geometry of the new and old versions of the updated objects and to specify the nature of the evolutions in a log. This is what we have named “updating deltas”.

In order to illustrate this delivery mode, results provided by the process of update extraction on the road network of the BDCARTO® (a 1:100 000 geographic database produced by the IGN which covers the French territory) are now presented.

Figure 5. Sample of the “updating deltas” processed with the road network of the BDCARTO® (City of Caen, Calvados)
The first figure (see Figure 5, left) is a geometrical representation of the up-to-date objects (i.e. included in the up-to-date version of the BDCARTO®). We have named it “delta +1”. The different colours on the figure indicate the nature of the evolutions. Objects stemming from a creation, merging, splitting, aggregation, semantic and/or geometrical modification operations are thus represented.

The second figure (see Figure 5, right) is a geometrical representation of the objects that have undergone an evolution (i.e. objects which have been updated and are included in the old version of the BDCARTO®). We have named it “delta -1”. As before, the different colours deal with the nature of the evolutions. Objects that have been deleted, merged, split, aggregated, semantically and/or geometrically modified are represented.

The nature of the evolutions and the detail of the semantic modifications are provided in a log (Figure 6). This update log uses the system of identifiers implemented in the BDCARTO®. Nevertheless, when such a system is not implemented, it is easy to deduce another structure for the delivery of the same information. For instance, for each evolution, it would be possible to deliver the geometry of the new and old objects, the type of updates, and eventually a list of the modified attributes with their old and new values. A temporary system of references, which allows the identification of the objects within the data exchange only, could be another solution.

**Update log: BDCARTO® v1 -> BDCARTO® v2**

**Creation:**
- NEUD_ROU -> BDC_ID: 991584043
- TRON_ROU -> BDC_ID: 992300803
- TRON_BAC -> BDC_ID: 269
- LIAISON -> BDC_ID: 153
- CARR_CX -> BDC_ID: 1375
- ROUTE -> BDC_ID: 39095

**Deletion:**
- NEUD_ROU -> BDC_ID: 140005977
- TRON_ROU -> BDC_ID: 140136378
- ROUTE -> BDC_ID: 24254

**Merging:**
- TRON_ROU : 140004577 -> TRON_ROU : 140004577
- TRON_ROU : 990157858 -> TRON_ROU : 992260352

**Splitting:**
- TRON_ROU : 140004259 -> TRON_ROU : 992260418
- TRON_ROU : 140002569 -> TRON_ROU : 992301214
- TRON_ROU : 991744402 -> TRON_ROU : 992300352

**Aggregation:**
- TRON_ROU : 140003594 -> TRON_ROU : 990746918
- TRON_ROU : 991520666 -> TRON_ROU : 992301382

**Semantic modifications:**
- NEUD_ROU : 990039893 -> NEUD_ROU : 990039893, TYPE: 4 -> 5
- COTE: 0 -> 9999

**Geometrical modifications:**
- NEUD_ROU : 990003893 -> NEUD_ROU : 9903893, TYPE: 4 -> 5
- COTE: 0 -> 9999
- NEUD_ROU : 140004360 -> TRON_ROU : 140004360, NBV_TOT: 1 -> 3

**Figure 6.** Sample of the evolution log processed with the road network of the BDCARTO® (Department of Calvados)
This delivery mode of updating information provides users with a minimal (in terms of number of objects) but detailed description of the evolutions undergone by the reference data sets provided by a producer. The integration of these updates in their information systems is made easier: they are able to choose the relevant updates and to control their effects especially on their own added information. A whole old or up-to-date version of the database could be delivered with the “updating deltas” in order to allow users to build the contextual information or to retrieve some unfortunately deleted objects. This notion of contextual information is important because it addresses the underlying problem of data exchange format. Indeed, not all the data exchange formats dedicated to the geographic information allow an easy delivery of the “updating deltas”. For instance, with EDIGéO [AFNOR, 1992], it is possible only if a non topologic mode is used and if a fictitious geometry for previous objects is provided. It may be a great danger for an unaware user. Reflections about structure and data exchange formats are presently in progress at the IGN for the delivery of the updates for the BDCARTO®.

5. Conclusion and outlooks

This process of update retrieval based on topologic and geometrical data matching tools allows the automatic extraction of the evolutions between two versions of a same geographic database. The implementation in the databases of a reliable and stable system of identifiers or time management is not necessary. All kinds of geographical entities stored in the databases may be processed with it. So, it can be considered as generic. Moreover, the updating information provided by this process is detailed and close to the modifications performed in the geographic databases. It allows users to automatically retrieve all the updates in the whole up-to-date version of a database delivered by a producer. The integration of these updates in their information systems is made easier: they are able to choose the relevant updates and to control their effects especially on their own added information. Users are thus not compelled to integrate the whole new version of the database in their information systems, which may induce significant risks of information loss or leave the database in an inconsistent state.

The use of this process in the definition of a new delivery mode of the updates for geographic databases has also been presented and detailed. The notion of “updating deltas” has thus been defined and illustrated by the results of a first experimentation. They are first tools which contribute to an easier integration of the updates delivered by a producer into the users’ databases.

The definition of this process is in fact part of a wider investigation engaged at the COGIT laboratory of the IGN, which aims at developing an automatic updating tool for the multi-representation geographic databases. The structure of this more general mechanism is particularly presented in [Badard, 1998].

References


AFNOR. (1992). Echange de données informatisées dans le cadre de l’information géographique (EDIGéO), AFNOR, ISSN 0335-3951.


Poupart-Lavoie, G. (1997). Développement d’une méthode de transfert des mises à jour de données à référence spatiale. Mémoire de maîtrise ès sciences (M. Sc.). Centre de Recherche en Géomatique, Université Laval, Québec, Canada.


Abstract

In a world where information sharing is more and more important, a strong demand for direct access to heterogeneous and large volumes of geospatial data drives the development of powerful softwares resolving many problems related to geospatial data sharing. Among those softwares, OGDI, allowing direct access to heterogeneous data and geospatial data warehouse, providing the end-user with a homogeneous, integrated, seamless and voluminous geospatial database, are proven solutions to overcome the geospatial data barrier. This paper provides an analysis of the geospatial data barrier, highlighting major transformation and integration problems related to the nature of georeferenced data. It also provides a description of some of the major geospatial data warehouse technology, indicating how this technology and related products offer a fast and efficient access to integrated geospatial data.

Introduction - Why Using a Geospatial Data Warehouse?

In a world where more and more people have access to more and more data and information, notably from the Internet, it is completely normal to observe a greater request for geospatial products. Although this demand is very keen, it runs up sometimes against major obstacles. Just like for the office and graphics applications (but with a higher degree of difficulty), the access to the geospatial data is complicated by lots of different and incompatible file formats. More serious are the problems related to the heterogeneity of the geospatial referencials. Moreover, management of geospatial data often requires high capacity storage hardware, as its nature (complex and lengthy geometric definition along with descriptive data) creates rapidly large data stores. As Gouin stated [Gouin 97], we can easily rule that the whole set of these problems constitutes a significant barrier for the diffusion and the use of geospatial data.

Two initiatives, lead by the Canadian Department of National Defence (DND) in collaboration with the industry, offer an appropriate solution to this problem. The first initiative, the Open Geospatial Datastore Interface, (better knew under name OGDI and supported by Global Geomatics (www.globalgeo.com) and by the Interoperability Information Institute (http://132.156.33.161/iii)), is targeted to break the geospatial data barrier, while the geospatial data warehouse, whose development is lead by Cubewerx (www.cubewerx.com) is targeted to deliver to end-users a coherent, uniform and voluminous geospatial data store. These two projects are particularly important for DND as the military context implies use and sharing of lots of geospatial data, which often are heterogeneous. Some military standards have been proposed and are widely used (e.g., Digest), but there is still many situations where standardise data are not available, both in war / peace keeping operations and in domestic operations (e.g., Ice Storm Crisis in January 1998 in Eastern Canada). Data warehouse is very important for DND in situations where simultaneous access is required to a complete and integrated geospatial database. This paper will focus on geospatial data warehousing technology, although some discussions will be held on OGDI, as the two products are closely tight.
An analysis of the problems of access and diffusion of geospatial data.

Definition of the problem

Since the development of computer-assisted cartography and geomatics technologies (remote sensing, digital photogrammetry, GIS, etc.), lots of money have been invested to produce geospatial data sets. In the early days of geomatics, many organisations were building their own data sets, based on a particular geospatial representation (referential) that fit their needs. Billions of dollars were invested on those initiatives and many of these data sets are still very valuable. The general scarceness of money for public organisations forces them to seek to re-use the existing data rather than to recreate them. As a consequence, many organisations need to exchange data between them.

The Internet has created a whole new market for geospatial products. Formerly, access to geospatial data was difficult and limited through the use of high-end software. The arrival of desktop GIS improved the situation, but this kind of software remained nevertheless out of the range of general public. The Internet brings geospatial data to the users through the use of simple but efficient data browser and navigator, and it democratises the access to geospatial data. Consequently, every new GIS product line produced nowadays is implementing Internet functions, in such a basic or more complex way. Serving data through the Internet for GIS is not a must, it is de rigueur!

Traditionally, GIS and computer-assisted cartography softwares were using internal databases to store both geometric and descriptive information. Although some of these databases were very powerful, none of these last offered as many functionality and performance that the high-end commercial relational DBMS. Bigger data sets, of an order of magnitude of more than 100 GB, require a new paradigm for the storage and management of geospatial data. Data warehousing became a natural avenue to store efficiently large data sets, more especially as this technology would allow to resolve some of the problems related to the geospatial data barrier as well as the Internet access to the data through client/server services. In the following paragraphs, we will emphasise the technical and conceptual problems related to heterogeneity and access to geospatial data while indicating how data warehousing technology may be used to solve these problems.

Few words about OGDI

OGDI is an application programming interface and a network protocol designed to be integrated into GIS client applications. It allows to connect to multiple data sources, supporting native data. Proprietary translators are no longer needed as OGDI provides native access mechanisms to numerous data formats through a client / server architecture. Thus, a GIS application or a thin web client may establish several connections to different data sources, either locally or remotely, from the Internet or an intranet. OGDI supports vector, raster and matrix format, from several different data sources.

OGDI has built-in mechanisms that allow projection, coordinate system and datum transformation. Thus, heterogeneous data, with different datum, projection and coordinate system may be integrated into a single and coherent GIS application. Figure 1 shows a generic architecture for OGDI. OGDI is an open-source initiative similar to Linux where members of the III may propose improvements of their own and thus make evolve the OGDI.
Heterogeneity of geospatial documents

As indicated earlier, geospatial data formats are numerous and heterogeneous. Integration of different maps or coverages coming from different platforms into one spatially and geometrically coherent GIS application is a challenging process that requires a lot of time and planning. Not all GIS applications do offer import capabilities to the more common GIS formats. Consequently, users may need to export their data to an intermediate format and then import it into the targeted format. There are strong risks that geometrical information is lost during the export and import process, thus decreasing the quality of the data final. Moreover, some data formats are based on a single data file access (shapefiles, dgn, dxf, geotiff, ...) while others are based on a structured database and indexes (vmap, cadrg, ...), thus complicating the management of these different types of data.

The other major problem, when dealing with multiple data sources is related with heterogeneous geospatial referentials. Indeed, in the case of building a spatially and geometrically coherent GIS from numerous and different data sources, it is likely that the system designer encounters difficulties of integration of the data related to heterogeneity of the geospatial referentials. Problems related to the integration of heterogeneous geospatial data are the following:

- Datum heterogeneity
- Projection heterogeneity
- Coordinate system heterogeneity
- Lack of meta-data (e.g., do not provide projection, datum, etc.)
Projection conversion problems are often encountered when trying to integrate differently projected data sets into a particular projection system. Not every GIS has the algorithms to convert the projection. Though some GIS have these capabilities, their limited number of supported projection transformations limits their capabilities to transform a large number of projections.

More fundamental problems may appear when integrating different data sets that share some identical features. These problems are related to the content of the data sets and the representation of the data. For example, the designer may get three different and incompatible representations of the same feature coming from different data sets when getting data from various sources. Or he may find the same object symbolised differently. Charron [Charron 95] offers a detailed analysis about this problem. Considering these problems, it is difficult to define an automatic mechanism that would clean the integrated data and choose the best data.

Other problems appear when the designer seeks data or try to access them. It is difficult to know all the data available and even if modern and powerful mechanisms (in particular Georeferenced Digital Libraries - GDL) do exist to find and locate data, lots of problems are still present in current GDL, as Létourneau [Létourneau, 98a, Létourneau et al., 98b] described them in different publications. Among other problems, we may cite direct access to retrieved data sets, which is not a standard function of current GDL.

Different mechanisms may be used to overcome those difficulties. In the next sections, we will describe some of them, which are part of a continuous R&D effort in geomatics at DREV and from various other agencies of the Department of National Defence and Natural Resources Canada.

A geospatial data infrastructure based on a data warehouse

Three major projects were conducted at DREV in geomatics (OGDI, Geospatial Data Warehouse and GDL). Each of those projects leads to the development or enhancements of products that may be use in the context described above. These products have been developed in order to simplify data access, overcome heterogeneous geospatial referencials, offer a consistent and coherent data set and facilitate geospatial data retrieval. In the following paragraphs, we will mainly focus on data warehousing technologies, while indicating how OGDI and GDL may be linked to the geospatial data warehouse.

Geospatial data warehouse - definition and architecture

Numerous definitions have been used about standard data warehousing. One of the nicest has been defined by Devlin [Devlin 97] and is stated as this: a data warehouse is a single, complete, and a consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context. Rawling [Rawling and Kucera 97] defined the geospatial data warehouse this way: a geospatial data warehouse is a collection of subject-oriented spatial data that is of known quality, is non-volatile, is time variant and includes the basic tools to access and extract information. Keighan [Keighan 99] indicates that a data warehouse is a database environment that organises operational data in one or many distributed repositories for efficient access.
From our point of view, a geospatial data warehouse has the following characteristics:

- Transformation and integration mechanisms allow the data to be stored in a coherent and consistent geospatial referential.
- Data is stored in a standard commercial RDBMS, allowing terabytes of data to be stored.
- Data is accessed through a client / server mechanism, allowing Internet or intranet access.
- Required time to access data is directly proportional to data extracted.
- Geospatial data is integrated in order to provide a seamless geospatial database.
- Generalisation mechanisms provide a relatively constant access time when loading data from wider area.

Of all these characteristics, the first is indisputably the most significant, as it solves almost completely the heterogeneity of geospatial referencials, at least for vector products and the majority of raster products. In the warehouse, geospatial data share the same projection, geodetic reference system and coordinate system. There are also lots of advantages to use a warehouse in order to disseminate geospatial data. Indeed, the updates are facilitated, as the administrator has only to update data in the central warehouse instead of sending updates to all sites. The warehouse may also be used for multiple purposes, for either data analysis, general usage within a GIS or simple visualisation. Finally, the client / server services allow multiple users to connect to the warehouse through TCP/IP protocol, from the Internet or an intranet. It also uses standard commercial RDBMS to store data in spatial tables. The geospatial data warehouse is also designed to store efficiently attribute data as well as meta-data. This architecture is very powerful to build an integrated application, where all data is located in the same database. The geospatial data warehouse architecture is summarised in figure 2. In the following section, we will describe more in-depth the main functionnalities of a geospatial data warehouse.
Functionalities of a data warehouse

1- Data loading

Without a robust and powerful data loading mechanism, a geospatial data warehouse would be useless. Indeed, the complexity involved when dealing with geospatial data requires a very powerful and strong mechanism, capable of dealing with projections, geodetic reference system, coordinate systems, etc.

Two different approaches have been used to populate the data warehouse. The first one, which is very interesting, uses the OGDI as a mechanism to access heterogeneous geospatial and descriptive data and Aconvert@ it to a consistent geospatial referential. While this is very functional for vector data, it is more challenging for raster and matrix data, as the OGDI was built to create a composed, sampled and scaled version of the original raster layer. Although this characteristic is very convenient because the image sent by the OGDI to the client is exactly the image that should be displayed on the screen, it challenges a data-converter environment, since it requires from the client to define a bounding rectangle and a resolution of the image that fit exactly the whole data set in order to import the whole data and not only a composed, scaled and sampled version of the data. The converter will use available meta-data from the raster data set to define those parameters, but some data sets do not provide enough meta-data to set correctly the parameters. Consequently, some additional drivers, accessing directly the data have been built, in order to access more efficiently raster data.

2- Data generalisation

As data sets became larger and larger, some problems arose when trying to access data covering large territory. For example, in the data warehouse, we have a worldwide coverage of elevation data. Although it is possible to query the warehouse to view the entire data set, this query would require some minutes to execute, even with very powerful computers, which is contrary of data warehouse theory, which require fast access to data. Some solutions are available to overcome this problem. The first one is to maintain different data set for the same feature for different scales. This approach is functional but may cause some problems, such as how to manage updates, features not consistent from a data set to another, etc. Some features may appear in one data set but not on the other one, which may cause confusion for the user. The other solution, which is more elegant, uses generalisation and aggregation with one data set and derives generalised data from this data source, hence providing a consistent data set whatever the scale is. Although generalisation, in theory, works well, it still requires a lot of R&D efforts to be fully functional.

Basic but efficient generalisation mechanisms have been implemented in the current version of our data warehouse. It is very suitable for matrix data, like elevation or raster data like satellite images and works well with certain vector data (shore lines, lakes, rivers), but the implementation of the mechanism does not support topology. Thus, although we would be able to generalise a road coverage, that would not make sense because we cannot efficiently generate a national road network from all the available road segments. The generalisation mechanism creates pre-generalised data sets of different resolutions. Consequently, the data warehouse does not do on-the-fly generalisation, providing a faster access to the coverage, from worldwide coverage to regional one, as illustrated in figure 3. Attribute data may still be queried from the generalised data sets, as the query localisation expressed in longitude and latitude is sent to the original data set where the query is performed.
3- Data management and structure

The data stored in the warehouse is managed through the use of coverages and features. For example, a coverage would be transportation and features of this coverage would be roads, railroads, port, airport, etc. This structure helps the user to browse for data sets in the warehouse.

In order to provide fast access to large volumes of data with variable density, data partition has been implemented into the geospatial data warehouse. Data partitioning is a mechanism that breaks a whole data set in smaller, easier to manage blocks of data. The larger a data set is, the more it requires partitions to be created. Partitions are also a function of density of features. For example, figure 4 illustrates the partition map for road features. The partition mechanism is based on quad tree structure. More roads partitions are created in eastern North America and Europe than in NorthWest Territories, as there are several more roads in eastern North America and Europe than in NWT. Partitions are created for point, line and polygon vector data only, as matrix and raster data is evenly distributed over the territory.

A new feature that is implemented in the warehouse is a default symbology manager. This feature allows defining default symbology for features in the warehouse, allowing a consistent representation of the features through different clients. Many point features, like quarries and mines, buoys, reefs, church, telecommunication towers, . . . will be consistently symbolised through the warehouse, as well as roads based on their attributes (two lanes highway, dirt roads, etc.). This default symbology will be rendered through the use of rules based on current field of view or scale. The features symbolised through this mechanism are viewable on the Web output from the warehouse. Of course, other commercial client applications should be able to manage and interpret this information in order to use it efficiently.
A basic topology manager for vector data, based on the Eigenhoffer operations, is implemented into the warehouse. This topology manager allows the user to perform basic spatial analysis over vector data, e.g., intersection, inclusion, touch, ... Eventually, this topological structure would be used to build topological generalisation of a road network. This type of generalisation is based on attribute data and connectivity of road segments.

4- Data access

Three different mechanisms are used to access data. The first mechanism uses OGDI to allow client applications to access data directly within the warehouse. This is a very interesting approach, since the warehouse may be used as a key component in a GIS built from OGDI data sources. Attribute data and meta-data contained in the warehouse are also available from the OGDI. OGDI access to the warehouse allows the user to view vector, raster and matrix data, to perform queries on attribute data and to retrieve meta-data for specific data set.

The other access mechanism is through Cubestor CGI tools. These tools allow the publication of warehouse content on the Internet using basic HTML coding. Created for browsing and navigating, this service display an image of the data through JPEG or PNG (Portable Network Graphic) files and offers the user simple mechanisms to zoom, pan and query data. Users may also query multiple data warehouse, select coverages and features and reorder them. As it uses original and generalised data, this application is useful to browse data, whatever the scale, from worldwide view to regional view. As this application uses standard HTML form, the interactivity with the map is limited. However, through the use of Java scripts and applets, interactivity will be added to this access mechanism. Users will be able to choose different projections, perform custom zooms, define basic symbology, etc. The Java applets will simply wrap the actual CGI elements. Figure 5 shows this application. The third mechanism is used by various commercial GIS applications and provides a direct access to the warehouse content.

Figure 5. Basic HTML access to the data warehouse
5- Georeferenced Digital Library linked to the data warehouse

A georeferenced digital library (GDL) is a computerised geospatial index of georeferenced documents including meta data, making it possible to know the contents and nature and to visualise these complete documents or as abstracts. It has a graphic interface, which makes it possible to define or to select the area of interest to which relates the information retrieval, as well as the functions necessary to the definition of the search parameters. Traditionally, the GDL is a standalone application with its own database and graphic engine, requiring data retrieval and database population. In order to improve the efficiency of the GDL, we want to couple it with the warehouse, using the Oracle underlying database for meta-data management, query and retrieval and the Cubestor CGI product with Java applets as the graphical interface. The user would have also a direct link to the retrieved data and may either use Cubestor CGI or OGDI to access it, depending of his needs.

The main difference between a GDL and a librarian is related to the spatio-descriptive query of the GDL. The librarian provides only a list of coverages the user, while the GDL provides the same list in addition to providing the necessary tools to define spatial, temporal and meta-data query.

The integrated GDL would allow the user to query the warehouse over a particular region using either a rectangle, a point, a polyline, a circle or a place name, instead of using a bounding rectangle. This new approach will be particularly interesting for retrieval of documents along a linear element, for example retrieval of aerial photographs along a power line. Instead of retrieving the whole photos contained in the bounding rectangle, only the photos over or near the power line would be retrieved, avoiding unnecessary data to be retrieved.

The GDL will also allow the user to perform standard SQL queries over the meta-data or to use the best match query mechanism to retrieve data that would not be retrieved using standard SQL queries. Best match query mechanism is based on fuzzy logic and will allow the following query type. For example, a user may want to query an aerial photograph coverage to find aerial photos of scale of 1:20 000. If the database contains only photos of 1:25 000, a standard SQL query would not retrieve the photos. Instead, by using the best match approach, the user will receive results that may not be exactly what they requested, but at least, they will receive the best data available.

As the GDL is directly linked to the data warehouse, it features functions that allow the user to connect directly to the data in order to use it. This is a real advantage over classic GDL that send the user only a list containing references to documents. It represents a real step toward a fully integrated application.

Conclusion

The geospatial data warehouse is a powerful tool, which is one of the major foundations of a Global Geospatial Data Infrastructure. The ability of the warehouse to integrate heterogeneous geospatial data in a homogeneous geospatial referential offers a wonderful opportunity to build many geomatics applications around this infrastructure. Used wisely in combination with OGDI, geospatial data warehousing may unlock the potential of enterprise and government geospatial data. Many analyses, not feasible because geospatial data was unintegrated and heterogeneous, may now be possible, as GIS and decision support applications will access homogeneous and integrated data from either the warehouse or through OGDI drivers.
References

Charron J. (1995), Développement d’un processus de sélection des meilleures sources de données cartographiques pour leur intégration à une base de données à référence spatiale. Mémoire présenté pour l’obtention du grade de maître ès sciences (M. Sc.) Département des Sciences Géomatiques, Université Laval


Keighan E. (1999), CubeSTOR for Oracle, CubeWerx, 1999


Interactive, User-Oriented, Error Modelling Designed to Support Rural Spatial Analysis and Information Production

David Fraser
Department of Land Information, RMIT University
GPO Box 2476V, Melbourne, Australia,
david.fraser@rmit.edu.au

Abstract
In an attempt to understand our physical environment, models have been developed. The need to understand our physical world leads to the need to bring order to physical systems so that logic can be applied and analysis undertaken. This is done to determine whether a particular outcome is possible, to determine the interactions between particular features over space and time, or to map the study area so that other analysts can extract information from the products created. The success or otherwise of spatial modelling hinges on the ability of the analysts to identify key elements of the physical environment for the spatial modelling exercise.

Small changes in the identification of feature, the collection and processing of data and the display of information can result in quite different outcomes. The more specific the modelling the greater is the expectation relating to the data and the data modelling. General models have more latitude regarding accuracy and precision. Errors and uncertainties will interfere with the purity of the information obtained from the model and may lead to misleading interpretation by the user of the information produced.

The uncertainty, errors, accuracy and precision from all of the models that are used in the processing must be assessed and the validity of the data questioned before the next stage of a project is undertaken.

The onus is on the modeller to create a digital environment, which involves the use of generalisation rules to limit the use of the data and to limit the creation of information products by taking into account the geometric and attribute accuracy of the generic data. To think otherwise is to suggest that users, with little or no cartographic background, are able to impose such controls themselves. More and more spatial information systems now include controls to limit the use the data can be put to. It is the purpose of this paper to identify new ways of implementing these controls. A new conceptual model will be presented which demonstrates how the uncertainty, errors and imprecision from different data sources can be taken into account before the final spatial visualisations are produced.

Introduction
In an attempt to understand our physical environment spatial models have been developed. The key to successful spatial modelling is the identification of the most important characteristics of physical features under study and the recording of them digitally. The development and application of appropriate algorithms, which take into account the uncertainty, errors, inaccuracies and imprecision, which are inevitable parts of the modelling, is also a requirement. Often the surface of the earth, and the features upon it, are replaced by simple mathematical equations that aim to mimic a physical process or physical environment.
Small changes in the feature identification, data collection, data processing and the display of information can result in quite different outcomes. The more specific the modelling the greater the expectations relating to the data and the data modelling. General models can have more latitude in accuracy and precision. With all spatial models, errors and uncertainties will interfere with the purity of the information obtained from the model and may lead to misleading interpretation by the user of the information produced.

It is necessary to clearly identify the limits of the modelling and work within these limits. The initial conditions such as the conceptualisation, data collection method and the application of the logic associated with the modelling, will influence the outcome from the modelling. It is important to be able to identify outcomes from the modelling which can be related back to the physical environment and those outcomes that are produced by the system due to such factors as data generalisation or data error.

The analyst must decide what to select from the identifiable features so that a model of a specific aspect of the environment can mimic at least part of the physical world. This process is full of decision making which can lead to different outcomes depending on many factors such as:

1. What features can be identified?
2. How clearly can a feature be distinguished from other features?
3. To what level of positional accuracy can a feature be recorded?
4. How concisely can the features be defined?
5. What errors are associated with feature extraction?
6. What data is not recorded?
7. What information is lost in the modelling process?
8. What constraint must be placed on the use of the model output?
9. What assumptions have been made throughout the process?
10. How will the user interpret the output?
11. What other information may be obtained through the process of induction?

**Modelling**

The physical environment is constantly changing. The spatial scale considered for modelling and the date the data is collected for the modelling will influence the outcome from the model. What may be applicable to one study may not be applicable to another study. Errors introduced early in the process will be magnified as other errors, of both a random and systematic nature, occur at specific points in the process. The errors may occur because of different data collection and representation, and through inappropriate or inadequate analytical modelling. If a modeller is uncertain about the requirements of a modelling process, or does not take note of the uncertainties right from the start then the modelling is likely to be of minimal use even though the output may look sophisticated.

The more specialised a project the less likely it is that the results of the project will be applicable in other areas. If the study were reliant on precise measurements under controlled conditions then the same rigid parameters would need to be adopted in other study areas if comparisons were to be made.

The order we attempt to bring into our spatial modelling comes at a price. There is the need to confine, stratify, approximate, generalise, separate, segregate, categorise and simplify elements selected from the environment, so that they conform to a new order that allows processing in a digital environment. This foreign environment is contrary to the physical world where ‘Nature paints its scenes without regard for conventional order, for
straight lines or Euclidean shapes. .....The essence of the earth’s beauty lies in disorder, a peculiarly patterned disorder,..........’ [Russek J. (Ed.), 1991, p11]

To select specific elements from the physical environment and to disregard other elements may break important links that exist in the natural system. For example, to consider a river as a closed system is to ignore the external interactions which continually alter the course and composition of the river.

In an attempt to overcome some of the element selection problems, research and development should involve many professionals from different backgrounds. In this way the individual specialist may identify those factors critical to the success of a modelling exercise. For instance it is unlikely that one individual is knowledgeable in mathematics, statistics, spatial analysis, soils, vegetation, hydrology, geology, mapping, surveying and computer science. Yet it is quite possible that knowledge in all of these areas is required for the successful implementation of a model. ‘The specialist concentrates on detail and disregards the wider structure which gives it context.’ [Laszlo, 1972, p13] An overall methodology needs to be established so that the input from the specialists is integrated successfully into the processing.

It is the responsibility of the modeller to provide an approximation of the real world situation that is as close as practical to the situation on the ground. The success of the model in part relies on the knowledge of the model creator. According to Nash “our ‘knowing’ always affects our ‘seeing’. We learn not only how to see but also what to see; often we simply don’t see anything for which our learning has not, in some sense, prepared us.” [Nash, 1963, p9]

The outcomes from a modelling project are influenced by the processing stages, from the genesis of the model through to the final interpretation of the information products. William James writes: “What we say about reality thus depends on the perspective into which we throw it .......... By our inclusions and omissions we trace the field’s extent; by our emphasis we mark its foreground and its background; by our order we read it in this direction or in that. We receive in short the block of marble, but we carve the statue ourselves.” [Nash, 1963, p3] The end user will assimilate their interpretation of the modelling results with other knowledge they have gained over time. The intellect, upbringing, ethnicity, biases, education, professional background, environment may all play a part in determining what the user considers to be the important information derived from the model. This component of the modelling process is something the modeller has little control over, as Einstein observes: ‘Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world.’ [Nash, 1963, p47]

The lack of knowledge relating to our environment, the uncertainty associated with identifying unique features in the environment, the errors associated with data collection, the levels of accuracy and precision associated with data collection and storage, the assumptions made in current modelling all lead to the fact that we are still far short of fulfilling the kernel of Albert Einstein’s belief ‘in the possibility of producing a model of reality - that is to say, a theory which will represent things themselves.’ [quoted in Nash, 1963, preface]

If it is true that “...the vast majority of GIS users... have little or no cartographic background.” [BCS, 1995, p5] then the onus is on the modeller to produce information products which are unambiguous and fit for purpose. The onus is also on the modeller to create a digital environment, which involves the use of generalisation rules to limit the use of the data and to limit the creation of information products by taking into account the geometric and attribute accuracy of the generic data. To think otherwise is to suggest that users, with little or no cartographic background, are able to impose such controls themselves. More and more spatial information systems now include controls to limit the use the data can be put to.


Facts

Collecting digital spatial data, like collecting other facts about the physical environment, provides endless possibilities. If a feature can be culled from the environment then this feature will have many characteristics that need to be put through a sieve so that only the characteristics of most importance are retained. On the other hand Goldstein [Goldstein et al, 1978, p20] argues that “if one objects that most of the details are irrelevant, the answer must be, “How do we know?” Based on our accumulated knowledge and past experiences we must determine what is relevant.

A difficulty faced in spatial data collection and analysis is that different disciplines will consider different data to be important or central to the problem at hand. A soil scientist may consider soil to be central in a discussion on soil salinity. A hydrologist may consider water flow to be the critical component while a geologist may consider the underlying geology to be the central focus of the analysis. The uncertainty in the data begins at this early level of conceptualisation and continues on through the definition of data boundaries into the final representation as a map. The observer, or groups of observers, influence the outcome of any spatial data analysis in a major way. ‘Facts are not really independent of the observer and his theories and preconceptions. However, at any one time, in any one culture, it is usually possible for most observers to agree on them. To put it better, facts are what all observers agree on.’ [Goldstein et al, 1978, p21]

Developments in Spatial Modelling

The future holds many challenges for the professionals involved in spatial data analysis. Those analysing spatial data do so within an extremely dynamic environment. New technologies are not only improving the way we process and display spatial information they are changing the way we organise the projects themselves.

Research and development in the area of measuring and monitoring aspects of data quality is leading to more awareness of the limitations which must be placed on data use because of the uncertainties, inaccuracies and errors. The trend is to improve the data quality so that more reliability can be attached to the results of any analysis. In the beginning digital data mainly consisted of primitive files of numerical data which in the spatial domain did not have associated topology and was used to mimic the analogue map. “Very few people questioned the data being put on the screen. If the display looked impressive, then most users were prepared to accept that it was 100 percent accurate. This is particularly worrying when the data is output from a model which is, by definition, an approximation of the system.” [Watson et al, 1996, p431] On the other hand “while accuracy in measurement is a desirable goal, one should not conclude that the main concern of science is to measure all things with greater and greater accuracy. One important component of scientific judgment is to know when additional accuracy is worth the additional trouble it takes to get it.” [Goldstein et al, 1978, p234]

Data Integration

Many applications of digital data require the combining of data from multiple sources. The source of the data needs to be established and the integrity of the data needs to be determined. Along with this the properties of the data may show the datasets to be incompatible. The need to integrate data from multiple sources has seen some users, through ignorance, combining data in an inappropriate manner such as when digital data compiled at a scale of 1:25000 is combined with data from a 1:100000 scale, with the output being presented on a 1:25000 scale analogue map.
To overcome some of the problems faced in data integration different levels of generalisation have been built into the more sophisticated spatial analysis systems. Such systems allow the display of certain digital data only when specific levels of generalisation are reached. For instance the individual allotments in an urban area may not appear on a display until a zoom factor equivalent to 1:2500 is reached. Up to this point only the road network may appear.

**Other considerations**

The sophistication of current digital spatial information systems means that users of these systems can display and manipulate the digital data through a graphical user interface designed to be intuitive and to allow access to the system in a no fuss manner. There is then the tendency to believe that because such systems are simple to use and because digital data is available for a particular study area, it can be used for whatever purpose the user has in mind. The ability for these systems to produce excellent looking map products may be used to hide the user’s ignorance regarding the data and hence the uncertainty associated with the output from such systems. Because of the multi-user environment for many datasets it is necessary for the software to have the controls built in rather than leaving it to the user to make the decisions. An override of the software can be provided to allow a user with the required understanding and a specific application to intervene and display a non-standard visualisation.

Probabilistic reasoning is becoming more important. Instead of rigid boundaries being formed between different classes of the same feature (e.g., vegetation cover), the probability that a geographical area is in one class or another class is stated and this allows the fuzziness associated with the physical environment to be introduced into the database. The introduction of knowledge based systems and the rule-based query of spatial databases has provided analysts with the ability to undertake deductive query processing. Self-interrogation software for spatial databases provides the database manager with tools to detect and correct for topological inconsistencies in the data.

As computer systems become more and more sophisticated and research takes database design and management into new areas, there is the opportunity for many of the database management tasks to be automated. In a similar way that a spelling and grammar checker is used in a wordprocessor, data interrogation software is being used on spatial databases. The aim being to identify duplications, identify problems with the topology, or determine the suitability of data for specific purposes or for addition to an existing database. The data interrogation software uses information about the data, stored with the data.

**Rural Spatial Analysis**

Spatial analysis in rural areas ranges from specific identification of land use/land cover characteristics on a paddock by paddock basis through to generalised classification based on the most dominant land use/land cover for a region. Analysis within parcels (precision farming) is also possible. The more general the classification, for instance at a regional scale, the less opportunity there is for detailed analysis.

The first stages of a mapping project relate to the sourcing of digital spatial data and the determination of the sources of expert knowledge related to a specific geographical system. The geographical study area then becomes the focus. Knowledge relating to the site is acquired and this forms the knowledge domain. The problem is rigorously defined. Data is selected and analysed according to a predetermined methodology and information products are produced. The interpretation of the information adds to the knowledge relating to the specific site and topic area (Figure 1).
Agricultural modelling systems must be capable of operating at different levels, such as global, regional and local. The results of analysis at the global level may not be directly applicable at the local level due to generalisations that are inherent in the data.

**Data modelling**

According to Dutton [Dutton et al, 1985] a good model deals with the locational, temporal and thematic aspects of a phenomenon. The modeller should be aware of, and should document, the assumptions and limitations built into the model. The modeller should also be aware of how the user of the model will interpret the output from the model. Not all users will interpret the results the same way and the breadth of interpretation should be understood. In this way the model can be changed or instructions provided to guide the end user. “No model is better than the assumptions and data it relies on. In circumstances where the assumptions are valid and/or forecasts are being made within the range of data used to establish the models, then it will probably give reasonable results. Where one is forecasting outside this area of tested reliability, no model should be relied on without further testing.” [Kirkby et al, 1993, p4]

The ideal is to produce a database that can satisfy the needs of all users over space and time. This, of course, is an impossibility so a database creator should:

1. attempt to maximise the integrity of the data,
2. build topology into the data,
3. provide for error traps,
4. alert users to misuse of the data through ‘hints’,
5. automate much of the map generalisation tasks,
6. establish sound geometric and attribute links between data sets,
7. provide for multiple representations from the one database system,
8. reduce the need for user intervention,
9. provide a graphical user interface that is logical and consistent,
10. automate the information retrieval as much as possible,
11. provide a means of data certainty, error, accuracy and precision checking,
12. build a database around sound spatial reasoning,
13. build a database around sound temporal reasoning,
14. build a database around sound aspatial reasoning,

We store “approximate, error prone, characterisations of aspects of the real world” [Jones C.B et al, 1996, p904]. Hence it can be said that spatial relationships are derived from error prone data.

**Error, accuracy, precision and uncertainty**

Errors and uncertainty will always be present in digital spatial databases. The black box nature of many algorithms used in an information system may see the initial errors and uncertainties increasing in size as the processing continues. Software tools for measuring and assessing the significance of error are required.

There are a number of possible ways of controlling the development of errors and uncertainties or at least identifying what they are. Some possibilities are listed below:

1. Specify the level of decision making for which an output product may be used.
2. Undertake empirical research into the databases errors and uncertainties.
3. Develop a set of generic tests that can be performed on the data.
4. Test the database against a database of higher accuracy.
5. Build error-testing algorithms into computer software.
6. Provide information regarding the error propagation associated with processing through specific algorithms.
7. Provide information regarding the origin and properties of all data in the database.
8. Have available a digital orthophoto so that vector data can be overlaid on it allowing a visual check of the topological and generalisation characteristics of the data.

Multiple input data and multiple output forms after modelling can provide a means of determining the uncertainty associated with a model. The level of uncertainty needs to be quantified on some standard scale so that the user is aware of the appropriateness for use of the particular data and processing method.
Interactive, User-oriented Error Modelling

In an attempt to address the lack of tools for measuring and assessing the significance of errors in spatial databases, the author has developed error-modelling software. The software deals with many of the issues identified in this paper and it acknowledges that digital spatial data may come from a variety of different sources and may vary significantly in quality.

The logic is based on the method used for land capability assessment. In land capability assessment the suitability of the land for a particular land use is determined by comparing key parameters of a study area against a land use capability table. For instance, a specific geographical area will be partitioned into slope zones. Within each zone data will be collected about the soil, rock, flood potential, etc. This data will then be compared to set tables, which define the limiting physical characteristics for specific land use types. The land can then be rated on a scale of 1 (suitable) to 5 (unsuitable) for a particular land use based on the comparison of the land characteristics against the capability table.

Figure 2 provides a synopsis of the adaptation of the land capability approach to spatial data error assessment. The software has been developed using Visual Basic and provides a user-friendly interface to the data and its quality indicators.

The user provides the central focus for the software. If the user, for instance the landowner, is very familiar with the issue but is not a familiar with digital spatial databases, then the system takes the user through a plain English, question and answer session. This is designed to identify the key aspects of the issue under investigation and to determine the desired information required by the user. At the end of this session the system accesses the information relating to the quality of the data (metadata) along with the “capability rating table” to determine the information products available for the user. This determination is based on the uncertainties, error levels, accuracy and precision details found in the metadata. In this way the user is restricted by the system to products which will provide information, not misinformation, at the desired scale of the study.

The metadata related to the database will vary significantly from dataset to dataset. Data may come from a variety of sources and hence may vary in quality. By interrogating the metadata the system can determine what data can be combined to create the particular information products and what data must not be combined.
The capability-rating table rates the spatial data according to the sophistication of the data collection method and also the certainty in terms of the feature identification. For example, ground survey data would be rated high on the data collection rating. If the ground survey were undertaken to determine vegetation cover then it would have a low certainty grade, as vegetation boundaries are, in general, difficult to determine definitively. The technology focus provides a path which allows the users who have a sophisticated understanding of digital spatial data to interrogate the data themselves. Users can view the metadata and the capability-rating table and make their own decisions in relation to the information products which may be created. This path provides the user with more latitude in the use of the database but also assumes that the user has the knowledge required for exploiting the database wisely.

**Concluding Comments**

“The initial impetus to use the new technology to automate map production still remains an important thrust but at the same time cartography clearly has a role to play with the emergence of GIS analytical tools and the integration of spatial and non-spatial information.” [BCS, 1995, p3] The production of maps as output from spatial information systems still remains one of the key forms of presentation. As the sophistication of spatial information systems grows so does the sophistication of the maps created. The cartographic product is now only one part of what is termed visualisation of the data. Such visualisation opens up new ways of displaying the results of an analysis. Visualisation of the results of an analysis of spatial data will create a distortion of the real world. How much distortion can be tolerated depends on the use that exists for the output. Interactive, user-oriented error modelling, as outlined in this paper, has the potential to be a valuable tool for the provision of improved spatial information products.

**References:**


Goldstein M. & Goldstein I.F. (1978) How We Know- An Exploration of the Scientific Process, Plenum Press,


Session / Séance 07-A

Active Object Techniques for Production of Multiple Map and Geodata Products from a Spatial Database

P.G. Hardy
Laser-Scan Ltd,
Science Park, Milton Road, Cambridge, CB4 4FY, UK.
Phone +44 1223 420414, Fax +44 1223 420044
Email: Paul @ lsl.co.uk

Abstract

Historically, map production has been an isolated task, compiling information from various sources to produce a particular cartographic product. This traditional view is now challenged by the availability of new environments which bring together databasing, object data modelling, image analysis, active agent behaviour and automated cartography to produce a unified flowline for production of multiple maps, charts, geospatial data, and on-demand spatial visualisations such as Internet web mapping.

This paper overviews a modern production application built on an object-oriented geospatial database, and specifically highlights its capabilities for active representation, multiple geometry and active object selection, in order to explore the benefits and future directions of active object mapping. Recent advances in handling spatial objects allow for the first time the efficient storing and retrieval of a geographical feature as an object with a single set of attributes, but with multiple alternative sets of spatial information (geometries). This allows one continuous spatial dataset to hold data appropriate to multiple map and geodata products, reducing the overheads of update as the outside world changes.

Complementing the multiple geometry capability are new mechanisms for selection of objects from the database for product generation. Object database views are implemented as object behaviours, and are the means by which geographical feature objects can decide for themselves whether they should be included in a specific map or geodata product. Once selected for inclusion, further object behaviour techniques come into play. Active generalisation methods rely on message passing to objects to ask them to simplify or displace themselves. Dynamic representation implemented as active object display behaviours allows individual objects to draw themselves differently according to the surroundings.

Distributing the knowledge of selection, generalisation and representation into object behaviours in this way overcomes many of the problems previously encountered in embedding the skill of the human cartographer into a software solution. For the first time the cartographer’s dream of the scale-free map can become reality, not only for generation of a range of paper-based mapping products, but also for on-demand visualisation in new media.
1. INTRODUCTION

1.1 Why an object database for mapping?

A map is a model of part of the surface of the Earth, presented conventionally as a graphical illustration. Maps for different purposes will tend to exaggerate relevant features while minimising or suppressing irrelevant detail. The term ‘map’ is used in this paper to cover the whole range of mapping products such as topographic maps, thematic maps, charts, plans, atlases and geodata (e.g. CD-ROMs). Producing mapping products used to be a manual draughting task, but now relies on computer cartography.

Understandably, the first stages of the evolution of digital mapping mimicked the conventional production process, capturing and compiling the data needed to produce a particular map or chart, usually using file-based feature mapping or graphics software. Increasingly though, the wasteful nature of such one-off capture has been recognised, and there is a move to a database-centric approach in which a geospatial model of the world is captured, stored, and updated [Cameron & Hardy 1998]. Starting from the database, one can produce a range of products at differing scales and to different specifications, as described later in this paper.

Traditional relational databases are not designed for holding the complex data models and large volumes of variable length data involved in building and ensuring the ongoing integrity of a real-world geographic mapping database. Neither is it easy to produce a range of cartographic products from such data using the static representation facilities found in traditional GIS and mapping software. Now, Object-Oriented (O-O) geospatial databases and associated mapping products have appeared [Warboys et al 1990], which provide the technology for a new world of active objects and product-independent geodata storage. The later sections of this paper cover the O-O paradigm, and put forward its strengths for geographic databasing and map production. They use as an exemplar, the Gothic O-O database and LAMPS2 mapping system from Laser-Scan [Laser-Scan 1994], shown schematically in the product family diagram below right.

2. OBJECT DATA MODELLING

2.1 Object-Orientation and object data model

In an O-O database, real world entities are abstracted and held as objects. All objects belong to object classes. For each class there may be many objects, but each object belongs to only one class. The class defines what values can be held by an object. Values can be simple datatypes (integers, strings, dates, etc.) together with more specialist types (geometries, locations, rasters, and tables). Furthermore, objects can hold structural information or references between objects.

A key, and defining, concept of O-O is that of methods defined on objects. These methods are bound to behaviours. When a method on an object is invoked by sending a message to the object, the behaviour bound to it is executed, possibly using values and references held by the object. The ability to define behaviours as part of the database schema, rather than as part of the application, is a fundamental concept of the O-O paradigm.

A further key concept of O-O is that of inheritance, which provides the means to define a new object class in terms of existing classes. The new class inherits the characteristics (values, references, behaviour methods) of its parent class or classes, unless superseded or redefined. Using inheritance, hierarchies of classes can be created and easily maintained.
True O-O has gained much popularity in software engineering and computer graphics [Taylor 1990], and is appearing in GIS, cartography and geodata production. In reality, however, there are still few commercially available systems that support all the key elements to a level that can successfully support mapping, charting and geodata production applications.

### 2.2 Methods and behaviours

Methods are central to the O-O technology. Each object class will have inherited basic methods from its parent classes, and can have other methods and specific behaviours for standard methods defined on itself. Methods are of several types:

- **Value methods** return an answer to a message. The results appear as attributes on enquiry, e.g. area, length, description.
- **Reflex methods** occur automatically at milestones in an object’s lifecycle: creation, modification or deletion (before and after). They are used to set up consequences of actions.
- **Validation reflex methods** enforce integrity, and allow you to put your own rules on each object class (see 2.3 below).
- **Change to a referenced object** is a reflex method, which can trigger propagation of effects from one object to another.
- **Display methods** give active representation (see 2.4 and 3 below)
- **Process methods** happen at operator request. They are used to carry out data cleaning, data checking, polygon formation, and generalisation on defined sets of objects.

![Diagram of multiple inheritance of classes](image-url)

**Figure 1.** Multiple inheritance of classes
2.3 Validation methods for data integrity

Data integrity is a major issue for agencies who invest large amounts of money in capturing and maintaining a large geospatial database. The object data model allows the agency to define its geodata logic and business rules as reflex methods in the database schema. This means that the database will enforce these rules as the objects are entered into the dataset, and again whenever they are modified.

Whenever an object is being modified, messages are sent automatically to the object at the various stages:

- One before the modification is started, so that it can check that it is allowable (e.g. can’t move a lighthouse unless you are a supervisor).
- One after the modification is finished, so that it can check that it has been done validly (e.g. can’t edit a contour so that it crosses another contour).

If any of the validation reflex methods return a “not OK” result, then the complete transaction is rolled back as if it hadn’t started. Note that such validation methods in the database are not just applied during interactive edit, but also during other operations such as bulk data loading from external data sources (e.g. legacy data).
2.4 Display methods and active representation

In an O-O mapping system, the appearance of an object on the screen or on hardcopy is generated at draw time by execution of an arbitrary ‘display method’. Such methods are defined on the object class and stored in the database under the direct control of the customer. This contrasts with the traditional approach as indicated in the following table.

<table>
<thead>
<tr>
<th>Dynamic O-O Active Representation</th>
<th>Traditional static feature-based representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Objects can draw themselves differently each time, adapting to external influences (e.g. map scale, product)</td>
<td>1. Single unchanging appearance, requiring separate datasets and representations to achieve multiple products</td>
</tr>
<tr>
<td>• Behaviour defined in database by customer</td>
<td>2. Behaviour defined in application by supplier</td>
</tr>
<tr>
<td>• Can be influenced by combinations of attributes including attributes derived from other referenced objects</td>
<td>3. Influenced only by single feature code attribute, and each feature is represented in isolation</td>
</tr>
</tbody>
</table>

Figure 3. Topographic map active representation
Base map data courtesy of DOSLI South Africa. Copyright RSA.
The functionality and benefits of active object representation [Hardy & Woodsford 1997] are applicable to a wide range of cartographic products, from topographic mapping to nautical charting, as shown above.

### 2.5 Object Versioning and Long Transactions

One of the problems of traditional relational databases is that their transaction model is designed round the rapid lock-update-unlock scenario common in financial and business transactions. However, completing the update of geodata and mapping for an area is often a long drawn-out process, taking several hours, days, or even weeks [Hardy 1995]. In the interim, the half-updated state must not be allowed to be used for production tasks, but also the unchanged data must not remain locked.

The object database, with its encapsulation of all the data and behaviour for each object, lends itself to a different transaction model. In this, each user has a stable view of a ‘version’ of the dataset. Only changes made by that user are stored in the version, the unchanged objects are accessed from the previous version. Versioning of datasets solves the problem of long transactions and allows the sharing of very large data volumes between multiple users needing write access [Woodsford and Hardy 1997].
Figure 5. Dataset versions
Figure 6. Long transactions

In the diagram above, two users each reserve a segment of a continuous dataset for update. After several checkpoints (e.g. stops for lunch), the changes are merged to give an updated mainstream version. This is discussed further in previous papers on spatial databasing [Woodsford 1996].

Within mapping, charting and geodata production flowlines, versioning allows efficient multi-user update access to a continuous dataset of master features. In addition, it allows product-specific versions to be subset from the continuum, e.g. to include all the sheet furniture (grids, borders, titles, legends, North arrow, etc.), without having to copy all the unchanged geodata.

2.6 Topology and Structure

The Gothic object spatial database uses methods and object references to implement in-built support for topology structure. The user can choose Spaghetti or Structured for each class, and then define snapping tolerances between pairs of classes. The database will then apply these rules and create the necessary links and nodes as objects are digitised or imported.
In addition to line topology, which is made up of links, areas can be topologically structured, referencing links as their boundaries and one or more faces as their interior. Uniquely, the topology is maintained automatically as the data is edited, avoiding risks of overshoots, undershoots, slivers etc., and obviating the need for subsequent error-prone building of coverages. Polygons can be formed out of existing linework, either directly or as a set of faces (the atomic entities of area).
4. OBJECT DATABASE VIEWS

4.1 Why do we need object database views?

The new database-centric approach to map production means that a single continuous dataset that models the real world is created and maintained. The contents of this dataset (sometimes called the Master Features Dataset or MFD) have to be the source for all the products to be generated, which may range from topographic mapping, through navigation atlases to thematic maps.

Any one of these products will depict a subset of the MFD - topographic maps may not show postcode boundaries; navigation atlases may not show contours, and thematic maps may not show roads.

In practice, selection criteria are much wider ranging than just by feature class. Topographic maps may need to show contours only if their heights are a multiple of 100m, or may show urban areas if their population is greater than a minimum value. Atlases may show villages only if they lie on a through route. A thematic map may only show points of interest if they lie within 2km of a trunk road. The efficient and flexible mechanism for implementing all these types of selection is the object database view.

4.2 What are object database views?

A database view is a way of hiding some of the information in a database, and presenting a simplified subset to the user. Relational database views are traditionally read-only, and can only subset by hiding specific columns of tables, or hiding rows by value range of specific fields. In contrast, object database views are read/write, and the selection depends on the result of any true/false value method defined on the object class. Hence, whether a specific object is in the view or not is determined when the method behaviour is invoked at draw time (or whenever). View methods are defined in the database schema (or can be automatically created based on query forms).

The view method can check multiple attribute values, can follow pointer references to related objects to retrieve information, and can use the power of the spatial index and spatial toolkit to calculate whether this object should appear in this view or not.

Because views are defined and held on the database, rather than in the application, they are applicable to a variety of kinds of database access, whether for screen display, object search, hardcopy plotting, or export to external product formats. The richness of the object database views capability allows the embodiment of a cartographic or geodata product specification as a saved specification, which can then be used to produce successive versions of products.

5. MULTIPLE GEOMETRY OBJECTS

5.1 Why multiple geometries?

The object database holds a model of the real-world state of the geographic features it includes. However in any particular map product, it may be necessary for cartographic reasons to show the feature in a different position, or in a modified shape.

One way to handle this might be to have separate product-specific datasets, one for each product, containing the modified features. However this would be difficult to maintain and would necessitate much labour to keep the master and derivatives in step in the light of change in the real world.
5.2 What are multiple geometry objects?

A better solution is made possible by encapsulation, which is one of the basic tenets of object-orientation. In the object database system, access to all properties of the object can only be via methods defined on the object class. This is true not just for simple values such as attributes, but also for more complex values such as the point, line, or area geometries which contain the defining coordinates for the features.

Hence, by overriding the two methods ‘set-geometry’ and ‘get-geometry’, it has proved easy to implement in Gothic an efficient mechanism for storing multiple alternative geometries on the object, only one of which is active at a particular time. This mechanism is transparent to the application and allows normal editing commands to be used for setting the alternative geometries, and normal display and hardcopy facilities to be used for output. The facility for multiple alternative geometries on objects in LAMPS2 allows a single dataset to hold not just the master geometry (the real-world position), but alternatives suitable for a range of derived products. At the same time, only a single object exists with a single set of attributes, so space is conserved and the complexities of parallel update are not needed to control multiple objects.

5.3 How are multiple geometry objects used?

This facility can be used for the many situations in cartography where alternatives are needed. In particular it has shown worth in:

- Sheet dependent information, where the constraints of the sheet edge necessitate cartographic amendments to the map data.
- Scale dependent ‘patches’ of modified data, stored in the main dataset, but replacing the true-to-life basic scale data in cases where cartographic generalisation for scale has forced deviation from reality because of constraints of clarity.

The left illustration below shows the master geometry for all objects, while the right illustration shows the appearance when a product alternative has been selected. The purpose of the alternative is to produce an atlas page covering the top right quadrant of the screen. Note the movement of some features (e.g. “Stoughton”), and the change in shape of others (e.g. the road near “Funtington”). All these changes are to alternative object geometry, so attributes (such as the village name “Stoughton” are common and not stored twice.
Figure 9. Display for map sheet using Master geometry
   (Base data courtesy of Automobile Association, UK)
6. GENERALISATION

6.1 Map Generalisation

Map generalisation is the science (and art) of exaggerating those aspects that are important for this particular map purpose and scale, and removing irrelevant detail that would clutter the map and confuse the user. The following example shows the same area at three different scales, showing different levels of detail, and different objects for the same entity (e.g. area at detailed scale goes to symbol at small scale).

Figure 10. Display for atlas page using Alternative geometry
(Base data courtesy of Automobile Association, UK)
Generalisation has traditionally been a hard task to automate, being dependent on the skills of the human cartographer. People have tried for years to build centralised ‘knowledge bases’ of generalisation rules, with very limited success. In such systems, the map features themselves have just been passive items containing coordinates and attributes, acted upon by the centralised rules [McMaster, 1991].

In the object-oriented world, this is turned upside down. The map features themselves become objects that have generalisation behaviours defined in the database schema. The application itself becomes much thinner, and contains no knowledge about what, how, or when. It merely provides a framework for invoking and sequencing the generalisation processes by sending messages to selected objects.

Each such object inherits behaviours from its object class definition and from superclasses in its class inheritance hierarchy. These behaviours allow the object to decide for itself what to do when receiving a message to generalise itself, e.g. it can inspect its relationships with its neighbours to decide whether to move itself. However, as the object modifies itself, any objects that are directly linked to it or spatially adjacent can also be told to reassess themselves, so that effects propagate.

One of the fundamental tenets of O-O is polymorphism, in that different object classes may respond to the same message by different method behaviours. For generalisation, this has particular strengths in that the ‘simplify outline’ generalisation method may have very different behaviours defined for man-made objects like buildings, to natural objects like lakes, even though they are both area objects [Ormsby and Mackaness 1999].

The advent of the object-oriented paradigm therefore opens up new strategies for generalisation [Buttenfield 1995]. These apply particularly to single datasets used for multiple products, but also for maintaining a series of related but distinct datasets [Kilpeläinen 1997], [Harrie 1998].

LAMPS2 includes an object-oriented generalisation facility, which allows the user to define the strategy for generalisation in terms of methods on the object classes [Hardy 1996]. Generalisation base classes are provided which supply generalisation process methods for multi-object combinational operations (aggregation, typification, displacement) and others for single object generalisation (collapsing, refinement, exaggeration and simplification). Note that these are implemented as behaviours of the objects in the database, not as commands within a program.

Preparation for generalisation using these methods is aided by a visual interface to setting controlling parameters (see figure below left). This allows the map designer to see in real-time the effects of tuning parameters. Once set up, then a process sequencing mechanism allows unattended execution of complex generalisation runs on multiple object classes for chosen areas of the continuum to produce generalised products.
Figure 12. Visual interface to generalisation parameters

Figure 13. Geographic objects become co-operating agents
6.2 Multi-Agent Generalisation

Further dramatic developments of the LAMPS2 generalisation facilities are under way, driven by the AGENT project on multi-agent generalisation. This project [AGENT 1997] is a collaboration under the ESPRIT programme (LTR/24939) involving Laser-Scan as providers of object technology together with a national mapping agency (IGN) as prime contractor, and academic partners (Edinburgh & Zurich, INPG). Some partners provide in-depth knowledge of generalisation algorithms, while others provide insight into multi-agent modelling. The contract involves 48 person years of effort over a 3-year period.

In this context, agents are self-aware active software objects that co-operate, subject to a set of constraints, to achieve a goal. For map generalisation, it is the geographic objects such as houses and roads which become active agents and co-operate through simplification, typification and displacement of themselves to achieve a cartographically acceptable generalised result [Baejis et al 1996]. The illustration above right shows co-operating meso-agents handling urban blocks, communicating with the micro-agents that are the buildings and roads.

One outcome of the AGENT project is to be a set of base classes in LAMPS2, the methods of which embody the measures, constraints, algorithms and goals of map generalisation for specific feature classes. Future papers will expand on this new approach to automated generalisation.

To allow true on-demand mapping in response to user requests over the Internet, then the decisions that traditionally have been made by the human cartographer will in future need to be made by the automated geodata web server. This automation has proven insurmountable for traditional feature-based digital mapping systems. However, the active object generalisation paradigm has shown the way forward and Laser-Scan and the AGENT project are working towards the goal of on-demand adaptive Internet mapping through O-O and multi-agent generalisation.

7. MAPPING ON THE INTERNET

7.1 Web client-Server

The Web typically relies on a multi-tier architecture. The most common formulation consists of three tiers:
1. Data storage
2. Business logic
3. Presentation

The Internet is used to remote the ‘presentation’ tier from the data and business logic. Furthermore, it is common to expect the presentation tier to be thin, both in its footprint and in its bandwidth requirements in communicating with the business logic tier. The ‘business’ behind most successful geospatial applications, Internet based or otherwise, is a sound and extensible model of the real world. Object-orientation provides the most successful software modelling tool to date. Thus, one can regard an object-oriented database as a natural fit for the data and business logic tiers in a Web-enabled geospatial solution.

To keep the presentation tier thin means that higher level abstractions are required for communication over the Internet. The most common abstraction with geospatial databases is the feature. However in many cases it is necessary to communicate a ‘map’ to the presentation tier, and this is ideal work for the O-O selection, generalisation and representation mechanisms described previously. An object-oriented database therefore is not only capable of answering web queries about the properties of individual features, but can also serve up fully-symbolised maps.

Java applets using a feature communication protocol provide one important way of delivering an easy-to-use presentation tier. A great deal of user-interaction can be handled locally, from simple tasks like drawing out a
zoom-box to manipulating individual vertices in a feature geometry. For more analytically rich and cartographically complete situations, it is sometimes better to let the object database take the responsibility for validation and preparation for presentation, and let the O-O server supply a symbolised map as an image to a very thin client. Both types of client are not mutually exclusive, and it is common to server the same object database in both vector and image modes to clients, depending on their capabilities and needs.

7.2 Web presentation of feature attributes and context

Object-oriented techniques also facilitate web presentation of attributes and context. Benefits come because the client does not even need to know what type of feature is being manipulated. All features in an object-oriented database might support a common method that returns an HTML page describing the feature. Furthermore the ability of an object-oriented database to deal with a variety of data-types means that it can be readily customised to return HTML that is generated as required. The illustrations below show web-based access to a Gothic object dataset, and a generated web page showing derived properties of a particular feature.

Figure 14. Web applet displaying Gothic object data (Data Crown Copyright)
Postal Sector: CB 1 5

Area analysis

- **Area**: 56.0 sq kms
- **Population**: 4104 [0.07 thousands / sq km]
- **% Urban**: 5

Car Ownership analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households with no cars</td>
<td>269</td>
<td>17</td>
</tr>
<tr>
<td>Households with 1 car</td>
<td>715</td>
<td>45</td>
</tr>
<tr>
<td>Households with 2 cars</td>
<td>497</td>
<td>31</td>
</tr>
<tr>
<td>Households with 3 or more cars</td>
<td>121</td>
<td>7</td>
</tr>
</tbody>
</table>

Dwelling Type analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flats</td>
<td>89</td>
<td>5</td>
</tr>
<tr>
<td>Terraced Houses</td>
<td>312</td>
<td>19</td>
</tr>
<tr>
<td>Semi-detached Houses</td>
<td>597</td>
<td>36</td>
</tr>
<tr>
<td>Detached Houses</td>
<td>678</td>
<td>40</td>
</tr>
</tbody>
</table>

Radio Station analysis

- No data
- BBC Radio Cambridgeshire
- Chiltern Radio Eastern Network Total
- Q103

Based upon the 1997 Ordnance Survey Meridian Map with the permission of The Controller of Her Majesty's Stationary Office
(c) Crown Copyright

Laser-Scan

**Figure 15.** Derived attributes as a web page. Generated as HTML on-demand

The O-O paradigm is increasingly fundamental to the evolution of the Internet, as witness the dominant role of Java and CORBA.

A recent workshop was held at the AGI (Association for Geographic Information in UK) on a Java/CORBA approach to delivering geographic information over the Internet, and this approach is described further in [Laser-Scan 1998].
8. BENEFITS OF O-O

8.1 Strengths of Object-Orientation for Cartographic Production

The description of O-O given in earlier sections applies to almost any geospatial data application [Woodsford 1995]. The particular strengths of O-O in respect to mapping and cartography relate to:

- Data storage and retrieval of a model of the real world in an object database, including object versioning, long transactions, complex data modelling, validity checks and data integrity.
- Data visualisation and cartographic product generation, including active representation, multi-geometry alternatives, automated generalisation, and on-demand web access.

8.2 Benefits of O-O Mapping

The benefits that arise from the above strengths include:

- The object data model allows accurate modelling of the real world, including behaviours.
- The dataset versioning and long transactions allow efficient multi-user access to a true continuous map dataset.
- The validation methods of the O-O data model can prevent invalid data being captured, allowing immediate rectification of operator error.
- Active representation allows efficient generation of a range of cartographic quality products from a common database.
- Multi-product alternative geometries stored on single objects allow sheet-specific modifications to be stored in the master dataset. This obviates the need for tracking of similar changes through multiple product datasets, and hence reduces the costs of product update.
- An object-oriented database provides the ideal middle tier for Internet based geospatial solutions. Object-oriented databases not only provide an ideal framework in which to implement new functionality but their open architecture allows them to be readily adapted to work with the emerging standards of the Web.
- O-O generalisation methods provide an automated solution to what is historically a labour-intensive and expensive task. This new ability for dynamic generalisation is timely, given the increasing requirements for on-demand mapping, such as for presenting responsive maps in an Internet web browser.

9. CONCLUSION

Database-centric object-oriented mapping and charting software as typified by Laser-Scan’s Gothic LAMPS2 provides a set of versatile capabilities that enable a range of visual and data products to be generated from a common spatial database.

Recent advances include object database view, multiple geometry objects, and web interfaces. These supplement the existing database versioning, active representation and automated generalisation capabilities to give a complete and cost-effective flowline for multi-product generation.
REFERENCES


Note: Some background material used in this paper is updated from that in Hardy, P.G., 1996, “Map Production From An Active Object Database, Using Dynamic Representation and Automated Generalisation”, British Cartographic Society Annual Symposium, Keele, UK.
Session / Séance 51-A

Application of Method of Object-Oriented Analysis for Multi-Detailed Representation of Cartographic Data

Michael O. Govorov
Dept. of Surveying and Land Studies, The Papua New Guinea University of Technology
PMB, Lae, Papua New Guinea
Phone: 675-4757439, Fax: 675-4757483,
E-mail address: mgovorov@survey.unitech.ac.pg

Aleksej G. Khorev
Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences,
6 Ac. Lavrentjev Ave., Novosibirsk, 630090, Russia
Phone: 7-3832-350280, Fax: 7-3832- 351242,
E-mail address: aleksej@adm.ict.nsc.ru

Elena L. Kasjianova
Dept. of Cartography, Siberian State Academy of Geodesy
10 Plahotnogo, Novosibirsk, 630108, Russia
Phone: 7-383-2-432577, Fax: 7-383-2-433977

Abstract

Drawbacks and limitations of the use of batch, interactive and knowledge-based approaches for model-oriented and cartographic generalization as they are, call for necessity to create and support multi-detail representation of spatial and cartographic data. The latest can be widely used in GIS-applications that require processing of a large volume of spatial data in real-time mode, e.g. in the course of restoration of required cartographic representation in WWW.

Multi-detailed representation of data can be achieved through the results of interactive generalization system with the means of interactive logging of data, methods, and rules. At that, results of some data processing can be encapsulated into the data structure, and some of the rules - into methods.

Of interest are methods of development of multi-detailed representation of data. Such methods should be of the same nature with the ones used in the course of object and cartographic generalization. Generalization is an intellectual process that is why it is advisable to use artificial intelligence (AI) for its description. One of such approaches, which ideas are widely used in the AI, is the object-oriented one. Such approach includes several methods, such as object-oriented programming, design, analysis, and building of databases.

The overall process model of generation of OO database includes several stages, including the stage of analysis. In the present work, we have mostly dealt with this stage of analyzing of an OO database. Object-oriented analysis (OOA) permits the system to be described in the same terms as the real world.

In this article, we shall examine some approaches of OO analysis with the use of semantic object modeling approach method for notation of a multi-detailed representation.
Introduction

In recent years, the breach between the needs of cartographers and the potential of existing means of spatial data processing has been becoming more visible. To a large extent, this is explained by a sharp increase in volumes of spatial data, complex structural links between informational objects as well as by the necessity to have combined representation, processing, analysis, and usage of spatial data.

The most important of yet unsolved problems of cartography is the automated implementation of the process of model-oriented and cartographic generalization (MOCG). In the past 10 years there has been no vivid progress in system solutions of the given problem, though there is a large number of developed algorithms for realization of specific techniques of MOCG.

Model-oriented and cartographic generalization is the intellectual process of modeling, consisting in abstracting of spatial objects, links between them and their cartographic representation [Muller et al., 1995]. As all previous experience of cartographic generalization automation shows, what it is possible to model such a process only with the use of intellectual approaches. One of such approaches is the object-oriented one (OOAp). MOCG is a complex and diverse process, which requires technologies and methods that provide formality and preciseness to this process. Object-oriented approach provides such technologies for step-by-step constructing of static and dynamic elements of MOCG. OOAp comprises several methods, such as OO programming, design, analysis and building of databases.

OOAp methods can be successfully used for computer realization of MOCG by creating and supporting multi-detail representation of spatial and cartographic data. At present, this very method is considered the most promising in automation of model-oriented and cartographic generalization. It consists in storing of the basic version of spatial and cartographic objects and restoration of smaller-scale or low-detailed layers (versions of objects) at the request of user.

Multi-detailed representation of data can be achieved through a system of interactive generalization with the help of interaction, logging of data, methods, and rules. At that, results of some data processing can be encapsulated into the data structure, and some of the rules - into methods.

The overall process model of generation of object-oriented database includes several stages, including the analysis stage. In the present work, we have mostly dealt with this stage of analyzing of an object-oriented database. The model of spatial and cartographic generalization built as the result of OO analysis, later is converted into a project and then into a code. In the OO multi-scale spatial and cartographic database the data are stored together with their processing methods, and can not be processed without them. These methods become available at the moment of request. Data on all objects are encapsulated into the objects. The data are in general active, not passive.

Many facilities of traditional DBMS are important and included into OO database. Some useful for MOCG modeling peculiarities of OO database are such that it allows simultaneous storage of alternative versions of reality. This can be very useful for representation of multi-scale or multi-detailed layers of a geographic object. Another peculiarity of OO database is shared transactions. Shared group operations support workstation groups, which can then co-ordinate their actions in real-time mode. Shared transactions allow several users take part in one single group operation. One of the major advantages of OO techniques is the use of the same conceptual model for analysis, design and construction of database.

An analysis within the context of the present article means the process of specification of user requirements and system structure and functions independently of the means of implementation or physical decomposition into modules or components [Martin and Odel, 1992].

There exist rather many OOA methods or OO models for representation of object domains [Graham, 1994b]. In the present article we shall examine some semantically rich approaches of object-oriented analysis with the
use of models proposed by James Martin and James J. Odel [Martin and Odel, 1992] and SOMA of Jan Graham [Graham, 1994a] for notation of a multi-detailed representation. Application of such methods for analysis of a multi-detailed representation obtained through generalization of spatial and cartographic data is useful because they allow modeling of all operations of a cartographer, both with traditional and interactive approaches to generalization of an entity of spatial and map objects. Thus, SOMA combines a unitary notation for OOA with knowledge-based system-style rules for describing constraints, rules, global system control, database triggers, and quantification over relationships. SOMA provides support for classification, composition, general associations, pre-, post-, invariance conditions, and inheritance. SOMA is also unique in supporting fuzzy classification, which is important for requirement specification in some domains such as modeling of MOCG. It adds expert-system-style rules to objects. At design time, these rules are converted into logical assertions. In SOMA, model of an object encapsulates not only attributes and methods, but also rules.

The first chapter of the article deals with stages of OO modeling and provides an overview of model types that are created during OOA. The second chapter comprises discussion of components of identifying structures of multi-detailed spatial and cartographic objects. The third chapter presents various components of behavior of multi-detailed spatial and cartographic objects within a defined structure. The last two chapters show examples of diagrams that are used as a language of communication for OO modeling of spatial and cartographic objects and the corresponding technologies. In the conclusion, similarity of the nature of behavior analysis and the process of MOCG is discussed, as well as the tools for co-ordination of different types of models—Information Computer-Aided Software Engineering.

Models of Object-Oriented Analysis

There are a number of requirements for a model. A model is an aspect of reality, and should be constructed in such a way as to help understand this aspect of reality. A model should resemble the original, be controllable to be able to research and change the system. In cartography, map is the model of geographical reality. One of possible actions on manipulating a map is map generalization, which can include both generalization of cartographic representation and generalization of a spatial digital model of reality. In OOA, ways of modeling reality are different from traditional analysis. Modeling of the world is done in terms of object types and their changes. Stages of building an OO system are analysis, design and programming [Martin and Odel, 1992]. At this stage of analysis a model of an object or an event is being done. Later this model will be converted into a design and then into a program code. This model should reflect the way the end user comprehends the application area, and the system actions necessary for the user. A model created with the help of OO analysis will represent the original in a more natural way compared with a model of traditional system analysis. Reality consists of objects and events, which change the state of these objects. Use of OO techniques allows creating of software that models reality in a more natural way. When reality changes, OO software can be changed easily. In cartography, OO modeling can be presented in a slightly different way, as object modeling affects not only the reality but also its cartographic representation – in fact, a graphic model (see Figure 1).

Figure 1. Stages of creating spatial and cartographic OO system
As shown in Figure 1, a multi-scale GIS should comprise two levels: the level that forms a digital data model in the required scale or with required details, and the level that represents this digital data model in graphic (cartographic) form.

OO analysis decomposes the world in terms of objects that have properties and behavior and events that trigger operations that change the state of the objects. Objects interact formally with other objects. There are three primary aspects of this system. These are respectively concerned with: data, objects or concepts and their structure; architecture or atemporal process; and dynamics or system behavior. I.e., there are three dimensions, such as data, process, and control. Object-orientation combines these aspects. In the course of OOA, we can make two types of closely related models: a model of object types and their structures, and a model that describes changes of objects.

Models are shown with the help of diagrams that are called schemes. Object scheme represents the structure of the object. Event scheme shows what happens to the objects. In OO modeling there are many standards for describing schemes. Notation for creating schemes in the present work uses the standard proposed in [Martin and Odel, 1992].

**Object Structure Analysis (OSA)** defines types of objects under investigation and relations between these objects. In OO analysis, attributes of an object are usually considered to be an identified link between this object and another object or a set of objects. In the course of OSA the following information is defined:

1. What are the **object types** and what are their reciprocal associations?
2. How are the object types organized into supertypes and subtypes?
3. How is the composition of complex objects performed?

Solutions to the above questions are shown in the way of diagrams. An object scheme combines three types of such diagrams.

One of ideas of structure analysis has been borrowed from SOMA [Graham, 1994a]. For example, each level of the object types hierarchy can be represented in the way of a layer. Within the framework of the present method, the notion of a layer has a wider meaning as compared with the one used in geo-relational GIS. Here a layer is not merely a way to decompose the problem region, but also a real object with its own rights and object semantics. Layers have two differences from simple objects:

1. They are the tops of a compositional structure.
2. Each their method must be implemented in the methods of some objects within this structure.

Each layer is a collection of objects and an agreed entity, which can receive messages from other layers or objects on an individual basis.

**Object Behavior Analysis (OBA)** depicts a scheme of events, showing these events, their succession and influence on changes of the object state. Scheme of events shows the programme for processing changes in the object state. Scheme of events is implemented in terms of object scheme. When creating object and event schemes, one should upkeep the conformity that exists between these two types of presentation.

In the course of object behavior analysis the following information is defined, later shown in schemes:

1. In what state can the object be? One object can be in several states. State of an object is a set of object types, which can be applied to the object, or, in terms of OO programming, a set of associations that the object possesses.
2. What transitions in the states of the object can have place?
3. What events can take place? An event is a change in the state of the object. E.g., the following types of events can be defined: object created, object terminated, object classified, object declassified, object reclassified, attributes of object changed, objects coalesced, object decoalesced, etc.
4. What kind of operations can be applied to an object? Specification for performance of an operation is called a method.

5. What interrelations arise between objects?

6. What analytical rules are used to regulate the events?

7. What operations are presented in methods?

8. How is the hierarchy of events in the hierarchy scheme defined?

Scheme of events can comprise specifications of operations that accompany interactions between objects. Such diagram can be used for implementation of OO programme structure: object types are realized as classes; operations become OO programme operations.

After OO analysis of application area, the following schemes can be created: object scheme, event scheme, object life cycle diagram (changes in the state of object), and object flow scheme (for strategical level of planning).

Definition of data semantics should rather be performed separately from definition of objects and attributes. In OO databases semantic information is provided in an evident form. Data semantics can be represented as relations between objects or clearly stored in the OO database. Addition of rules to objects allows not only avoiding ambiguity of multiple inheritance and polymorphism, but also defining of priority rules for defaults. The rules can be inherited and overridden. Global pre- and post-conditions can also be represented in the way of rules, if necessary, which are contained in a top-level object. The rules can belong to a specific method or to an object in general.

From the point of view of specific usage, the rules can be of several types: control rules, triggers, business or exception handling rules.

Control rules are basically related to realization of multiple inheritance. Business rules define dependence between attributes. Triggers link attributes and methods.

Everything that can be expressed in the way of rules, can be expresses in the way of pre-, post- and invariant conditions.

There are six types (or sets) of rules [Graham, 1994b]:

1. Rules that link attributes to attributes. For example, if the population of a settlement is over 10,000 but less than 50,000 inhabitants, then its label should be written using capital font of 5 mm. These rules are expressed as pre- or post conditions.

2. Rules that link methods to methods. For example, the operation of exaggeration of an island that is an important point can cause an event that will trigger the operation of omitting a neighboring island, of no importance and causing a geometrical conflict. These rules are better expressed as assertions than rules.

3. Rules that link attributes to methods. For example, if a cartographic representation of lake of less than 1 sq. mm has no inflows or outflows and is not an oriented point, such lake can undergo selection and omitting. These rules are expressed as pre- or post-conditions.

4. Control rules for attributes. For example, if a line is a common border for a hydrographic object and a vegetation object, then in the course of cartographic representation its visual variables will be inherited from the hydrographic layer. These rules are expressed as pre-conditions.

5. Control rules for methods. For example, to simplify a line, which is common for a forest border and a road, simplification method from super-class “road network” will be used. These rules are expressed as post-conditions.

6. Exception handling rules. For example, if a node is connected to only one arc, then the arc is either dangling or closed. A road object that inherits geometry of such arc can undergo selection. These rules are expressed as post- or invariant conditions.
That is to say, operations may contain, along with the details of the operation’s function, parameters, and type information - invariance, pre- and post- conditions that must hold when the operation is running, starts, and terminates respectively. In this way, for example, a part of the control structure is encapsulated in the object methods, and a part – into the object completely.

**Structural Analysis of Spatial and Cartographic Objects**

In the previous chapter, we have described stages and schemes that are used for analysis of OO systems. In the present and the following chapters examples of usage of these schemes for OO analysis of model-oriented and cartographic generalization are given.

Objects of geographical types can inherit attributes, behavior, and rules of objects of geometrical and cartographic types. Objects of these types can undergo semantic, spatial and cartographic generalization. In the process of OO structure analysis, it is necessary to identify existing object types and associations between them, as well as their hierarchy and composition.

A geometrical data model is described with the help of class hierarchy [Roussilhe and Peloux, 1996]. Let us view three levels of class hierarchy: zero-topology geometrical object classes (so-called spaghetti); object classes described with the help of graphs (networks); topological geometrical object classes (fully topological level – topology of polygons and (or) triangles).

1. Zero-topology geometrical object classes consist of classes of 0, 1, 2 or 3-dimensional geometrical objects, described with a list of points with given co-ordinates.
2. On the level that represents objects as non-planar graphs, topology of intersections between linear objects is described. On this level, 0-dimensional objects are nodes of the network. Topology of 2-dimensional objects is not described here.
3. On the fully topological level, the topology of areas and/or triangles is described. Areas are described as a set of arcs that have attributes of the left and the right sides. Topology between 0-topology objects is presented with the help of a set of triangles, supported by Constrained Delaunay triangulation. Triangles are built of edges, and the latter inherit geometry of vertices [Laurini and Thomson, 1996].

Cartographic data model is a composition of spatial objects, having its own cartographic attributes (cartographic visual variables: shape, size, orientation, color (hue, value, and chroma), pattern (arrangement, texture, and orientation)) and its own geometrical forms generated by cartographic generalization.

Objects of geographical classes are, in fact, spatial models of entities of the real world. They can be created by inheriting geometrical forms from geometrical types, and they have their own semantic attributes and methods. A digital model of reality is made from geographical class objects. In the course of visualization, objects of geographical classes are represented in a cartographic form, i.e. as objects of cartographic types.

Each level of geometrical class hierarchy can be represented in the way of a layer.

A cartographic model comprises such object types as point cartographic objects, linear cartographic objects, areal cartographic objects, text objects, etc. Operations on visual representation of these object types correspond to visualization of corresponding conventional signs. Complex cartographic objects are made out of simple cartographic objects through their composition, making up a cartographic layer.

Geographical objects are also divided into layers on the basis of their semantic similarity. E.g., in the course of topographic objects modeling of the real world a layer will combine classes of hydrographic objects or road network objects.

That is to say, geographical reality objects are represented with the help of two models: digital geometrical and digital cartographic ones. Object classes of the latter two sub-models have associations among each other. On
the basis of these models, two active object-oriented databases are created: one for digital representation of geographical objects, one for their cartographic representation. As a rule, the first OO database is the basic one, as it is utilized rather strict rules. Rather often cartographic representation of a digital model depends on aesthetic inclinations of a cartographer in terms of visual variables, as well on topic and purposes of mapping, level of generalization (detail level), scale, etc. From that point of view, a cartographic database will have to be created every time there it arises a need in this or that cartographic product. To do that, we can use ready-made types of cartographic objects, slightly changing their attributes. Cartographic geometry objects can be available from corresponding entities of object geometrical types.

Figure 2 shows possible variants of extended structural scheme of geometrical layers, and their corresponding object types. In our case, only most prominent object types are shown.

![Object scheme of geometrical object types.](image)

**Figure 2.** Object scheme of geometrical object types.

**Behavior Analysis of Spatial and Cartographic Objects**

The first step in constructing an event scheme is definition of *focus or domain of analysis* [Martin and Odel, 1992]. In the given case, it is model-oriented and cartographic generalization. The second step in definition of *goal event type*. Definition of goal event type can be creation or retrieval of a spatial or cartographic object from the database, its processing with the goal of obtaining a multi-detailed version or creation of a new object at a higher level of spatial, graphical or semantic abstraction.

Next stage of behavior analysis is *definition and named of event types* that arise in the course of MOCG, their reference to main types of the events, as well as association of these event types to pre- and post-event object types that are closely related to the very appearance of such event types. Table 1 shows examples of such definitions and associations.
Table 1. Examples of events that arise in the course of object-oriented and cartographic generalization

<table>
<thead>
<tr>
<th>Type of event</th>
<th>Example of event</th>
<th>Pre-state of event</th>
<th>Post-state of event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes of Object Changed</td>
<td>Simplified</td>
<td>Base Line Object</td>
<td>Simplified Line Object</td>
</tr>
<tr>
<td>Object Terminated</td>
<td>Omitted</td>
<td>Base Point Object</td>
<td>not Point on this Level of Detail</td>
</tr>
<tr>
<td>Objects Coalesced</td>
<td>Aggregated</td>
<td>Two Base Objects</td>
<td>New Areal Object</td>
</tr>
<tr>
<td>Attributes of Object Changed</td>
<td>Collapsed</td>
<td>2D Areal Object</td>
<td>0D or 1D Object</td>
</tr>
</tbody>
</table>

The chosen types of events of MOCG should later be investigated in terms of their hierarchy conformity with other events that may take place in the system. For example, when studying an event type “simplification”, one can define hierarchy dependence among events of “area outline simplified”, “arc simplified”, “line simplified”, and “geographical object simplified” types. For each type of event, we can define its level in the corresponding generalization hierarchy of events. Results of event type generalization are required for system integration. Table 2 shows examples of event type generalization.

Table 2. Examples of event type generalization

<table>
<thead>
<tr>
<th>Level of Abstraction</th>
<th>Name of Event Type</th>
<th>Name of Event Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Objects Generalized</td>
<td>Object Generalized</td>
</tr>
<tr>
<td></td>
<td>Objects Coalesced</td>
<td>Attributes of Object Changed</td>
</tr>
<tr>
<td></td>
<td>Islands Grouped</td>
<td>Building Collapsed</td>
</tr>
<tr>
<td></td>
<td>Islands Aggregated</td>
<td>Areal Object collapsed</td>
</tr>
<tr>
<td></td>
<td>Areal Objects Aggregated</td>
<td>0D Object Created</td>
</tr>
</tbody>
</table>

On the next stage, operations and conditions of their implementation are reviewed. Operation is a process that changes the state of objects and causing events that, in their own turn, can cause implementation of further operations. MOCG operations, such as selection and omitting, simplification, exaggeration, smoothing, displacement, grouping (aggregation, spatial generalization and absorption [Govorov, 1995] (here the meaning of “grouping operations” is slightly different from those described in [Shea and McMaster 1989])), collapse, symbolization, smoothing, classification, etc., can generate various major types of events, ranging from creation of new objects to changing of object attributes. Many of these operations are complex ones, and can be represented by their own event schemes at a more detailed level of behavior specification, up to definition of methods of implementation of such operations. Figure 3 shows an example of such splitting.

Figure 3. Expression of “spatial generalization” operation in the way of a scheme of events
Operations can be internal or external type that is why it is necessary to determine them according to this notion. For example, an operation of object geometry simplification is of internal type, and an operation of object retrieval from the database is of external type, with reference to the reviewed analysis sphere. Internal type operations require an expression of the method of their implementation.

Operations are activated with the help of one or several triggers. Triggering rules are a process that may cause a specific operation after a certain set of event types has taken place and define a way according to which objects are incorporated into the operation arguments. For example, spatial generalization event of two heterogeneous areal objects triggers the operation of creating new areal geometrical and geographical objects, and later on – operation on classification of this geographical object.

For the operation to take place, sometimes it is necessary to satisfy some certain pre-conditions of such operation. The operation comprises implementation of invariant conditions and it is finished with post-conditions that cause certain events. Conditions (rules) for implementation of MOCG are having geometrical, procedural, structural, semantic and graphical character [Buttenfield and McMaster, 1991].

For example, we can omit a section of a road if the pre-conditions of such operation satisfy the following requirements: the arc that represents this sector’s geometry does not link two important geographical objects and is the dangling one.

Having defined the conditions, it is sometimes necessary to convert them into disjunctive normal form (DNF). For example, if methods for aggregation of areal and point objects are different, the following set of conditions can anticipate the operation of aggregation: the objects are located too close to each other; the objects are either of point or of areal type; the objects should be nearest neighbors; the objects should be from the same geographical layer. After normalization, this set of conditions will consist of two pre-conditions: (the objects are placed too close to each other AND are of areal type AND are nearest neighbors AND are from the same geographical layer) OR (the objects are placed too close to each other AND are of point type AND are nearest neighbors AND are from the same geographical layer).

Next step in the formulation of event scheme is definition of types of expected events so that one or more conditions are true. In a simple case shown in Figure 4a only one condition should be true for the operation to take place. In Figure 4b – several conditions are required for assessment of the true type of the condition. However, not all of the events required for satisfaction of the condition actually take part in triggering of the given operation. The example in the previous paragraph shows that only the “spatial proximity of objects assessed” event actually takes part in the triggering of aggregation operation. If only one or several events are required for an event to take place, then this condition is a redundant one (See Figure 4c).

If a condition consists of trigger rules groups and specific events can be applied only to specific groups of events, then disjunctive normalization of triggering event types is required in accordance with DNT of their events. In such a case a condition is expressed as a set of disjunctive conditions, i.e. re-grouping of trigger rules takes place (See Figure 4d).
Figure 4a, b, c, d. Simple, complex and redundant conditions and normalized triggering under complex conditions

After normalization, it is necessary to identify objects that are arguments of events that trigger operations. For example, several arguments may be required for mapping one operation into another one. To perform the operation of selecting objects for further omitting it is necessary first to perform the operation of assessing density of objects. The event “density of objects assessed” requires more than one argument, for example, for assessing several objects of the geographical object type “well”.

Processing, performed at the closing stage of specification of an event scheme, may cause changes in the corresponding structural scheme as well. Such specification can comprise, if necessary, generalization of event type triggers, specialization of goal type events, deletion of event duplicates. An operation with the same name, e.g. grouping, may have different methods of implementation: aggregation, spatial generalization, and absorption. Event types, generated by such operation, should be specific, which requires specialization of event “object grouped”.

A cycle of the above steps of the behavior analysis is performed for each internal event type.

An object can be in different states. A scheme of events reflects changes in the state of objects, generated by performance of operations. Another vivid means of showing possible states of an object is the state-transition diagram. It shows the lifecycle of an object, i.e. successful events that have taken place. An example of transitional states of a “building” object is given in Figure 5.

Figure 5. Fence diagram of possible transitional states of a “building” geometrical object.
Summary

By its own nature, model-oriented and cartographic generalization is a process. In that respect, behavior-oriented specification of the given process provides more advantages in understanding and formulating of OOCG as compared with structure static approach. A structure, defined and detailed by an OOA process, can be more similar to the original. Advantages of the use of OO analysis lie in its behavioral nature.

Though technologies for creating behavior schemes and other diagrams include some elements of integration with object schemes, special automated tools are required to keep up the interface between the schemes. Such tools are Information Computer-Aided Software Engineering (I-CASE) tools. Such tools give us the possibility to graphically show plans, models, projects, as well as develop software by generating programme code. [Graham, 1994b; Martin and Odel, 1992]. At the same time, OO languages and tools of support should not limit OO analysis methods thereof.

When using I-CASE, information is input into one of diagram types and is automatically represented in other types of diagrams. I-CASE and the information archive (IA) enable expression of uniform information by different diagrams. Informational archives are useful tools for automation of OO analysis and software designing. IAs store large volumes of knowledge and catalogues of OO classes that have high reusability. IA also contains the full code representation of objects that are used in planning, analysis, designing, and coding. OO methods are used to preserve unity and integrity of such knowledge.

References


Multi-scale representation of raster drainage network based on model

Yaolin Liu and Martien Molenaar

International Institute for Aerospace Survey and Earth Science (ITC)
P.O.Box 6,7500 AA Enschede, The Netherlands
Email: [Yaolin,Molenaar]@itc.nl

Abstract

In the development of geographical information system, the need for hierarchically-organized data structure has evolved. Automated organization of drainage network is the key for the multi-scale representation of drainage network. A numerical algorithm is proposed for automated interpretation of drainage network from raster image, indexing of network nodes, and ordering of channels by the Horton model. Channel ordering and indexing is fundamental to the automation of multi-scale representation. The program was developed by the Visual C ++ 5.0.

Introduction

In the development of geographical information system, the need for hierarchically-organized data structure has evolved. Each river system has an intrinsic hierarchical structure that can be described by various stream ordering procedures (Horton 1945, Straler 1957, Shreve 1966) and the hierarchical structure can be utilized as the basis for a feature elimination procedure. These ordering schemes reflect the topological order of links in the river tree from the source to the outlet and can be thus be usefully exploited for the multi-representation of drainage networks. Automated organization of drainage network is the key for the multi-scale representation of drainage network. Jurgen Garbrecht (1997) discussed the automated organization of drainage-network based on the Strahler’s system of stream ordering. E. Rusak (1990) applied Horton’s system of stream ordering in generalizing rivers for reference maps at a variety of small scales. But there is little article to discuss the automated organization of drainage network based on the Horton method. The multi-scale representation of drainage network could also benefit from the hierarchical structure, as the goal of the multi-scale representation of drainage network is to retain a recognizable representation of a feature, rather than a simple mathematical reduction. We are here concerned with raster images.

Geographic and cartographic requirements are fulfilled by only one classification—Horton’s classification allows for the preservation of aggregated river entities or linkages under one proper name. In the case of Horton’s approach, the initial classification must be conducted by a person knowledgeable in the system. For this reason, Horon’s classification is more complex and demanding. Horton’ method is the most suitable for carrying out a generalization process.

In this paper an algorithm to analyze images of raster channel networks, index network nodes, and order the channels by the Horton method is presented. The computational steps are presented in detail to permit the implementation of the algorithm by the reader. The initial network definition and other related raster data that are needed by the algorithm are produced by scanning the map. The data format is bmp.
Horton’s system of ordering (Horton, 1945) has been shown to be the most useful for establishing a database structure that is amenable to multi scale representation procedure that can be used in computerized data bases where the system may be required to generate different maps at quite different scales.

**Objects and Relationships Between Objects**

**Objects**

A channel network consists of a set of channel links connected by network nodes (Horton, 1945; Shreve, 1966) hereafter simply referred to as nodes (Figure 1). Three types of nodes are encountered in channel network: the watershed outlet node, upstream tips of the channel network where channel links originate (source nodes), and points at which two or more channel links join (junction nodes). The channel links can be ordered according to Horton (1945).

Drainage networks can be schematized as sets of basic unit: links and nodes (Figure 1). The channel links is a basic objects for data organization. In appearance, they resemble a topologic tree with a main trunk and several branches. From these, many smaller branches stem and from these branches even smaller branches diverge, and so on. A topologic graph unveils the basic organizational structure that is hidden by the complexity of the naturally existing system. It is this structure which provides a basis for ordering the river network, establishing of commonalties and differences between various networks by which they can be described and compared, and subsequently, and by which they can be generalized in maps (E. Rusak, 1990).

**Figure 1. Basic units of drainage network**

**Relationships Between Objects**

Let N be a set of nodes, L be a set of links within the framework of the FDS (Molenaar, 1995), thus (before the construction of a Horton):

- for each \( l_i \in L \), there exists at most one node \( n_b \in N \) for which:
  \[
  \text{Begin}(l_i, n_b) = 1 \quad \text{and thus} \quad \text{End}(l_i, n_b) = 0 \quad \text{if} \quad l_i \text{ does not form a loop;}
  \]
- for each \( l_i \in L \), there exists at most one node \( n_e \in N \) for which:
  \[
  \text{End}(l_i, n_e) = 1 \quad \text{and thus} \quad \text{Begin}(l_i, n_e) = 0 \quad \text{if} \quad l_i \text{ does not form a loop;}
  \]
- for each \( n_i \in N \), there may or may not exist an link \( l_i \) for which:
  \[
  \text{Begin}(l_i, n_i) + \text{End}(l_i, n_i) = 1;
  \]

**Adjacency Relationship between Nodes**

Two cells are adjacent if they meet the requirement of 8 neighborhood.
**Extended adjacency Relationship between Nodes**

Two nodes are adjacent if they are connected by a link in the network W:

For two nodes $n_i$ and $n_j$ if there exists a $l_k$ such that:

$\text{Begin}(l_k, n_i) + \text{End}(l_k, n_i) = 1$, and $\text{Begin}(l_k, n_j) + \text{End}(l_k, n_j) = 1$,

then

$\text{Adjacent}(n_i, n_j) = \text{Adjacent}(n_j, n_i) = 1$.

**Adjacency Relationship Between Links**

Two Link $l_1$ and $l_2$ are adjacent at an node $n_i$ if their join point is the same junction node.

$\text{Adjacent}[l_1, l_2 | n_i] = 1 \iff n_i \in l_1 \cap l_2$

Otherwise $\text{Adjacent}[l_1, l_2 | n_i] = 0$.

**Hierarchical Relationship in the Drainage Network**

The creation of a river network data base requires the hierarchical structure of a river network through the application of Horton’s ordering method. The rules specified by Horton (1945) become the basis for hierarchically coding links. In point form, They are:

- **order 1** — fingertip tributaries,
- **order 2** — tributaries that receive order 1 stream
- **order 3** — tributaries that receive at least one order 2 stream and may also receive order 1 stream.
- **order 4** — with at least one order 3 stream and usually other lower orders and so on.

the main river becomes of the highest order shown in Figure 2.

![Figure 2. Horton’s stream ordering scheme](image)

**Automated organization of Drainage network based on Horton’s system**

The algorithm consists of two computational steps. First, the raster image of the network is processed. A Horton order is assigned to each channel link and a table of channel attributes is created. This table contains the coordinates of the beginning and ending node of each channel. And thereby captures the topology of the network. The numerical interpretation of the raster network begins at the source nodes and proceeds in a downstream direction along the channels to watershed outlet node.
Input Data

The data of network can be gotten by scanning drainage map. The data format by scanning is bmp. The program can transform the bmp data into raster image. The only outlet node is needed to be given as initial input for the proposed algorithm. The other node types are automatically recognized by the program according to the properties of nodes. The channel network is identified in the network raster as strings of connected raster cells having a value of 100, on the background of cells having a value of 0.

Identifying the node types

According to 3x3 models and the properties of three types of nodes, three types of nodes can be identified automatically. The cells at the tips of the network (source nodes) are given a value of B, and the cells at the junction of the network (junction nodes) are given a value of C, and the cell at the outlet of the network (outlet node) is given a value of A. In the example network raster shown in Figure .

Determining the main stream of drainage network

Generally speaking, the main stream has the longest length in drainage network. The distances from source nodes to outlet node are computed. The longest one is selected as main stream. The junction nodes in the main stream are put into the stack for tracing the other links and ordering them.

Determining the order of channel links

The numerical interpretation of the network begins at the junction node in main stream to the source node. The automated organization of Horton system can be accomplished by a recursive procedure. At first, the main-stream of drainage network are found out. And all junction nodes in the main stream are put into the stack. All the cells except the junction nodes in main stream are assigned order 1. Then, the junction node in the main stream is popped from the stack. The tributaries connected the main stream are researched. There are two cases. Case 1: If there is only one source node shown in Figure 3. All cells in tributary or branch are assigned order 1. At the same time the order of the stream joined at the tributary or branch is order of tributary +1. Case 2: If there are more than two source nodes in the tributaries shown in Figure 4. The distance from each source node to the junction node is computed. The longest tributary or branch is put into the stack, and All cells of the longest tributary or branch except the junction cells are assigned order 1. The order of the stream connected longest tributary or branch by the junction node is the order of the longest tributary or branch +1. The algorithm will end if the stack is empty. The algorithm is as following:

Figure 3. Case for only one source node

Figure 4. Case for two or more than two source nodes
Order ( )
{the junction nodes the main stream of drainage network are put into the stack if (the stack is not empty) then
{the junction node is popped, the tributaries are researched and the source nodes are found. If (source node = 1)
then the order of the branch is assigned as 1
else {computing the distance between junction node and source node.
determining the longest branch and the junction nodes are put into the stack.
assigning the order of the longest branch and adjusting the order of the stream connected the longest
branch by junction node.
Order ( )} } }

Multi-resolution representation

It is easy to implement the multi-resolution representation of drainage network after the drainage network is
automatically organized based on the Horton system. The specific criteria for channel links selection will
ultimately depend on the purpose of the application. As such, They may related to the order of channel link, the
volume of flow, length, historical significance, economic importance, or some combination of these or other
criteria. In other word, any number of valid criteria may be applied in a given case.

Selection of channel links for different scales depend on purpose of application, scale representation, quantitative, quality and location of objects. The 5-tuple is defined for the selection of channel link of multi-scale representation.

Selection(l_i)=(app, sca, quan, quali, loca)

Where quan represents the quantitative properties such as the length or width of channel link.

Where qual represents the quality properties such as order, forever river, season river and so on.

Loca represents the location of channel link in the network

app represents the field of application.

sca represents the scale

Conclusion

An algorithm that automatically determining channel Horton order and node indices for raster channel networks is presented. Horton orders are determined by a cell-by-cell trace of the raster network in a downstream or upstream along the channels beginning at junction nodes to source nodes and outlet node. The automated ordering of drainage network based on Horton method is more complex than the one based on other ordering method. The Horton ordering system has proven to be the most useful one for multi-scale representation of drainage network, because it combines topological order with metric properties. The 5 tuple for selecting channel links is used for implementing multi-scale representation of drainage network.

Acknowledgement

This work was supported by International Institute of Aerospace Survey and Earth Science and Fund of Education Ministry of P.R. China.
References


Session / Séance C3-B

Production and “Intelligence” of Raster Maps

C. Armenakis and E. Siekierska
Centre for Topographic Information
Mapping Services Branch, Geomatics Canada
615 Booth St., Ottawa, Ontario, Canada K1A 0E9
e-mail: {armenaki}{siekiers}@nrcan.gc.ca

F. Hegyi and P. Pilon
Hegyi GeoTechnologies International Inc.
707-170 Laurier Ave. W., Ottawa, Ontario, Canada K1P 5V5
e-mail: {fhegyi}{ppilon}@hegyi.com

Abstract
Rasterized versions of paper maps are becoming more and more popular. The digital rasterized topographic map is an evolutionary step of the conventional paper map towards the digital world. It satisfies a large number of users who prefer the familiar cartographic and easy-to-interpret view of paper maps with the advantages of the digital environment and its tools. At the Centre for Topographic Information, rasterized versions of topographic maps are produced through a vector-to-raster (digital-to-digital) conversion process as a by-product of digital map production. These georeferenced raster maps are primarily used as a backdrop or base layer for other information. Certain “intelligence” can be added to the pixels of raster maps in the form of linking pixels and image objects with vectors and attributes from existing spatial and thematic databases. The implementation of this kind of “intelligence” eventually leads towards the creation of “smart” maps, where topological and other type of queries, usually associated only with vector data, can be handled as well. The “smart” raster maps can be a key layer in a geospatial infrastructure and in GIS systems, can be integrated with vector data for analytical spatial operations, can support GPS-based position and navigation operations and can be used for terrain visualization. In this paper, the production of raster maps at Geomatics Canada and several value-added applications of “smart” raster maps are described.

1.0 INTRODUCTION

As Geographic Information Systems (GIS) and Global Positioning Systems (GPS) have become popular tools for handling spatial information, users of topographic maps have started to move towards the digital environment. As a first step, conventional hard-copy maps were scanned into raster products for use in computer assisted applications. A number of national mapping organizations have started producing raster maps by scanning their existing paper versions [5],[6]. In this process, some problems were encountered in terms of quality of products and up-to-date information, and their use in the digital environment. In addition, users started to formulate requirements regarding the levels of spatial intelligence that are needed within digital raster products.
The digital rasterized topographic map is an evolutionary step of the conventional hard copy version in its transition to the digital world. As such, it will satisfy the large number of users who prefer the familiar cartographic and easy-to-interpret view of paper maps with the advantages of the digital environment and its tools. In digital form raster maps can be enhanced, modified and electronically transmitted and viewed. It is also possible to integrate raster maps with other digital georeferenced spatial data to provide a more complete information package as the raster structure of these maps allows retrieval of the positional and thematic information of each pixel.

In an effort to provide solutions to the above challenges, the Centre for Topographic Information (CTI) of Geomatics Canada started a study on alternative raster products, including opportunities for attaching various levels of intelligence to these new products. In this paper, the results of this study are documented including examples of commercial applications resulted from the cooperation between CTI and industry (Hegyi GeoTechnologies International, HGI).

2.0 PRODUCTION OF RASTER MAPS

The primary objective of this study, by the Centre for Topographic Information of Geomatics Canada, was to investigate the feasibility of producing digital raster map versions of the National Topographic System (NTS) maps as a by-product of its current digital cartographic and map production process. Based on the basic property of the raster map product - an image representation of the map - investigation was carried out to evaluate the feasibility and capabilities of digitally producing raster topographic maps.

The National Topographic System (NTS) map contains topographic information coming from the National Topographic Database (NTDB) and information from secondary source data not included in the NTDB. The Cartographic Editing System (CES) is then used to integrate the NTS files for final map symbolization, to perform the map assembly, quality control, and for exporting the cartographic NTS map files in Postscript format to the Cartographic Imaging and Printing Services (CIPS) for publishing the map. Based on this map production process the generation of raster map files is done through a digital-to-digital process without involving any scanning.

2.1 Production process

The final cartographic file (CES file) of the interior of a map is exported in Postscript (PS) format to the CIPS’ DALiM publishing prepress software, which as a Raster Image Processor (RIP) converts the PS files to a raster file format in order to print digital image files. To capture the map details and maintain map accuracy the vector-to-raster conversion (rasterization) is done at 600ppi (0.085mm) and with 24-bit per RGB pixel. The raster file is converted to TIFF format for export.

The TIFF raster map file is imported to the PCI image analysis and processing software for colour adjustment and georeferencing. Colour adjustment is applied by manipulating the RGB histograms, in order to visually match them with the paper map colours (this adjustment may not be necessary, as the RGB colour model is device dependent). The georeferencing is performed using the UTM grid intersections as control points since their cartographic coordinates are known. The geometric transformation is the affine (1st degree polynomial) and its parameters are determined by measuring the UTM grid images and projecting them to their known ground coordinates. The resampling algorithm used is the bilinear interpolation. It was selected to ensure line continuity and smoothness, sharpness of text, lines, and colour, and minimization of any artifacts. The output ground pixel size is set to 4.25m, which is equivalent to the final resolution of 300ppi. Presently, nine well-distributed points are used as control points and four as check points. Due to the digital nature of the cartographic files, four corner control points and four check points in the interior of the map can also be sufficient.
The georeferenced raster map file is converted from the application specific format (.PIX) to GeoTIFF format, which carries the georeferencing information (reference ellipsoid, projection, scale, cartographic and geographic coordinates). GeoTIFF is supported by many types of GIS and image analysis software vendors and packages such as: CARIS, ESRI, Intergraph, ERDAS, ER Mapper, Laser-Scan, MapInfo, TNT MicroImages, Map Maker Pro, PCI and SoftDesk. It can be used with TIFF viewers as well.

At 300ppi and 24-bit pixel, the dimensions of the raster map file at middle latitudes are approximately 8470x7060 pixels resulting in file size of about 180Mb. Due the nature of the raster map, which contains large areas of constant colour, data compression achieves high reduction rates. For example, using the LZW lossless compression algorithm the size of the compress file is about 11Mb (16:1). Similar compression rates are achieved with the WinZip utility. To facilitate the use of the file with micro-computers the file can be also resampled to 150ppi resulting in file size of about 45Mb. This resolution still allows for 2x zoom-in for the Macintosh and PC monitors, which have resolutions of 72ppi and 96ppi respectively.

2.2 Advantages over scanning existing paper maps

Considering the complete digital map generation process, the advantages of this approach are:

- No need to scan the paper product, an analog-to-digital conversion process.
- Easily obtained and cost-effective by-product, since all new editions of the topographic maps are produced digitally.
- Higher quality end-product than the one produced by scanning the paper maps, leading to sharper/crisper and cleaner map images.
- Better geometric and radiometric fidelity.
- Timeliness product as the raster map files can be available before the printed map is in circulation and even before it is printed.
- Elimination of the various artifacts introduced during the scanning of paper maps.

2.3 Metadata

A map contains both the graphic (map window) and the textual information about the graphic. In the NTS map series the textual information is located in the marginalia of the graphic (surroundings) and on the reverse side of the graphic (map legend). Excluding the legend as being common to all maps, rasterization of the map can be done for both the graphic and the surrounding or only for the graphic. In the former, the user can access/view the map information by panning around the map image, which is a not very effective way. In addition by including the surroundings in the raster map, the size of the file increases significantly and does not also support the concept of creating a seamless raster database. By rasterizing and georeferencing only the graphic part of the map and including the map metadata as a support file linked to the image file, the product becomes much more versatile for meeting users’ present and future needs. It effectively and efficiently supports search and retrieval on both metadata and image, smaller size files and the creation of geographically continuous raster map database.

The metadata file contains a large amount of information about the digital raster product. This information includes: identification items (map id, name, province), publication dates, data sources, projection, scale, accuracy, format, raster dimensions, file size, spatial and radiometric resolutions, cartographic and geographic coordinates of the raster and map spatial windows, and contact address. The metadata serve for production purposes but mainly serve the raster maps user both in using the raster file and in searching and retrieving information. For example, the user is able to search for raster map availability based on: map id information, spatial point and window geographic information, temporal ranges and accuracies. The metadata database was created using MS ACCESS 97 and it is accessible via the Internet.
2.4 Technical issues

The rasterization procedures, CTI’s digital-to-digital and the analog-to-digital (by scanning) are strongly influenced by the spatial resolution, as well as the speed. Both approaches result in relatively large data files, which in turn slow down the display of raster files. Technical issues associated with these problems are summarized below:

1. File size compression is important for applications utilizing digital raster maps [4]. In the future, raster map files may be downloaded to users via the Internet, in which case file compression is an absolutely necessary. To conserve storage space in the raster map database, on-the-fly image compression and decompression would be ideal. It may also be useful to compress files when cutting a user requested CD-ROM. However, this may complicate easy access to the raster map files since the files will have to be decompressed. Both the compression and decompression processes take time.

2. Many image handling software systems include file compression routines. The PCI Image analysis software system allows users to choose LZW compression when converting from its image file format (.PIX) to the GeoTIFF format. Graphics software packages such as Adobe PhotoShop also include LZW compression routines. The LZW compression results in large reductions in file size. However, it is important to note that the LZW compression software is patented. Therefore, only graphical and map viewers that have paid licensing fees to the UNISYS Corporation are able to view LZW compressed files. Raster files can also be compressed using standard compression routines such as PKZIP and WINZIP.

3. Several companies have developed specialized image compression software. One such company is LizardTech, with its MrSID (Multiresolution Seamless Image Database) compression software, which is based on the Discrete Wavelet Transform technology and on the MCICR (Monte Carlo Image Conversion & Representation) colour quantization process both developed at Los Alamos National Laboratory in Los Alamos in New Mexico, USA. The MCICR process has been used to covert 24/32-bit images into a standard 8-bit file format without loss of image quality. This allows for files to be 1/3 of their original size. Images compressed with MCICR can be written in all standard graphic file formats and do not require decompression.

4. The fast transmission of large raster map files over the Internet has been another issue. In this direction a company called HMR uses a multi-resolution-tiled image architecture and image compression to minimize the amount of image data which needs to be transferred on the client’s request [3]. Therefore only those data covering the region displayed on the screen are transmitted.

5. The DALiM publishing software shares with many other software packages the function of converting Postscript files to a raster file format in order to print digital image files. These systems are known as raster image processors (RIPs). Several alternatives to the DALiM publishing software can be found. These can be categorized as falling into two types: those associated with printing any types of graphic images and those which specialize in producing as maps and other geographical images.

6. Presently, the measurements for the georeferencing of the raster map files are performed manually. Since, the UTM grid intersections are measured it is possible to automate the measuring process by automatically driving to the UTM grid locations and matching their images with a pre-defined grid template [2]. This approach will expedite the georeference process.

7. Presently, the metadata file is in text format. To facilitate its use it can be incorporated with the viewer(s) of the raster map file in the form of pull-down/pop-up window including legend information.
3.0 INTELLIGENCE

In a digital vector environment, there is practically an open list of attributes that can be linked to graphic entities through either an automatic process or by deriving combinations of characteristics of the vector data. In the digital raster world, on the other hand, the assignment and derivation of attributes or “intelligence” to graphic entities is more complicated and often time consuming, as the raster structure allows only for the retrieval of the positional and thematic information of each pixel.

With the increasing demand for digital raster maps, the requirement for attaching various levels of intelligence to graphic map entities and for performing various spatial operations also increase. In particular, the requirement of users is summarized below:

1. Digital raster maps with basic intelligence (level I). These georeferenced products are generally required by users whose primary interest is to view the maps through a computer for positioning and measurements, in a manner similar to how paper maps are being viewed but with various tools (i.e., position, distance measurements, zoom-in). In this environment, the display of GPS positions on the raster images is rapidly becoming a major requirement. Examples of this application are found in fleet management and monitoring, security monitoring, delivery of valuable goods and money with armoured cars, delivery of goods by vans and trucks, natural resource management, tourism, and in monitoring cars to prevent theft.

2. Digital raster maps with attribute intelligence (level II). These georeferenced maps have map objects link to attribute databases. They are often used in combination with GPS by those who require the location of important “addresses”, such as clients, suppliers and administrative offices. In the digital raster environment, these addresses are generally located manually on the map images and linked to a database. In operational applications, users can type in an identifier of each location, such as name of client, the icons of these locations are displayed on the map, showing accurate georeferencing. Examples of this application, in addition to the ones listed above under (1), are in taxis dispatch, law enforcement, and courier services.

3. Digital raster maps integrated with vector data intelligence (level III). This product is rapidly becoming the most desirable option with the sophisticated users of GPS as various spatial operations are performed using the vector data and the raster map is used for input and for display of the results. The vector data, registered to the raster map image, offers the advantage of the following application opportunities:

   - Location of points on the map with accurate georeferencing directly through a data base, such as addresses in a city, business locations, offices, path route analysis, as well as planimetric features such as roads, rivers, lakes, forest stands, agricultural fields, farms and other relevant areas. Examples of this application include those listed above, especially in areas where the intelligence associated with the digital maps involves a large database, such as all the addresses in a city.

   - Navigation between selected points, such as the GPS position of a vehicle and an address, two addresses defined through the database, as well as selection of routes for security and delivery vehicles. This application is mostly with GPS and value-added software packages that integrate GPS and wireless communication technologies to provide the monitoring of mobile units with reference to assigned routes. Examples of applications include those listed above, with special focus on the use of road network in cities by taxis, security vehicles and law enforcement mobile units, as well as in cross country travel by truckers. In addition to location of points, there is frequent requirement to measure distances between points and along routes at different zoom levels.

   - Draping the raster image maps over digital elevation models to create 3D views for terrain visualization (Figure 1). This type of application is used mostly in natural resource management and landscape architecture designs. Examples of applications include the creation of perspective views to simulate the effects of logging on watersheds, examination of line-of- sight for wireless data transmission in mountainous terrain, fly-bys, and the planning recreational activities where 3D views are required.
4.0 EXAMPLES OF VALUE-ADDED APPLICATIONS

By adding intelligence to the products described above a wider range of value-added applications can be created. In fact, there is a dynamic interaction between product development and value-added applications. Product developments are largely driven by applications, and as these products get into operational use, new value-added applications are formed.

Value-added applications resulting from the development of digital raster map products are highlighted below as follows:

1. Automatic Vehicle Locator (AVL) software packages. There is a rapidly expanding industry, using mostly digital raster maps. The value added components include software packages operating at monitoring centres for tracking vehicles, including:
   - Display of mobile units on digital raster maps.
   - Sounding alert at Monitoring Centres and displaying the position of vehicles when they deviate from assigned routes, activate panic buttons in case of emergency, or signal for routing assistance.
   - Displaying pre-programmed and pre-recorded messages on a base computer at the Monitoring and Dispatch Centre when the appropriate button is pushed on a particular AVL hardware unit and the corresponding signal is transmitted back through wireless infrastructures.
Creating new dispatch systems for taxis, law enforcement and security vehicles, and delivery trucks, based on spatial information available through the products using raster map image with attached intelligence.

Creating new electronic bidding systems for truckers to pick up and deliver loads, using spatial information based on digital raster maps and their associated intelligence.

Creating new electronic monitoring systems for private and commercial automobiles to prevent or minimize theft.

Creating new compliance monitoring systems in natural resource management to ensure that environmental protection guidelines are maintained.

2. Mobile Office software packages on laptop computers. As natural resource maps are becoming available in digital form, users are taking advantage of the opportunity to view these maps in the field on laptop computers with GPS and wireless communication technology. In this environment, field samples are often collected and their attributes recorded on electronic forms with GPS linkages to the digital maps. In addition, vector maps about natural resources draped over raster images and DEM’s are frequently used in this environment, resulting in further value-added products.

3. Digital raster maps on pen-based hand-held computers. There is an increasing number of applications where digital raster maps are viewed in the field on hand-held computers linked to GPS to show exact locations. Further new GPS receivers are loaded with maps. Applications of this option include building inspectors, sales staff, and persons working in the health care sector. In each area, additional software is being developed to integrate the spatial information with the specific application focus.

4. Wireless communications. Currently, a prototype unit is being released that includes a cell phone, GPS unit and a specialized micro-processor. With this system, it is possible to push one of the pre-programmed keys on the unit to request a map segment through the Internet, showing the exact position of the person. This type of application will likely bring about extensive value-added technology and software development.

5. Data access through CEONet. With the prototype unit referred to in (4) above, a new value added software package has been created by HGI. The package searches data bases registered in CEONET through metadata, accesses their appropriate web sites, extracts maps segments around GPS positions, and down-loads digital raster maps with some level of intelligence to users in the field.

5.0 CONCLUSIONS AND RECOMMENDATIONS

Digital raster maps have a major role in applications that are computer-based and involve GPS and wireless data transmission. They are preferred in most cases to the vector maps produced by a GIS in the traditional manner, as they resemble conventional maps, which have been designed for most effective visual communication based on the principles of graphic semiology [1]. The main reason for this is that users have got accustomed to paper maps that already display classified (interpreted) features, are standardized and are esthetically pleasing. Raster maps contain colour-enhanced highlighting of important features, the roads are presented with realistic dimensions, and key thematic details are illustrated with appropriate symbols. Examples of such maps are city maps, provincial, state and national road maps, recreational maps, and thematic charts.

In order to use traditional paper maps in computer-assisted applications, they need to be converted into digital form. This conversion is often achieved by scanning the hardcopy products at resolutions that are appropriate for the targeted applications. An alternative method for producing digital raster maps is through software, as discussed in this paper. This approach uses the vector cartographic files and the resultant raster images are generally of higher quality than their scanned counterparts. In addition, the software generated raster maps
offer greater flexibility for producing by-products at different resolution, an important consideration when different levels of detail are required in applications. For example, when an entire city map is displayed in the computer, both products show acceptable images. However, when it is necessary to zoom in to examine a particular city street in order to locate an address, the software generated products show a crisper image, maintaining the illustrations in a readable form, as well as minimizing the discontinuity of linear features.

A key requirement in the use of digital raster maps is the availability of some level of intelligence. The most cost-effective method of attaching intelligence to digital raster maps is by overlaying and registering vector files on them. Once vector files are registered, they can provide all available intelligence, while at the same time users can view the raster images. With some application software, it is possible to turn off the display of the vector lines, while at the same time the intelligence is retained under the raster images.

The increasing availability of digital raster maps is generating many value-added applications, especially in the automatic vehicle tracking niche market. Initially, efforts were focused on displaying the raster maps on desktop computers at central locations. During the past years, value-added software products have made it possible to use these digital maps in the field on laptop computers, and in combination with GPS and wireless data transmission. Recently, there are strong efforts being made to make digital map segments available in mobile units, as well as for individuals working in hazardous environments, on AVL hardware units equipped with graphic display screens, GPS and wireless data transmission through the Internet.

Finally, there is a rapidly growing value-added industry that is focused on making seamless digital maps available via the Internet. In particular, requests for digital maps will be made more and more by specifying a location such as an address or GPS coordinates, and users will want the map segment around such locations. This niche market involves desktop, laptop and hand-held computers, as well as AVL hardware units. Hence, it is recommended that increased research and development efforts be devoted to these types of application, especially as cable companies are beginning to provide high speed Internet access to homes and businesses. In addition, as the speed of wireless data transmission technology is being improved, accessibility of Internet web sites from the field and from mobile units will increase, thus creating further demands for the downloading of digital maps or map segments through wireless infrastructures.

Acknowledgements

The authors wish to thank J. Handy, B. Kelso and M. McLean, all from CIPS, for their contributions to the tests and production of the raster maps.

References

Enhancing a Database Management System for GIS with Fuzzy Set Methodologies

Emmanuel Stefanakis and Timos Sellis
Knowledge and Database Systems Laboratory
Department of Electrical and Computer Engineering
National Technical University of Athens, Zographou 157 73, Athens, Greece
{stefanak, timos}@cs.ntua.gr

Abstract
The methods used in commercial GIS packages for both the representation and analysis of geographic data are inadequate, because they do not handle uncertainty. This leads to information loss and inaccuracy in analysis with adverse consequences in the spatial decision-making process. The incorporation of fuzzy set methodologies into a DBMS repository for the application domain of GIS should be beneficial and will improve its level of intelligence. Focusing on this direction the paper addresses both a representation and a reasoning issue. Specifically, it extends a general spatial data model to deal with the uncertainty of geographic entities, and shows how the standard data interpretation operations available in GIS packages may be extended to support the fuzzy spatial reasoning. Representative geographic operations, such as the fuzzy overlay, fuzzy distance and fuzzy select, are examined, while a real world situation involving spatial decision making is presented.

1. Introduction

Uncertainty (sometimes the terms imprecision and vagueness are used instead) refers to the imperfect and inexact knowledge concerning some domain of interest [Goodchild and Gopal, 1989; Openshaw, 1991, Kruse et al., 1991]. The uncertainty is an inherent feature of geographic data and may arise through [Altman, 1994]:
a) incomplete information associated to them; b) the presence of varying concentrations of attributes; and c) the use of qualitative descriptions of their attribute values and relationships. Currently used methods for the representation and analysis of geographic information are inadequate, because they do not tolerate uncertainty [Wang et al., 1990]. This is largely due to the underlying membership concept of the classical set theory, according to which a set has precisely defined boundaries and an element has either full or no membership in a given set (Boolean logic).

The representation of geographic data based on the classical set theory has a tight effect on reasoning and analysis procedures [Stefanakis, 1997; Stefanakis et al., 1999]. Specifically, the employment of a sequence of basic GIS operations to support a real world situation, such as that of residential site selection, is accompanied with all problems of an “early and sharp classification” [Stefanakis et al., 1999, 1996; Stefanakis, 1997]. The overall decision is made in steps which drastically and sharply reduce the intermediate results. Any constraint is accompanied with an absolute threshold value and no exception is allowed. For instance, if the threshold for a level land is slope = 10%, a location with slope equal to 9.9% is characterized as level, while a second location with slope equal to 10.1% is characterized as non-level (steep). Moreover, for decisions based on multiple criteria, it is usually the case that an entity (i.e., an individual location), which satisfies quite well the
The majority of constraints and is marginally rejected in one of them, to be selected as valid by decision-makers. However, based on Boolean logic, a location with slope 10.1% will be rejected (as non-level), even if it satisfies quite well all other constraints posed by decision-makers. In addition, decision-makers are obliged to express their constraints through arithmetical terms and mathematical symbols in crisp relationships (e.g., slope < 10%), since they are not allowed to use natural language lexical terms (e.g., level land). Finally, another effect of classical set theory is that the selection result is flat, in the sense that there is no overall ordering of the valid entities as regard to the degree they fulfill the set of constraints. For instance, dry-level layer highlights all locations which satisfy the constraints: dry land (threshold 20%) and level ground (threshold 10%). However, there is no clear distinction between a location with moisture = 10% and slope = 3% and another with moisture = 15% and slope = 7%.

These impediments call for a more general and sound logical foundation for GIS. In the literature several attempts can be found to overcome these impediments, such as the adoption of statistical and probabilistic models. However, most of these attempts handle imprecision in geographic data as randomness. It is argued that there are several aspects of geographic data that cannot be attributed to randomness [Leung and Leung, 1993; Altman, 1994]; such as, complexity, missing information, and the use of natural language. Fuzzy set theory, on the other hand, seems to be an appropriate means of modeling uncertain data and provides methodologies to support reasoning based on these data. The question is basically how to incorporate fuzzy analysis into GIS [Openshaw, 1991]. This paper addresses two issues regarding the incorporation of fuzzy set methodologies into a DBMS repository for the application domain of GIS. First, a representation issue: it is shown how uncertainty, which characterizes geographic features, may be incorporated into the spatial data model; and second, a reasoning issue: it is shown how fuzzy logic methodologies may be incorporated into the basic data interpretation operations available in GIS packages. An extended version of this work can be found in [Stefanakis et al., 1999].

The discussion is organized as follows. Section 2 presents a general spatial data model and the basic operations available in GIS packages [Stefanakis and Sellis, 1996, 1997; Stefanakis, 1997]. Section 3 examines a simplified real world application (i.e., site selection for a residential housing development) and shows how data-interpretation operations may be combined to compose a composite procedure and support the task posed to decision-makers. After a brief introduction of fuzzy set theory in Section 4, the incorporation of fuzzy set methodologies into a DBMS repository for the application domain of GIS is examined. Specifically, Section 5 extends the general spatial data model considered to accommodate the fuzziness of geographic entities, while Section 6 shows how the standard data interpretation operations may be extended to support fuzzy spatial reasoning. Section 7 reexamines the simplified example of site selection, presented in Section 3, in order to highlight some of the advantages provided by the incorporation of fuzzy set methodologies in spatial decision making process. Finally, Section 8 concludes the discussion by summarizing the contributions of the paper and giving hints for future research in the area of fuzzy set methodologies for geographic data handling.

2. Spatial Data Modeling and Operations

According to the model first introduced by Tomlin [Tomlin, 1990] geographic information can be viewed as a hierarchy of data [Samet and Aref, 1995; Stefanakis and Sellis, 1996, 1997]. At the highest level, there is a library of maps (more commonly referred to as layers), all of which are in registration (i.e., they have a common coordinate system). Each layer corresponds to a specific theme of interest and is partitioned into zones (regions), where zones are sets of individual locations with a common attribute value. Examples of layers are the land-use layer, which is divided into land-use zones (e.g., wetland, river, desert, city, park and agricultural zones) and the road network layer, which contains the roads that pass through the portion of space that is covered by the layer.
There is no standard algebra defined on geographic data. This means that there is no standardized set of base operations for geographic data handling. Hence, the set of operations available in GIS varies from one system to another and heavily depends on the application domain. However, their fundamental capabilities can be expressed in terms of four types of operations: a) programming, b) data preparation, c) data presentation, and d) data interpretation operations. Data interpretation operations are those that transform data into information and as such they comprise the heart of any system for handling geographic data. Data interpretation operations available in GIS characterize [Aronoff, 1989; Tomlin, 1990; Samet and Aref, 1995; Stefanakis and Sellis, 1996, 1997; Stefanakis 1997]: a) individual locations, b) locations within neighborhoods, and c) locations within zones; and constitute respectively the following three classes of operations:

- **Local operations**: they include those that compute a new value for each individual location on a layer as a function of existing data explicitly associated with that location.

- **Focal operations**: they compute new values for every individual location as a function of its neighborhood. A neighborhood is defined as any set of one or more locations that bear a specified distance and/or topological or directional relationship to a particular location (or set of locations in general), the neighborhood focus.

- **Zonal operations**: they include those that compute a new value for each individual location as a function of existing values associated with a zone containing that location.

All data interpretation is done in a layer-by-layer basis. That is, each operation accepts one or more existing layers as input (the operands) and generates a new layer as output (the product), which can be used as operand into subsequent operations. Hence, data interpretation operations may be combined to compose one or more procedures (a procedure is any finite sequence of one or more operations that are applied to meaningful data with a deliberate intent) and accomplish a composite task posed by the spatial decision-making process. A simplified real world example is presented in the following Section. Table 1 summarizes the basic classes of data interpretation operations accompanied by representative examples [Aronoff, 1989; Tomlin, 1990; Stefanakis and Sellis, 1996, 1997).

### Table 1. Basic classes of data interpretation operations.

<table>
<thead>
<tr>
<th>Classes of Operations</th>
<th>Examples of Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Local Operations</strong></td>
<td></td>
</tr>
<tr>
<td>- Classification &amp; recoding</td>
<td>re-code, re-compute, re-classify</td>
</tr>
<tr>
<td>- Generalization</td>
<td>generalize, abstract</td>
</tr>
<tr>
<td>- Overlay (spatial join)</td>
<td>overlay, superimpose</td>
</tr>
<tr>
<td><strong>Focal Operations</strong></td>
<td></td>
</tr>
<tr>
<td>- Neighborhood</td>
<td></td>
</tr>
<tr>
<td>- window &amp; point queries</td>
<td>disjoint, zoom-out, point-in-polygon</td>
</tr>
<tr>
<td>- topological</td>
<td>meet, equal, contains, inside, covers, overlap</td>
</tr>
<tr>
<td>- direction</td>
<td>north, north-east, weak-bounded-north, same-level</td>
</tr>
<tr>
<td>- metric (distance) &amp; buffer zones</td>
<td>near, about, buffer, corridor</td>
</tr>
<tr>
<td>- nearest neighbor</td>
<td>nearest-neighbor, k-nearest-neighbors</td>
</tr>
<tr>
<td>- Interpolation</td>
<td></td>
</tr>
<tr>
<td>- location properties</td>
<td>point-linear, (inverse) distance-weighted</td>
</tr>
<tr>
<td>- thiessen polygons</td>
<td>thiessen-polygons, voronoi-diagrams</td>
</tr>
<tr>
<td>- Surfacial</td>
<td></td>
</tr>
<tr>
<td>- visualization</td>
<td>contours, TINs</td>
</tr>
<tr>
<td>- location properties</td>
<td>height, slope, aspect, gradient</td>
</tr>
<tr>
<td><strong>Connectivity</strong></td>
<td></td>
</tr>
<tr>
<td>- routing &amp; allocation (network)</td>
<td>optimum-path-finding, optimum-routing, spread, seek</td>
</tr>
<tr>
<td>- intervisibility</td>
<td>visible, light-of-sight, viewed, perspective, illumination</td>
</tr>
<tr>
<td><strong>Zonal Operations</strong></td>
<td></td>
</tr>
<tr>
<td>- Mask queries (spatial selection)</td>
<td>select-from-where, retrieve</td>
</tr>
<tr>
<td>- Measurement</td>
<td>distance, area, perimeter, volume</td>
</tr>
</tbody>
</table>
3. Site Selection Based on a Sequence of GIS Operations

The purpose of this Section is to present a sequence of data-interpretation operations which may compose one or more procedures to accomplish the task of site selection for a residential housing development [Stefanakis and Sellis, 1996]. The basic approach to this is to create a set of constraints, which restrict the planned activity, and a set of opportunities, which are conducive to the activity. The combination of the two is considered in order to find the best locations.

In the simplified situation that follows the set of constraints and opportunities consists of: a) vacant area (i.e., no development), b) dry land, c) level and smooth site (e.g., slope < 10%), d) nearness to the existing road network, and e) south-facing slope. In addition all candidate sites should have an adequate size to satisfy the needs of the planning activity (e.g., between 1 and 1.5 sq km). The whole task requires as input three layers of the region under examination: 1) hypsography layer: the three-dimensional surface of the region (altitude values); 2) development layer: it depicts the existing infrastructure of the region (e.g., roads, buildings, etc.); and 3) moisture layer: it depicts the soil moisture of the region (e.g., lakes, wet-lands, dry-lands, etc.).

The procedure of site selection, based on the sets of constraints and opportunities determined above, may consist of the sequence of operations shown in Figure 1. In that Figure rectangles represent operations (1st row: operation class; 2nd row: operation subclass), while parallelograms their operand (on the left side) and product (on the right side) layers.

![Figure 1. Site selection for a residential housing development.](image-url)
4. Fuzzy Set Theory

In classical set theory (Boolean logic) an individual is a member or it is not a member of any given set. Specifically, the degree to which an individual observation $z$ is a member of a set $A$ is expressed by the membership function, $MF^A$, which can take the value 0 or 1, i.e., $MF^A(z) = 1$, if $b_1 \leq z \leq b_2$; and $MF^A(z) = 0$, if $z < b_1$ or $z > b_2$, where $b_1$ and $b_2$ define the exact boundaries of set $A$. For instance, if the boundaries between “gentle”, “moderate” and “steep” land were to be set at $b_1 = 30\%$ slope and $b_2 = 60\%$ slope, then the membership function defines all “moderate slope lands”. Notice that classical sets allow only binary membership functions (i.e., TRUE or FALSE).

Fuzzy set theory [Zadeh 1965, 1988] is an extension of the classical set theory. A fuzzy set $A$ is defined mathematically as follows: If $Z = \{ z \}$ denotes a space of objects, then the fuzzy set $A$ in $Z$ is the set of ordered pairs: $A = \{ z, MF^A(z) \}$, $z \in Z$, where the membership function $MF^A(z)$ is known as the “degree of membership (d.o.m.) of $z$ in $A$”. Usually, $MF^A(z)$ is a real number in the range $[0,1]$, where 0 indicates no-membership and 1 indicates full membership. Hence, $MF^A(z)$ of $z$ in $A$ specifies the extent to which $z$ can be regarded as belonging to set $A$. The choice of the membership function, i.e., its shape and form, is crucial and strongly affects the results derived by the decision-making process. In correspondence to classical set theory, two options are available for choosing the membership functions for fuzzy sets [Burrough 1996]: a) through an imposed “expert” model; and b) by a data driven multivariate procedure.

Fuzzy set theoretic operations [Zadeh 1965; Turksen 1991] provide the counterpart operations to those of classical set theory. In other words, logical operations with fuzzy sets are generalizations of usual Boolean algebra applied to observations that have partial membership of more than one set. The standard operations of union, intersection, and complement of fuzzy sets $A$ and $B$, defined over some domain $C$, create a new fuzzy set whose membership function is defined as:

\[
\text{Union: } MF_{A\cup B}(z) = \max\{ MF_A(z), MF_B(z) \}, \forall z \in C \\
\text{Intersection: } MF_{A\cap B}(z) = \min\{ MF_A(z), MF_B(z) \}, \forall z \in C \\
\text{Complement: } MF_{\complement A}(z) = 1 - MF_A(z), \forall z \in C
\]

Consider the classification of individual locations on a layer based on the slope values and a second classification based on the land moisture with the following lexical values: [dry, moderate, wet, water]. For each individual location $l$ (e.g., d.o.m. for level $= 0.8$ and d.o.m. for dry $= 0.4$) the d.o.m. value which provides an overall measure regarding: a) level ground and dry land is derived by: min$\{MF_{\text{level}}(l),MF_{\text{dry}}(l)\}$, (e.g., min$\{0.8,0.4\}$ =0.4); b) level ground or dry land is derived by: max$\{MF_{\text{level}}(l),MF_{\text{dry}}(l)\}$, (e.g., max$\{0.8,0.4\}$=0.8); and c) non-level ground is derived by: 1- $MF_{\text{level}}(l)$, (e.g., 1-0.8=0.2).

A problem that arises in this case is that only one of the participating d.o.m. values dominates by assigning its value to the whole decision criterion. In this way the contribution of the other d.o.m. values is eliminated. In the literature several other functions have been proposed for handling the logical operations on observations that have partial membership on more than one set. These functions are more complex to compute and interpret, however they provide more expressive and accurate results. One of those is the energy metric [Gupta et al., 1988], in which $k$ fuzzy sets $(A_1,A_2,…,A_k)$ defined over some domain $C$, create a new fuzzy set $E$ with membership function given by Eq. 4, where $q$ is a positive integer. By applying this equation (e.g., for $q = 2$; quadratic measure) the big weight values (d.o.m.) are amplified, while the small values are nearly eliminated. Assuming the previous example, the overall measure characterizing each individual location ($l$) of a region, regarding level ground and dry land using the energy function, is given by Eq. 5. Notice that the energy measure derived by the previous formula should be normalized in the fuzzy domain $[0,1]$ [Gupta and Yamakawa, 1988].

\[
MF^E(z) = \sum_{i=1}^k [MF^A_i(z)]^q
\]

\[
MF_{\text{level-dry}}(l) = [MF_{\text{level}}(l)]^2 + [MF_{\text{dry}}(l)]^2
\]
Eq. 4 is more flexible and expressive than Eq. 2, because its overall measure is explicitly affected by the d.o.m. values of individual locations on both fuzzy sets A and B. Specifically, assume two individual locations $L_1$ and $L_2$, with the following d.o.m. values: $L_1$ (d.o.m. for level = 0.8 and d.o.m. for dry = 0.4) and $L_2$ (d.o.m. for level = 0.6 and d.o.m. for dry = 0.4). The overall measure provided by Eq. 2 is 0.4 for both locations, while that derived by Eq. 5 is 0.80 for $L_1$ and 0.56 for $L_2$. Clearly the energy metric provides an ordering of the two locations. It says that $L_1$ satisfies better the two criteria posed by decision-makers (i.e., level and dry land). This feature of energy metric is very beneficial for decision criteria which combine multiple sets and lexical values, while ordering of the qualified entities (i.e., individual locations) is required (e.g., find the five most level and dry locations of a region).

Apparently, in fuzzy logic methodologies, contrary to traditional logic, reasoning is based on a “late and flexible classification”, and consequently several of the problems caused by the employment of the latter are overcome. The following Sections examine: a) how uncertainty, which characterizes geographic entities, may be incorporated into a spatial data model (representation issue); and b) how fuzzy logic methodologies may be incorporated into the basic data interpretation operations for geographic data handling (reasoning issue).

5. Representation Issue: Extending the Data Model

The purpose of this Section is to show how fuzziness, which models uncertainty related to geographic data, may be incorporated into the general spatial data model presented in Section 2. In that model individual locations constitute the basic entities and attribute values regarding various themes (layers) are assigned to them. Each theme is described through a set of attribute values and each individual location on it is assigned only one of these values. The assignment of an attribute value to an individual location indicates its full membership regarding this feature in the corresponding layer.

In fuzzy set theory the concept of full membership is replaced by that of partial membership and consequently the representation of individual locations should change. The incorporation of fuzziness into the spatial data model forces the redefinition of the components forming the hierarchical data model. Specifically, while in conventional set theory the individual locations in a layer are assigned the attribute values (e.g., soil, grass, fruit-trees, forest) characterizing a theme (e.g., vegetation), in fuzzy set theory they are assigned d.o.m. values regarding each attribute value (e.g., 0.1 for soil, 0.6 for grass, 0.3 for fruit-trees and 0 for forest) characterizing a theme (e.g., vegetation). These values are derived by applying both the appropriate membership functions chosen by decision-makers and the knowledge provided by the experts. Hence, earth measurements and results derived from sampling techniques are processed and transformed into d.o.m. values for the predefined attribute (lexical) values characterizing a theme. Apparently, the number of layers increases, since each theme is represented by as many layers as the number of attribute values associated to it. The concept of zone is also changing in the fuzzy representation. Specifically, a fuzzy zone is defined as the sets of individual locations with a d.o.m. value greater than zero regarding a specific attribute value characterizing a theme. For instance, a forest area (road) consists of all individual locations that are assigned a non-zero d.o.m. value in the layer “forest” (“road”) characterizing the theme “vegetation” (“development”). Hence, all individual locations in a layer with non-zero d.o.m. values constitute a fuzzy zone.

6. Reasoning Issue: Extending the Operations

Following the incorporation of fuzziness into the spatial data model, the reasoning issue should be examined. This involves the incorporation of fuzzy logic methodologies into the basic data interpretation operations available in GIS packages [Openshaw, 1991; Stefanakis et al., 1999, 1996; Stefanakis and Sellis 1997; Stefanakis
Adopting the classification of these operations, as presented in Section 2, it is shown how the individual operations may be extended to support fuzzy spatial reasoning. Specifically, after a redefinition of the basic classes of data interpretation operations, one representative operation for each class is examined, accompanied with examples that commonly appear in spatial decision-making process. The three classes of data interpretation operations are redefined as follows in order to incorporate fuzziness:

- **Fuzzy local operations**: they include those that compute new fuzzy values for each individual location on a layer as a fuzzy function of existing fuzzy data explicitly associated with that location; e.g., fuzzy overlay operation.

- **Fuzzy focal operations**: they compute new fuzzy values for every individual location as a fuzzy function of its neighborhood; e.g., fuzzy distance operation; and

- **Fuzzy zonal operations**: they include those that compute new fuzzy values for each individual location as a fuzzy function of existing fuzzy values associated with a fuzzy zone containing that location; e.g., fuzzy select operation.

### 6.1 Fuzzy Overlay Operation

The overlay operation is analogous to join operation in conventional database systems, and is defined as the assignment of new attribute values to individual locations resulting from the combination of two or more layers. The fuzzy overlay operation takes a more general form and is defined as the computation and assignment of an overall measure (d.o.m. value) to each individual location, which is derived from the consideration of d.o.m. values on two or more layers and the execution of appropriate fuzzy operation(s). The overall measure is also expressed in the fuzzy domain $[0,1]$.

Consider the situation where all individual locations that belong to forest areas with a steep ground slope are searched. In conventional overlay operation the input layers are: a) “vegetation” layer, and b) “ground slope” layer. The product is the “forest & steep” layer, whose individual locations are assigned the value of 1, if they satisfy both constraints, or 0 otherwise. On the other hand, in fuzzy overlay operation the input layers are: a) “forest” layer, and b) “steep ground” layer. In both layers individual locations are assigned the corresponding d.o.m. values for forest and steep ground, respectively. The product is the “forest & steep” layer, whose individual locations are assigned new d.o.m. values, characterizing their membership in the new fuzzy set, and derived by applying the appropriate fuzzy operation (Eq. 2 or 4) and normalizing the result.

### 6.2 Fuzzy Distance Operation

A distance metric is usually required in order to analyze the spatial relationships between entities in GIS. Several metrics are available [Preparata and Samos, 1985; Gatrell, 1991]. The choice depends on the application domain. Two cases of fuzzy distance are examined next. They show how individual locations on a layer are classified based on their distance from: a) a given location, or b) a given fuzzy zone.

In order to characterize an individual location $X$ based on its distance from a given location $L$ the following procedure is executed. First, the distance $d$ (e.g., Euclidean) from $L$ to $X$ is computed. Then a membership function is chosen to transform distances into d.o.m. values on the predefined attribute (lexical) values characterizing the theme “nearness” (e.g., close, near, moderate, far, far away). Finally, the distance from $L$ to $X$ is transformed into d.o.m. values. Hence, the product of the fuzzy distance operation consists of a set of layers and each layer accommodates the d.o.m. values regarding a specific attribute value (i.e., close, near, moderate, far, far away) characterizing the theme “nearness to location $L$”. On the other hand, in order to characterize an individual location $X$ based on its distance from a given fuzzy zone $Z$, which consists of a set of individual locations $\{L_1, L_2, \ldots, L_n\}$ with different d.o.m. values in the fuzzy zone, the following procedure is executed.
First, the distances \( d_i \) from all \( L_i \) \((i = 1, 2, \ldots, n)\) to \( X \) are computed and transformed into d.o.m. values on the predefined attribute (lexical) values characterizing the theme “nearness” (e.g., close, near, moderate, far, far away). That is, for each attribute value \( A \), the individual location \( X \) is assigned a set of pairs \((MF^X_A(X), MF^X_A(L_i))\) (for \( i = 1, 2, \ldots, n \)), where \( MF^X_A(X) \) is the d.o.m. value for \( A \) characterizing the theme “nearness”, and \( MF^X_A(L_i) \) is the d.o.m. value of location \( L_i \) in fuzzy zone \( Z \). Finally, a fuzzy function, chosen by the experts, is applied in order to map the set of pairs into single d.o.m. values (i.e., overall measures) on the predefined attribute (lexical) values characterizing the theme “nearness” (e.g., close, near, moderate, far, far away). Apparently, the product of the fuzzy distance operation consists of a set of layers and each layer accommodates the d.o.m. values regarding a specific attribute value (i.e., close, near, moderate, far, far away) characterizing the theme “nearness to fuzzy zone \( Z \”).

Some rather common queries where fuzzy distance operation is applied, in combination with fuzzy select operation, are: “find all areas that are near to the existing road network”, “find all areas that are far from schools”, etc. Notice that similarly to fuzzy distance operation, other focal operations, such as fuzzy direction (with example lexical values: north, east, south, west) and fuzzy topological operations (with example lexical values: disjoint, overlap), may be defined.

### 6.3 Fuzzy Select Operation

The scope of fuzzy select operation is to highlight individual locations on a layer based on their d.o.m. values regarding a single or composite attribute characterizing one or a combination of layers. Hence, depending on the constraints posed by the query, the fuzzy select operation may highlight: a) those individual locations that have a d.o.m. value within a range of predefined threshold values; and b) the \( n \)-individual locations that are superior to others based on their d.o.m. values (notion of ordering).

### 7. Site Selection and Fuzzy Reasoning

Focusing on the simplified example of site selection for a residential housing development, presented in Section 3, some of the advantages provided by the fuzzy data interpretation operations may be highlighted [Stefanakis et al., 1996; Stefanakis, 1997]. For the constraints posed by decision-makers the following lexical values could be considered: a) ground slope: \{level, gentle, moderate, steep\}, b) development: \{vacant, semi-developed, developed\}, c) soil moisture: \{dry, moderate, wet, water\}, d) accessibility: \{close, near, moderate, far, far away\}, and e) aspect: \{north, east, south, west\}.

Transformation functions will be adopted to map ground measurements to d.o.m. values characterizing individual locations of the area under study. By performing a fuzzy classification one layer of d.o.m. will be generated for each lexical value characterizing a theme. Examining the layers which correspond to the lexical values of interest (i.e., vacant, dry, level, near, south) a fuzzy overlay will provide a new layer which classify all individual location of the area under study based on the degree they fulfill the constraints posed by decision-makers. A fuzzy select operation will highlight the best sites for the planning activity. Figure 2 illustrates the procedure. Obviously, in this scheme, contrary to conventional way, reasoning is based on a late and flexible classification with several advantages as shown in Section 4.
Steps 1 to 5: generate layers with d.o.m. values for vacant, dry, level, south facing, near to highways areas

Step 6 and 7: generate a layer of good sites and isolate those with adequate size for the planning activity.

**Figure 2.** Site selection for a residential housing development.

**8. Conclusion**

Fuzzy set methodologies seem to be instrumental in the design of efficient tools to support the spatial decision-making process. This paper examines the incorporation of these methodologies into a DBMS repository for the application domain of GIS. It is shown how the useful concepts of fuzzy set theory may be adopted for the representation and analysis of geographic data, whose uncertainty is an inherent characteristic. The contribution of the paper can be summarized as follows: After a short introduction to the concept of uncertainty in geographic information and the negative effects of the employment of classical set theory in both the representation and analysis of geographic data, the alternative logical foundation provided by the fuzzy set theory is presented and its advantages over the standard logic are highlighted. A general spatial data model is extended to accommodate the uncertainty of geographic entities and it is shown how the standard data interpretation operations may be extended to support fuzzy spatial reasoning. Several simplified real world examples are given while a real world situation involving spatial decision making is examined. Future research in the area includes the design and implementation of a prototype DBMS repository for the application domain of GIS with fuzzy set methodologies incorporated in both the representation and analysis of geographic data; as well as an extensive theoretical and experimental study on the choice of the appropriate membership functions to simulate physical phenomena and fuzzy operations for the set of constraints posed by decision-makers (experts) in real world applications.

**Acknowledgments**

Work supported by the European Commission funded TMR project CHOROCHRONOS.
References


A Model-based Approach to Geological Map Representation

Boyan Brodaric
Geological Survey of Canada
234B-615 Booth St., Ottawa, Ontario, K1A 0E9
brodaric@gsc.nrcan.gc.ca.

Abstract
Digital geologic maps have been traditionally constructed from the cartographic or feature-based points of view. Cartographic representations typically focus on a map’s geometric entities and permit symbolization and semantic content to be associated with each entity; i.e. a geologic map is a collection of symbolized and attributed points, lines, polygons, etc. Feature-based representations invert this emphasis by prioritizing the semantic concept over its geometry and cartographic appearance: i.e. a geologic map is a collection of categorized geologic features that contain geometric, symbol and semantic attributes. In each case the relationship between entities and between features is primarily spatial, understating the potential for complex semantic and temporal relations to exist. Traditional paper geologic maps describe these and other relations, as well as feature content, in the map’s surrounding text and legend, and it is clear that digital representations must capture the same knowledge for effective geologic map usage. In capturing this knowledge the interrelated nature of a geologic map is affirmed, and the author’s explanatory model is fully expressed. However, in this model-based view, a geologic map is largely a conceptual construct that can be understood only if the underlying scientific framework is comprehended. That comprehension is possible within the geologic community implies the existence of a common cognitive schema that is shared by geologists, and that is typically expressed in map, text, and oral form. The digital representation of this requires the cognitive schema itself to be modeled. To accomplish this, prevalent geographic space-time models must in turn enhance their semantic components and model the nature of a scientific premise. An approach to this was developed and implemented in the context of prototype digital geologic map standards at the Geological Survey of Canada and the U.S. Geological Survey.

1. Introduction

The explosion in the availability of digital information in all parts of society is placing intense pressure on scientific agencies to make their knowledge digital. This is occurring in tandem with a changing socio-political climate in which public sentiment now openly demands science to be more aligned with broader societal issues on economic, environmental, political, and other fronts. For many science agencies meeting this societal imperative, and its loosely associated digital challenge, is a daunting task that strains entrenched operational frameworks, attitudes and even mandates. Science agencies are becoming more outward looking as the need to demonstrate relevance to non-experts becomes more apparent, and the need to integrate research across disciplinary boundaries and within diverse scientific teams becomes compelling. At the same time the prevalence of computing technology is affecting how science is undertaken and even conceptualized. The combination of ever-improving digital technology and its penetration into every aspect of our lives, is seducing and coercing science agencies to adopt digital approaches to corporate knowledge management.

Geological mapping organizations are not immune to these forces, and they are responding with vigor. However, their ability to harness the full potential of digital technology is hampered by the lack of a rigorous description and formal understanding (for computing purposes) of geologic knowledge. A geologist’s knowledge
is traditionally bundled into maps, reports, charts and diagrams, where the subjective nature of their contents can be often be viewed as degree of opinion as much as fact [Burrough, 1992]. The geologic map serves a unique role in this knowledge milieu as it has over time developed into a cartographic synthesis of an author’s geological understanding of a geographic region. Digital map representation efforts that are rooted in corporate databases [POSC; PPDM; Ryburn and O’Donnell, 1998], map production [Laxton and Becken, 1996], or map data models [Bain and Giles, 1997], capture much of this. However, there is scope to improve the representation of the scientific premise behind a map, which would lead to enhanced knowledge representation and more sophisticated map usage. As disparately conceived geologic data flood the Internet, and as pressing societal dictates demand more conceptual accessibility to, not just physical availability of, geologic map data, explicating scientific premise becomes a priority.

This paper presents an approach to incorporating geological premise within digital geologic maps. The approach is described firstly, from the geologic viewpoint, secondly from the viewpoint of cartographic theory, and thirdly, from the perspective of geographic representation. Lastly, an implementation derived from work on U.S. and Canadian digital geologic map standards [Johnson et. al., 1998] is presented.

2. The Geological Map

The traditional geologic map contains measurements, descriptive observations and inferred geologic phenomenon, symbolized according to a loose but time-honored tradition. The categories (e.g. geological units) ascribed to the geologic phenomenon are described in the legend and surrounding text. Spatial relations are depicted on the map face, and spatio-temporal relations between geologic categories, and even individual phenomena, are described textually and pictorially in the map’s legend and surround. A geological history for the area is constructed by the map reader who must synthesize the legend, surround, and the map face through a filter of general geologic heuristics (e.g. Walthers law of superposition) and personal experience. The result is an understanding of the four-dimensional geologic evolution of a specific geographic region. Not every geological map reaches this ideal as original purpose, or a surfeit of data or resources, prevent a full geological explanation to be constructed. The map then takes on a descriptive rather than explanatory role.

3. Cartographic Representation

The triadic sign model of semiotics [MacEachren, 1995, p. 245] describes how map elements are understood through the relation of map sign, referent (actual phenomenon) and concept (interpretations). Emphasizing specific aspects of this relation suggests approaches to mental category formation [Nyerges, 1991] that variably suit geologic map representation.

Positioning the map sign as the primary vehicle for map element understanding (Figure 1), binds sign understanding more to the inherent visual competency of the sign than to its interpretation. As a result, some signs will be more effective than others, leading to a search for optimal signs. The existence of concepts and referents is not questioned in this, and it only remains for the sign to draw them together. Nor is the strength of the concept-referent pair questioned, for if a sign could equally associate several concepts with a referent then its interpretive value is lost; if it could link several referents to a concept, then its spatial validity would be suspect. Which is not to say that misunderstanding is impossible, but only that it results from poor sign selection rather than from variation or inconsistency in how phenomena are interpreted.

Tight concept-referent pairings correspond to classical theories of mental category formation in which phenomena (referents) possess inherent meaning [MacEachren, 1995, p. 151]. Differences in categorizing an object can then be attributed to faulty sensory mechanisms, or in the realm of computer maps, to syntactic diversity where the concept has many names. Resolving the latter leads to the development of a standard
catalogue of categories, a popular idea in many geographic standards efforts [Usery, 1993]. Likewise, strong sign-concept pairings enable the addition of standard map signs to such catalogues, causing map sign and category to be fused until they function indistinguishably – indeed many GIS operate in this framework. Fitting a geologic map into this sign-centric configuration is problematic in that tight referent-concept pairings disallow uncertainty, variation and error to be associated with the categorization of phenomena. Such discrepancies are only appropriate for the sign relations (sign-concept and sign-referent), to measure their optimal nature. Furthermore, strong sign-concept pairings, and the GIS that advocate them, discourage the re-use of signs for different scientific purpose. It is difficult to imagine geologic and geographic thought evolving in such environments, where phenomena, and the truth of their properties, are self-evident. Mapping activity is then demoted to simple perception (i.e. data collection), and map-making to a quest for visual efficacy, causing scientific discourse and progress in the geological mapping sciences to be discounted. Geographic representations that adopt a sign-centric approach are perhaps best suited to portraying experiential space [Smith and Mark, 1998] where the existence of concepts and referents, and their unique relation, is not overly suspect.

A concept-centric semiotic formulation that has mental concept mediating map sign and real world referent (Figure 2) is quite different. Meaning and understanding is then interpretative and thus cultural. Sign conventions may exist within cultures [MacEachren 1995, p. 261], but they are learned. Thus multiple referents and map signs may co-exist for any one concept, and this enables debate about concept validity and promotes scientific discourse. Improving map understanding in this framework requires the map to better explicate the author’s intentions, such that the author’s cognitive elements and synthetic processes are induced in the reader. Therefore the details outside the actual map face, such as the legend and explanatory notes as well as unwritten but assumed discipline-specific heuristics, become critical to understanding the map. Indeed these constitute the intensional knowledge [Nyerges, 1991], comprised of concepts, their interrelations and rules, that aid in the comprehension of the map’s extensional phenomenon (i.e. spatio-temporal objects and relations) depicted on the map face. Representational frameworks must model both the intensional and extensional aspects to convey and contain scientific meaning.

**Figure 1.** Sign mediating concept and referent and implied objects of cognitive representation.

![Figure 1](image1)

**Figure 2.** Concept mediating sign and referent Derived from MacEachren [1995, p. 246].
To better represent the author’s categorization process, the concept relations (referent-concept, concept-sign) must be further explored. Applying direction to these relations yields the activities of mapping and map-making (referentàconceptàsignàreferent), as well as map reading (signàconceptàreferentàsign). Mapping consists of sensing and interpreting phenomenon. This can be seen as a gradient from lesser to greater abstraction along the referent-concept link, involving perception (description), spatio-temporal relations, categorization, and hypothesis and geographic model construction, and resulting in a holistic understanding that is neatly contained in the mapper’s mind. Map-making consists of the expression of this understanding, firstly, through the concept-sign link by establishing a legend through the symbolization of categories, and secondly, through the referent-sign link, where the symbolized phenomenon are spatially distributed on the map face. In map reading, the opposite sequence is followed in that map signs and concepts are understood via a legend, which clarifies the referents. These correspondences are depicted in Figure 2.

The importance of this discussion lies in the correspondence between the elements of the concept-centric semiotic model and the elements of the geologic mapping and map-making process. Because the concept-centric formulation embraces scientific category formation and its visual expression on a map, its seems logical to use its elements to form a representation framework for digital geologic maps. The critical elements are: phenomenon (referents), perceptions/descriptions (referent-concept link), hypothesis and models (referent-concept link), categories (concepts), symbols (signs), and relations (spatial, temporal, causal, and semantic). These must be coordinated within a geographic representation.

4. Geographic Representation

Layers

Pursuing a cognitive approach to geological map representation is a departure from historical geographic representation strategies that are primarily experiential. For instance, layer-base approaches to GIS (Geographic Information Systems) depict geographic space as a set of thematic layers that are related only through spatial overlap. Layers consist of geometric entities (e.g. points, lines and polygons) with uncontested shape, position or theme. Occupants of a geographic space could belong to multiple themes, possess multiple shapes and positions by existing on more than one layer, but they could not easily be united into single features whose identity reaches across layers. The inability to form intra-layer relationships reinforces the spatial thrust of the layer (i.e. its geographic position and relations) and de-emphasizes its semantic aspect (i.e. its meaning), thereby reducing the ability to represent interpretive ‘large-scale’ information. In terms of the geologic map this discourages a full explanation of the geological categories that are typically found in the legend, elaborated in map surround, and which provide ‘large-scale’ context.

Features

More recent feature-based approaches overcome some of these problems by representing geographic reality as a collection of phenomenon instead of layers [Usery, 1996a]. Features typically posses [after Tang, 1992]:

- **State**: properties ascribed to the feature, normally described as space, time and theme aspects.
- **Identity**: a unique identifier maintained through changes of state.
- **Relations**: semantic, spatial and temporal relationships between features (to form aggregates) and feature properties.
- **Behavior**: the rules of behavior and other functional characteristics of a feature.

Operationally, each of the space, time and theme axes may also consist of distinct subsystems capable of managing representational and functional traits [Tang, 1996; Usery, 1996b]. For instance, the spatial aspect of a feature may be most appropriately represented as a continuous field versus discrete object, and as a three
dimensional octtree instead of a two dimensional grid. Analogous strategies can be defined for theme and time (and other potential feature dimensions), thereby providing a holistic conception of a feature that equally weighs space, time and theme. This holistic approach is primarily semantic in its concern with the meaning of geographic phenomenon, in contrast to the layer-based approach that was predominantly space-centered [Usery, 1996a].

However, the cognitive aspect of the feature-based approach is often diluted by systems that do not permit a feature to simultaneously possess multiple values for its spatial, thematic or temporal parts. Cognitively this implies that phenomenon cannot be variously perceived or categorized. Contemporary category theory supports this in part, by postulating the existence of basic-level categories to which sensory perception naturally converges [Lakoff, 1987], confirming our notions of ‘experiential’ space. Most current geospatial standards efforts [OGC Technical Committee, 1998; TC/211, 1998] are rooted in this approach.

Changing the scale of attention from this experiential scale, to the microscopic or to the macroscopic ‘large-scale’, causes the resulting categories to be seen much more as products of laboratory process or human culture and intellect, respectively, than of primal sensory cognition. Notwithstanding the importance of laboratory results to the geological categorization process and the need to include them within a representation framework, at the scale of the geologic map it is the ‘large-scale’ categories that are the primary concern. These are fundamentally not basic-level, as they are an artifact of the author’s education, culture and personal experience, and are thus eminently debatable. This does not mean to suggest that geologic theory is inconsistent, only that it can be variably and selectively applied. For example, in measuring the 3d orientation of a rock body, two geologists may or may not agree on its disposition in space, or on its geological category, thereby generating more than one description and category for the phenomenon. This points to an explicit subdivision of theme into category and description. Indeed, it is sometimes more profitable to treat these, and other, feature components as unrelated disaggregates [Gray, 1997] combined by users at their time of need [Gahegan and Flack, 1996]. Various research suggests enhancing feature definition beyond the standard identity, space, time, theme, relation and behavior components:

- **Category and Description**: to replace theme, as discussed above.
- **Prototype**: an idealized template for a feature (i.e. its intension) [Lakoff, 1987].
- **History**: a record of a feature’s origin and evolution [Gahegan, 1996]; required for cognitive and changing features.
- **Process and Logic**: process determines a feature’s reaction to physics (external forces), and logic maintains internal constraints and a reasoning mechanism to apply them [Smyth, 1998, after Davis, 1990]. These replace behavior.

Thus, a feature is composed of identity and state: space, time, category, description, relations, behavior (process + logic), and history. Separating feature identity from state is often necessary where a feature’s state is debatable or changing and thus potentially multi-valued: e.g. changing river X’s course (in nature or in cartography) does not alter its identity as X. Multi-valued states draw attention to other fundamental issues related to the definition of a feature, which are not clearly differentiated. It is convenient to use (liberally) object-modeling terminology for this differentiation:

1. **Identity (feature composite)**: a phenomenon with one identity but multiple states; e.g. geological boundary Y is both a fault and a geological contact [Brodaric, 1998].
2. **Instance (feature instance)**: one state of a multi-state phenomenon: e.g. boundary Y* as fault or contact, but not both.
3. **Class (feature prototype)**: a prototype feature, usually possessing no extension in space or time, though possibly possessing spatio-temporal and other relations that define it: e.g. a mathematical definition of a coastal landform.
4. **Metaclass (category)**: a category. Categories are the most abstract type of feature and are intensional in
nature: e.g. a thrust fault defined in a geology text. They can be seen as classical concept [Nyerges, 1991] or scientific vocabulary.

**Feature composite:** \( F (i, \langle s \rangle, \langle t \rangle, \langle c \rangle, \langle d \rangle, \langle r \rangle, \langle b \rangle, \langle h \rangle) \).

**Feature instance:** \( F^* (i, \langle s^* \rangle, \langle t^* \rangle, \langle c^* \rangle, \langle d^* \rangle, \langle r^* \rangle, \langle b^* \rangle, \langle h \rangle) \). \( F^* \) represents one state of \( F \).

**Feature prototype:** \( P (i, \langle s \rangle, \langle t \rangle, \langle c \rangle, \langle d \rangle, \langle r \rangle, \langle b \rangle, \langle h \rangle) \). \( s, t \) are relative (relations).

**Category:** \( c (i, d, r, h) \).

\( i=\text{identity}, \ s=\text{space}, \ t=\text{time}, \ c=\text{category}, \ d=\text{description}, \ r=\text{relations}, \ b=\text{behavior}, \ \text{and} \ h=\text{history} \).

The distinctions between these often blur. For example, an instance of a feature may also be its prototype class—in geology this corresponds to a definitive type locality. Note the increase in abstractness in the sequence of feature types—composite, instance, prototype, and category—as each becomes progressively less associated with a specific geographic phenomenon and more with a specific mental category—this neatly coincides with the transition from referent to concept in the semiotic sign model.

**Models**

Features with multiple states are problematic in that their diverse content undermines the coherent understanding of a geographic region. Their discordant elements must be purposefully filtered out, and the remaining elements appropriately assembled, for understanding to occur. This corresponds to a holistic view of geographic understanding [Harvey, 1997] and scientific thinking [Ahl and Allen, 1996], and suggests that representation frameworks must aggregate features into coherent compound structures to capture the geographic totality of a region. These compound structures are referred to as models [Zeigler 1990] reflecting their conceptual integrity and abstract nature. Moreover, representational structures that can host not one model, but multiple and diverse scientific models, are attractive as they enable model-level operations to occur: e.g. the evaluation of scientific models against a data pool. As models encapsulate scientific theory they are critical to scientific discovery activities such as visualization, which build and refine scientific theory. However, model-based systems in GIS [Bennet, 1997; Raper and Livingstone, 1995] are not explicitly cognitive. Differentiating the definition of a model according to cognitive feature type yields definitions that provide greater clarity in terms of scientific cognition and map representation:

1. **Dataset:** a set of unfiltered feature composites; e.g. a database of observations.
2. **Model:** an integral assembly of feature instances—a geographic model; e.g. a 3-d ore body in a GIS.
3. **Hypothesis:** an integral set of prototype features. This permits datasets to be tested against hypotheses for scientific fit, and once achieved, would signify a new model. A symbolized hypothesis defines a legend, and a legend is thus fundamentally tied to scientific premise.
4. **Taxonomy:** an arrangement of categories, often hierarchical (i.e. partially ordered); e.g. a rock type classification system.

\[
\begin{align*}
\text{dataset} & : \langle F, h \rangle. \\
\text{model} & : \langle F^*, h \rangle. \\
\text{hypothesis} & : \langle P, h \rangle. \\
\text{legend} & : \langle P, h, m \rangle, \ m \neq \text{null}, \ m \text{ denotes a cartographic mark (symbol).} \\
\text{taxonomy} & : \langle c, h \rangle.
\end{align*}
\]

These distinctions lead to two powerful (for computing) definitions, one conceptual and the other cartographic:

- **Explanatory model:** is a model that possesses a hypothesis and draws its concepts from a taxonomy.
- **Geologic map:** is an explanatory model that possesses a legend (i.e. its hypothesis is symbolized).

These form the basis for a cognitive representation scheme for a geologic map:

\[
\text{EXPLANATORY MODEL (model, hypothesis, taxonomy, h)} \\
\text{MAP (EXPLANATORY MODEL (model, legend, taxonomy), h)}
\]
Table 1. A summary of the components of the model-based approach to geologic map representation, their definitions, and their relation to the semiotic triadic sign model where concept mediates sign and referent.

<table>
<thead>
<tr>
<th>Model-based object</th>
<th>Definition</th>
<th>Semiotic correspondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite Feature</td>
<td>A feature with possibly multiple states.</td>
<td>Referent</td>
</tr>
<tr>
<td>Instance Feature</td>
<td>A feature with one state, or a single-state view of a composite feature.</td>
<td>Referent-Concept</td>
</tr>
<tr>
<td>Prototype</td>
<td>A general scientific term that could classify a feature.</td>
<td>Concept</td>
</tr>
<tr>
<td>Dataset</td>
<td>A collection of feature composites.</td>
<td>Referent</td>
</tr>
<tr>
<td>Model</td>
<td>A set of feature instances.</td>
<td>Referent-Concept</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>A set of feature prototypes.</td>
<td>Referent-Concept</td>
</tr>
<tr>
<td>Taxonomy</td>
<td>A set (partially ordered) of categories; a classification system(s).</td>
<td>Concept</td>
</tr>
<tr>
<td>Explanatory Model</td>
<td>A scientific explanation = model + hypothesis + taxonomy.</td>
<td>Referent-Concept</td>
</tr>
<tr>
<td>Legend</td>
<td>A symbolized hypothesis.</td>
<td>Concept-Sign</td>
</tr>
<tr>
<td>Map</td>
<td>An explanatory model with a legend.</td>
<td>Referent-Concept-Sign</td>
</tr>
</tbody>
</table>

5. Implementation

The model-based representation framework is a superset of the national digital geologic map data model prototype [Johnson et. al., 1998] of the USGS (United States Geological Survey) and the AASG (American Association of State Geologists). Though the USGS-AASG effort is primarily concerned with defining a database design for geologic map content, and the model-based approach is a broader and more theoretical treatment of the same issues, they are in the main congruent. They differ in three main respects: database systems do not support dynamic feature behavior, the temporal part is reduced, and the database design does not differentiate map from model. Figure 3 depicts the model-based approach as a data model.

Figure 3. A data model view of the model-based approach to geologic map representation. Note that Taxonomy is embodied by Category.
A large database of geologic maps for the Canadian Coordillera [Brodaric et. al., 1999] was constructed according to the USGS-AASG data model design. The database represents a digital library that contains six main geologic map series ranging in scale from 1:250,000 to 1:32,000,000 and covering a large geographic area including all of British Columbia, the Yukon, and parts of Alberta. Over 210,000 individual feature instances are contained in the library, as well as about 3500 prototype features. Each prototype feature is described according to absolute age range (e.g. 300-400 million years), geologic age range (Devonian-Silurian), rock type composition (shale, siltstone, etc.), and is also linked to a cartographic symbol within a legend. The definitive Decade of Geology of North America [Gabrielse and Yorath, 1992] was converted to digital format and compartmentalized, with text components indexed according to their geologic age and subject, and related to relevant figures and images, and to prototype features. Users are thus able to recall authoritative descriptions, including text and images, when viewing any feature instance on a map; users are also able to enter the archive from the text and image perspectives, with all interrelationships maintained. The library is constructed to operate in an Internet environment, by utilizing the Autodesk MapGuide software to display and contain its spatial aspects; the remaining (semantic) aspects of the database are hosted within the MS SQL Server relational database environment. The web site will be publicly available by the time of this publication, though its definitive web address is unknown at the time of writing. It is anticipated that the site will enhance geological research and geological education through the digital interconnection of the various map, text and image components.

6. Discussion

Ideal implementations of the model-based approach should host multiple scientific (cognitive) models and maps. Certain functional power is gained from this as model-level operations that compare, contrast and merge models into new formulations can then be defined and applied. In a cartographic sense these operations can be viewed as map transformations [Tobler, 1979; Moellering, 1991]. The geological application of this is quite attractive as it suggests mechanisms that first conceptually transform scientific concepts, and secondly, display them for visual presentation as a map. Indeed, conceptual transformations are the underpinning to model-based generalization [Muller et. al., 1995], which strives to generalize features on a complete semantic basis, as opposed to the spatial bias of cartographic generalization. Semantic transformations would also be required for the harmonization of disparately conceived and symbolized map products, found, for example, on the various Internet sites. Defining geologic transformations that enhance map utility is an important future challenge.

A related issue is the compound model, or the embedding of a model within another model (i.e. a map within a map), when the scale of scientific focus has become more general. It is convenient in these instances to replace the embedded model with a feature that summarizes the underlying model while maintaining knowledge of it. Thus a model is replaced by a feature instance, a hypothesis by a feature prototype, and a taxonomy by a category. This not only preserves the integrity of the previous definitions, but is also in line with work in the fields of modular simulation [Ziegler, 1990] and the philosophy of science [Ahl and Allen, 1996], where the scale of endeavor justifies the more general concepts and summaries of the replacement features. With the addition of semantic transformations to facilitate this change of scale, it is possible to envision map databases that are able to dynamically generate maps at generalized resolutions, while maintaining awareness of their origins. The impact of this on geological map providers would be enormous, as it would eliminate the need to divert resources to the creation of regional scale maps at regular time intervals, and would instead enable the immediate (or timely) integration of detailed maps into regional products [Colman-Sadd, 1996; Haugerud, 1997].

Lastly, it is possible to imagine using the legend (as defined herein) as an on-screen interface to the building and updating of scientific models during visual exploration, dynamic simulation, or geologic map modification.
7. Conclusions and Future Directions

A model-based approach to geologic map representation was presented in the form of abstract definitions and data model diagram. The approach was influenced firstly by cartographic theory, and secondly by geographic representation methods. Cartographic theory provided a functional model for map understanding, and suggested the basic objects for cognitive map representation. These were further explored within geographic representation strategies, with emphasis placed on the feature-based approach. This was extended to accommodate a more holistic representation of geographic reality, dominated by human perception and purpose and which is essentially model-driven. Definitions for cartographic components, such as maps and legends, were framed in a scientific thinking and general cognitive context, engaging the notions of hypothesis, model and taxonomy. The approach was implemented, somewhat simplified, in a relational GIS environment and upon a large dataset of geologic maps and substantial related information. The general applicability of the model-based approach was confirmed, but there is much left to do.

The translation of abstract definitions into a computable structure is a data modeling exercise. Most contemporary data modeling languages describe an aspect of reality in three dimensions: as objects, as functions and as dynamic processes: “the functional model specifies what happens, the dynamic model specifies when it happens, and the object model specifies what it happens to” [Rumbaugh et. al., 1991, p. 123]. In this discussion, cartographic theory provided the functional model and suggested the object model, which was completed using geographic representation theory. The temporal model, describing how objects (feature, model, map) change over time, is beyond the intention and scope of this discussion. It is important to note, however, that the model-based approach does emphasize a multi-state description of a feature that includes temporality and a subsystem for its management. It is thus able to represent the effects of events and processes on a feature, through the feature’s states. Indeed, the definition of a feature includes a behavior element, that, though underutilized to date, could provide the basis for a process component; it has been suggested that process should be another state dimension [Cheng and Molenaar, 1998]. However, the current approach is largely phenomenon-centric and ignores the actual processes acting on a feature, as well as the lines of causality between a feature and process. It is more suited to portraying a static display of a dynamic earth system, as per a traditional geologic map, and is not suited to the representation of an animated geologic map. Apart from such representational issues there are several outstanding functional issues that remain unresolved. Paramount among these is the definition of semantic transformations to enable the generalization and harmonization of disparate maps. This will require the tight integration of a reasoning engine into the framework, and much more work on geological map reasoning [Harrap and Helmstaed, 1998].

8. References


Maintaining Parallel Map-Datasets

Peter Højholt
National Survey & Cadastre
Rentemestervej 8, DK-2400 København N, Denmark
phs@kms.dk

Abstract
A set of concepts has been developed which can help to define and communicate problems which relate to the establishment, updating and analysis of parallel or mutually dependent maps. The set of concepts supports an organisation and formalisation of processes the management of which is currently left to the individual user to a great degree. The set of concepts is a first attempt, and further development and specification will be necessary.

Updating and maintaining parallel map systems
In practical situations, the preparation of several maps which cover the same physical area is often required. The maps often relate to one another. In these situations, there are only limited possibilities of finding guidance in the literature, and there exists no well-defined and commonly used set of concepts to specify and discuss these situations.

Map databases differ from other databases in that geometrical concepts play a major role. The definitions below relate primarily to geometrical concepts. Maps may contain a lot of other information but for maps there is a special need to be able to specify geometrical conditions.

The following is a first attempt to establish such a set of concepts. The author hopes that these introductory exercises may inspire others to more work on the concepts so that they become more precise and new rules can be derived. Some examples give an outline of how the new concepts may be used.

Words for which no definition is attempted

Reality
Is that which is to be described. It may be a landscape, the mess on my desk or whatever.

Geometrically relatable objects
Geometrically relatable objects are objects which, at the times at which they are observed, can be related to a geometrical space.

Geometrically related representation
A geometrically related representation is a representation of reality in which common geometrical concepts (such as distance) may be defined.
Map specifications

Map specification (subsequently just specification)

A specification describes how to form a geometrically related representation on the basis of physical or abstract, geometrically relatable objects or one or more other geometrically related representations observed in both cases at a random or specified time $T$ and possibly at other times $T'$ defined on the basis of $T$.

Collection of specifications

A collection of specifications consists of one or more specifications.

Defined relation to reality

A specification is said to contain a defined relation to reality if, in its description of how to form the geometrical representation, consideration must be shown for anything other than geometrical representations formed on the basis of other collections of specifications.

Defined relation between collections of specifications

A collection of specifications $A$ has a defined relation to another collection of specifications $B$ if collection $A$ describes that, in connection with the formation of the geometrical representations, consideration must be shown for geometrical representations formed on the basis of collection $B$. It must be possible to express a defined relation explicitly and it must be possible to decide objectively whether collection $A$ complies with its relation to collection $B$.

Independent collections of specifications

A collection of specifications $A$ is said to be independent of another collection of specifications $B$ if collection $A$ has no defined relation to collection $B$.

Mutually independent collections of specifications

Two collections of specifications $A$ and $B$ are said to be mutually independent if collection $A$ is independent of collection $B$ and collection $B$ is independent of collection $A$.

Dependent collection of specifications

A collection of specifications $A$ is said to be dependent on another collection of specifications $B$ if collection $A$ has at least one defined relation to collection $B$.

Derived collection of specifications

A collection of specifications $A$ is said to be derived from another collection of specifications $B$ if collection $A$ is dependent on collection $B$ and collection $B$ is independent of collection $A$.

Mutually dependent collections of specifications

Two collections of specifications $A$ and $B$ are said to be mutually dependent if collection $A$ is dependent on collection $B$ and collection $B$ is dependent on collection $A$. 

Ottawa ICA / ACI 1999 - Proceedings / Actes
**Completely derived collection of specifications**

A collection of specifications A is said to be completely derived from another collection of specifications B if collection A is derived from collection B and collection A only contains defined relations to collection B and collection A has no defined relations to reality.

**Maps**

Some of the terms defined in this section are illustrated in figure 1.

**Map**

A map is the geometrically related representation which is formed on the basis of physical or abstract, geometrically relatable objects or representations at one or more specific times (the time(s) of validity of the map) on the basis of a specification.

Where necessary, the specification on the basis of which the map was formed is indicated in parentheses after the name of the map. Moreover, the time(s) of validity of the map may be indicated in parentheses, for example A(S, T) indicates that map A was produced on the basis of specification S and represents reality at the times T.

**The empty map**

A map may be empty, i.e. the representation contains no objects.

The empty map with time of validity -oo represents, by definition, reality at time -oo in accordance with any specification and a specification may not, at time -oo, have any other representations than the empty map. (The empty map is introduced, among other things, to make it possible to handle establishment and updating in the same set of rules.)

**Collection of maps**

A collection of maps consists of one or more maps.

**Defined relations between collections of maps**

A collection of maps A has a defined relation to another collection of maps B if the specifications for A have at least one defined relation to the specifications for B.

**Independent collection of maps**

A collection of maps A is said to be independent of a collection of maps B if the specification for collection A is independent of the specification for collection B.

**Dependent collection of maps**

A collection of maps A is said to be dependent on a collection of maps B if the specifications for collection A are dependent on the specifications for collection B.
Derived collection of maps

A collection of maps A is said to be derived from a collection B if the specifications for collection A are derived from the specifications for collection B.

Completely derived collection of maps

A collection of maps A is said to be completely derived from another collection of maps B if the specifications for collection A are completely derived from the specifications for collection B.

Mutually dependent collections of maps

Two collections of maps A and B are said to be mutually dependent if the specifications for the two collections of maps are mutually dependent.

Figure 1. Illustration of terms concerning maps defined in paper

Derived rule:

If a collection of maps A is completely derived from a collection of maps B and collection B is completely derived from a collection of maps C, then collection A is completely derived from collection C.
Updating a map

Updating a map at time T’

Updating a map A(S, T) at time T’ means a process in which a new map A(S, T’) is formed on the basis of map A(S, T), specification S and reality at time T’ so that map A(S, T’) describes reality for time T’ (time of updating) for all relations to reality and, for all defined relations, complies with these relations as specified in specification S for map A.

*There may be defined rules on the relations between a map A(S, T) and an updated map A(S, T’) for the same specification S.*

In practice, such rules occur in particular where objects have only changed a little. In such case, the rule is that remeasurement is not necessary if the operator considers that the accuracy of the existing points in the object is within the specification of the map.

Derived rules:

*If a map A(S1, T) is dependent on a map B(S2, T), map A(S1) cannot be updated at times at which map B(S2) is not updated.*

If map A(S2) is updated at a time T at which map B(S2) is not updated, it is not possible to decide whether the defined rule is complied with. A later update B(S2, T) need not show consideration for the defined rule on the relations of map A to map B and the update of map B may, therefore, result in map A not complying with its defined relations to map B. Therefore, map A cannot be updated at time T until map B is updated at this time.

It follows from the above that:

*Two mutually dependent maps A(S1) and B(S2) can only be updated at a random time T if they are updated simultaneously.*

Examples of use

1. **Simultaneous search in one or more maps**

The creation of a set of search criteria is often used in GIS work. The search criteria may cover one or more maps. When the search is implemented, a search result is produced. The search result is a set of objects (possibly empty).

With the vocabulary defined in this paper, we will describe this process as follows:

We form a completely derived collection of maps (the search result) which has one or more defined relations (the search criteria) to a collection of maps (the maps in which the search was implemented). The defined relations together constitute the specification for the new collection of maps.

2. **Updating the road network in TOP10DK (see figure 2)**

TOP10DK is a Danish specification for the formation of vector maps in the scale of 1:10,000. The specification dictates a 5-year updating cycle. For a wide range of applications of this map product in counties and municipalities, updating of maps every 5 years is sufficiently frequent, but for, for example, bus companies, updating every 5 years is not sufficiently frequent. For route planning, etc., a bus company requires very frequently updated maps (for example, twice a year).
Figure 2. Different solutions to a map updating problem. Illustration to example 2.

A bus company is most interested in the road network and changes in the road network, while, as far as a bus company is concerned, the updating of the remaining contents of the map could well be done every 5 years. Separate updating of the road network could be done on the basis of satellite photographs or driving the roads with a GPS device. What possibilities exist for performing such separate updating of the road network?

We will look at the TOP10DK product which consists of several maps with various contents. A map contains the road network, buildings, areas of countryside, etc. For the objects in each of these maps, the TOP10DK specification describes a number of defined correlations with the objects in the other sub-maps. For example, objects from the road network must not intersect with buildings. Sub-maps are mutually dependent and it follows from the above that they can only be updated simultaneously. Therefore, the road network cannot be updated separately.

However, what problems would separate updating involve and can another solution be found?

Separate updating of the road network would mean that the user could no longer rely on the defined relations in the TOP10DK specification being valid. Map analyses which the user had previously prepared taking the TOP10DK specification into consideration would perhaps no longer be usable. A program which calculates the distance from a building to the nearest road would perhaps not be able to take into consideration the possibility of a road being inside a building and would fail.

Instead of separate updating of the road network, it is necessary to find another procedure which can solve the customer’s problem. In this case, it would be possible to prepare a specification for road update maps. A road update map could contain new roads and roads which no longer exist in separate subjects which did not coincide with the other TOP10DK subjects. In this way, existing programs which show consideration for the TOP10DK map as a whole would still function while customers who have special requirements for updated road networks could prepare analyses which take the road update map into consideration.

3. Division of road objects on the basis of municipal borders

In the TOP10DK specification, road objects are defined as going from a crossroads or a dead end to another dead end or crossroads. This is an expedient definition because it divides the road network into suitable small pieces and because the end points of the road objects can be fixed solely on the basis of the photographic material which is available in the production process during both the establishment and later updating of TOP10DK.

However, for several customer groups, this definition is not the most expedient. For example, authorities with responsibility for road administration want road objects divided according to the limits of their areas of responsibility. Road objects must be divided in such a way that all road objects are either completely outside or
completely inside the area of responsibility. In practice, the municipal borders delimit the areas of responsibility and are, therefore, required to be used to divide the road objects into smaller objects.

Therefore, it is required that the TOP10DK road objects comply with a defined rule on the relations to another map. Division of the existing TOP10DK road objects in the database will presumably be a feasible process but the problems increase in connection with updates.

Updating the TOP10DK database will no longer be possible exclusively on the basis of topographical material such as aerial photographs. Before an update, it is necessary to obtain an updated basis for municipal borders so that it is possible to determine from where and to where each road is to go.

Conversely, TOP10DK is precisely the basic map from which municipal borders are registered. Municipal borders can be registered on the basis of a topographical map because there are no defined rules on the relations from TOP10DK to the municipal map (TOP10DK is independent of the municipal map) whereas there are defined rules on the relations from the municipal map to TOP10DK (the municipal map is derived from TOP10DK). If defined rules on the relations from the municipal map to the TOP10DK road network are introduced, there will be defined rules in both directions between the maps and the maps will have become mutually dependent maps instead of independent and derived maps respectively.

For mutually dependent maps, updating can only take place simultaneously, which would mean a fundamental change of the working process in connection with the updating of TOP10DK.
Session / Séance 41-D

A Case Study of Propagating Updates between Cartographic Data Sets

Lars Harrie and Anna-Karin Hellström
Department of Real Estate Management, Lund University
Box 118, S-221 00 Lund, Sweden
E-mail: Lars.Harrie@lantm.lth.se

Abstract
Ideally a mapping organisation should only be required to maintain one detailed data set. The updates to this data set would then propagate automatically to other less detailed data sets. This article describes a case study of a prototype system for propagation of updates. The case study compares a map produced by the prototype system to a map produced manually by a cartographer.

Introduction

Much research during the last decades has been devoted to automation of cartographic generalisation [Müller et al., 1995]. It has focused on both generalisation methods for real-time applications (e.g. for use in a multiple scale navigation or geographic information system) or batch/interactive applications (for production of new cartographic data sets). The research has, among other things, led to the development of several commercial interactive generalisation systems for simplifying geometrical shapes (e.g. Laser-Scan’s Lamps2 Generaliser and Intergraph’s MGE Map Generalizer).

Updating of cartographic data sets is of major importance today; however, little research has been devoted to this area. Ideally, a mapping organisation should only need to be concerned with updating one detailed cartographic data set and then these updates should be propagated automatically to the other less detailed data sets. To enable this propagation new generalisation methods are required for the update process.

Kilpeläinen and Sarjakoski [1995] propose incremental generalisation for propagating updates between cartographic data sets using a multiple representation database, i.e. cartographic data sets at different scales together with connectivities between objects that refer to the same physical entities. Their major concept is the modularization of the generalisation process; a new object in the target data set should only affect other objects in the same module. However, Kilpeläinen and Sarjakoski never implemented their ideas.

This article presents a case study of a prototype system for propagation of updates between cartographic data sets. The first section provides a brief background of the prototype system and the second section describes a case study, where updates created by the prototype system are compared with those produced manually by a cartographer.
Description of the prototype system

Below follows a brief description of the prototype system (see Harrie and Hellström [1999] for details). The prototype system uses a multiple representation database. A multiple representation database consists of connected cartographic data sets at different scales. The connection is realised by storing connectivities between objects in the data sets that refer to the same physical entities. A multiple representation database has not yet been used in production, but has been studied by several researchers [e.g. Buttenfield, 1993; Kilpeläinen, 1997; Timpf, 1997]. There are several applications of multiple representation databases, but in this study we are solely interested in propagation of updates between cartographic data sets.

The platform for the prototype system is the object-oriented map-production software Lamps2, developed by Laser-Scan, UK (see Laser-Scan, 1997 for details). Lamps2 contains functionalities for storing a multiple representation database and has a function library including topological queries etc. required for the implementation of the prototype system. Furthermore, Lamps2 allows the application programmer to include their own functionality into reflexes; reflexes are methods that are triggered automatically at certain events. This was useful in the creation of the prototype system, since it enabled us to specify actions (in this case: based on propagation rules) when objects were created/deleted in the master data set. That is, reflexes are suitable tools for creating a fully automatic system for propagation of updates.

The prototype system is based on a conceptual framework of four steps: examination, propagation, generalisation of updates and solution of spatial conflicts.

The examination step determines which object should be created or deleted in the target data set. For example, if a new building object is created in the master data set rules in the examination step determines whether a building object should be created in the target data set or if a built-up area object should be modified. The rules in this step are gathered from the data set specifications (National Land Survey [1997a,b]).

The propagation step executes the outcome of the examination step. In addition, the latter step validates the update and maintains the integrity of the multiple representation database, e.g. sets connectivities between related objects.

In the third step the new or modified object in the target data set is generalised to suit the scale of the target data set. This step uses algorithms in Lamps2 Generaliser. The choice of parameter values is mainly based on studies of maps and cartographic data sets in the same map series.

Finally, the constraint method [Harrie, 1999] is used to solve the spatial conflicts caused by the update. This method will distort and/or move the objects according to user-defined constraints.

Case study

The objective of the case study was to compare a map created by the prototype system to a map created by a cartographer. To accomplish this comparison the following tasks were performed:

1) Creation of a multiple representation database.
2) Specification of fictitious updates for the master data set.
3) Propagation of the updates to the target data set using the prototype system.
4) Manual generalisation of the target data set according to the updates.
5) Comparison of the maps created by manual generalisation and the prototype system.

In this section the result of these tasks will be described.
Creation of a multiple representation database

A multiple representation database covering roughly 14 km² was created using two data sets provided by the National Land Survey of Sweden (at scales of 1:10 000 and 1:50 000). In Figures 1 and 2 a small part of the original data sets is shown.

Firstly a schema was created in Lamps2, where object classes, attributes and topological rules were defined. Data were then imported to these object classes. This step implied extensive work. Line objects were in link-node structure and therefore each link became a line-object in the object-oriented database. To create real line-objects the links had to be joined manually. Furthermore, some area classes had to be polygonized. Finally, to create a multiple representation database, the corresponding objects in the two data sets were identified manually, although automatic methods may also be possible [Jones et al., 1996; Devogele et al., 1996].

Specification of fictitious updates for the master data set

Instructions about the objectives of the study and a paper map (Figure 1 shows a small part of this map) were given to cartographers at the National Land Survey of Sweden. The cartographers then made fictitious updates according to the instructions. A main point here is that the persons who selected the updates did not have detailed knowledge about the prototype system.

Propagation of the updates to the target data set using the prototype system

The updates on the paper map constructed by the cartographers were digitised using the ordinary editing functions in Lamps2. The updates were automatically propagated to the target data set, with the last step (solution of spatial conflicts) still requiring some interaction. (See Figure 3.)

Manual generalisation of the target data set according to the updates

The same cartographers who had specified the updates to the master data set performed this task. They manually updated the target data set according to the updates made to the master data set. These updates were drawn on a paper map. Finally, this paper map was digitised in Lamps2 (see Figure 4).

Comparison of the maps created by the manual generalisation and the prototype system

The performance of the prototype system was judged by comparing Figures 3 and 4. These figures only show a small part of the study area, but the results are representative for the whole case study.

The prototype system performed well as regards Figures 3 and 4. However, the prototype system did not perfectly imitate manual generalisation.
Discussion

“The question is whether we can invent protocols to propagate changes (say through updating) from one level of abstraction to all others” [Müller et al. 1995, p. 6]. This question has major practical and economic implications in map production. In our studies we tried to answer that question by creating a prototype system. As a prototype the system performed acceptably, but it had certain limitations. Some of these limitations were due to general problems and are, accordingly, not easy to handle. Other limitations function as a guide to future improvements of the prototype system.

Future improvements

The main problem with the examination step was that all the updates were treated individually. That is, rules were defined for propagation of individual objects. This approach gives different results depending on the order of digitising new objects in the master data set.

The step generalisation of updates only used algorithms included in the Lamps2 generalisation module, and all the algorithms and parameters were set per object class (i.e. no shape analysis was performed). This step should be improved by including shape analysis and more algorithms.
Conclusions

In this article we have presented a case study of prototype system for propagating updates between cartographic data sets. We have shown that such a system can enhance the productivity in the update process.

Acknowledgements

The National Land Survey of Sweden has supported this work financially and through the provision of cartographic data sets, printed by permission 507-98-4091 (© Lantmäteriverket). Laser-Scan, Cambridge, UK, has contributed with software and implementation guidelines. Comments from Bengt Rystedt are appreciated. Special thanks to Fredrik Davidson, Bengt Härjung and Gunnar Jernberg, National Land Survey of Sweden, for help with the case study.

References


Managing Québec’s Cadastral Database

Michel Morneau
Archives and Distribution Division

Daniel Roberge
Analysis and Officialization Service

Gouvernement du Québec, Ministère des Ressources naturelles
Direction générale du cadastre
www.mrn.gouv.qc.ca/cadastre

Abstract
The management of Québec’s cadastral system has taken a new direction with the Cadastre Reform Program. The system, which for more than a century was based on the management of parcel plans, now uses a cartographical database known as the cadastral database, which is in fact a continuous, computer-based map of the four million lots that, together, make up Québec’s cadastre.
The computer version of the cadastral map has legal status as the original. It is updated systematically by means of a spatial reference information system known as the cadastral database management system.
Maximum use is made of the potential offered by the information superhighway for updating and data dissemination purposes.

The Cadastre Reform Program

In 1992, the Québec government entrusted the Direction générale du cadastre (Ministère des Ressources naturelles) with responsibility for implementing the Cadastre Reform Program.
The Program’s three main goals are:
• To constitute a complete and faithful picture of territorial parcellation;
• To ensure that the cadastre is updated continuously. All new parcels must be identified separately on the cadastre before being published in the land book (in Québec, the cadastre is the basic instrument for the publication of immovable rights);
• To ensure the versatility of the cadastre. The cadastre constitutes the basic point of reference for land division in Québec, and it should be possible to use it for other purposes too, such as establishing the boundaries of administrative regions, electoral ridings and municipalities, and for topographical, land valuation and public utility network purposes, etc. The computerized cadastre is consistent with the technical orientations of the government’s Geomatics Plan. (Since 1988, the Québec government has had its own geomatics plan to structure and channel the efforts of the government departments and bodies wishing to use geomatics, and to ensure that the field develops according to a comprehensive and coherent vision. - Internet site: www.pggq.qc.ca)
The first step in this process was to clarify the cadastral product. If the program is to function properly, the people concerned must have a shared understanding of the mission and scope of the cadastre.

New legal rules were also introduced. For example, an innovative legislative amendment was introduced, stating that the computerized version of the cadastre takes precedence over the written version.

Finally, information systems were designed to automate the quality control procedures applicable to the computerized version of the cadastral plan.

**From Plan Management to Cartographical Database Administration**

The renewal of the cadastral product’s form and legal status, as well as the control methods used, has completely transformed the way in which the cadastre is managed. Emphasis has shifted from management of parcel plans to administration of a cartographical database.

The cadastral database is in fact a huge, continuous map of all land in Québec that has been registered in the cadastre – in other words, an area of 117 000 km². The map shows the four million or so properties that make up the cadastre. It contains a set of geometric and descriptive data for each property (or lot).

**Geometric Data**

The geometric data describe the lot’s geometry. Lines represent the boundaries of each lot, forming a polygon identified by a number. The image of the lot is completed by other data, including the measurements of each boundary and the total surface area.

The lots are represented spatially using the official geodetic coordinates system, known as the Quebec plane coordinates system.

**Descriptive Data**

The cadastral database also contains a large quantity of descriptive data for each lot (see Figure 1). The data in question include the name of the municipality in which the lot is located, the name of the owner when the lot was officialized, the registration number of the title, the mode of acquisition (contract, succession, etc.), and correspondence between the old and new cadastral numbers.
Figure 1. Descriptive Data

The Key to Access: The Lot Number

The lot number forms a bridge between the descriptive and geometric components of the cadastral database. It is also the key that provides access to the land book. The land book, in which the rights registered in respect of a particular lot are recorded, are kept in registry offices throughout Québec.

As part of the cadastre reform process, lot identification has been standardized and simplified. The former method, based on the location of the lot and expressed in the form “lot 5-1 du rang Nord-Ouest Grande-Bostonnais du canton de Malhiot”, has been replaced by a system of non-significant numerical identifiers. Each lot within the cadastre now has its own 7-digit lot number, e.g. “1 234 567”.

Database, Standards, and Modelling / Base de données, standards, et modélisation
Since the lot number no longer provides details of the lot’s location, two pieces of descriptive data are used to situate the lot within the area. These are the map sheet number and the reference zone number. For example, for lot 1 260 111, the sheet number is 31H05-010-3435 and the reference zone number is A-8 (see Figure 2).

![Figure 2. Extract From a Sheet of the Computerized Cadastral Map](image)

**The Cadastral Data Management System**

The Cadastral Data Management System is a spatial reference system used to control access to the cadastral database and the updating process, protect the integrity of the database and disseminate its data. In this system, the geometric data are managed by means of the ARC/INFO localized data management system. It allows so-called “topographical” data to be stored and used to the full. Topology can be described as follows: each lot line is “aware” of the lots with which it is associated. The lines “know” that lot “X” lies to one side and lot “Y” to the other. Topology is also used to link annotations to a polygon – for example, a lot number can be linked to the appropriate polygon, or a linear measurement to the appropriate lot line.

The topological feature also links the representation of the lot (the polygon) to the appropriate descriptive data. The relational database management system ORACLE is used to manage the descriptive component. In the system, the geometric and descriptive components are highly integrated, to guarantee perfect synchronicity.
A Major Issue: Updating

Updating is one of the main issues of any cadastral system. Until recently, the means available meant that the process was long and costly, if not impossible. In Québec, where the cadastral system was introduced in the 19th century, the updating process was deficient. First, the legislator had not stipulated penalties for owners who failed to identify land parcels in the cadastre, and second, it was impossible to update the general plans originally produced for each town or village. The officialized parcel plans produced over the years were compiled from time to time, but their non-official status meant that they could only be used for reference purposes.

As part of its cadastre reform program, the Québec government took steps to ensure that the cadastre could be updated. First the computerized version of the cadastral plan was given legal status as the original. This now means that the official general plan can be accessed on a continuous basis, guaranteeing the authenticity of the data. Moreover, all landowners are now required to immatriculate the lots resulting from a new division of land in the cadastre. A right of ownership cannot be published in the land book if the immovable in question does not have its own, separate lot number (Civil Code of Québec, art. 3030). Finally, when a lot is parcelled out, all the resulting sub-lots must be immatriculated at the same time (Civil Code of Québec, art. 3043, 3rd paragraph).

Thanks to the development of computer tools, it has been possible to introduce quick, effective and systematic solutions to ensure that the cadastral data are updated regularly.

The Updating Process

The ministère des Ressources naturelles decided to use the request for a cadastral operation as the means of updating the cadastral database. The request is drawn up by a private land surveyor, on behalf of the landowner and in accordance with the instructions issued by the Minister responsible for the cadastre.

A cadastral operation is an action applied to one or more lots in the cadastral database (creation, correction, cancellation or replacement), i.e. the object of the updating (see Figure 3).

**Figure 3.** Updating Principle: Cadastral Database
The documents required for a cadastral operation are prepared by the land surveyor, and then sent to the ministère des Ressources naturelles in computerized and written form. When the documents are judged to have satisfied the requirements, the operation is officialized and entered into the cadastral database during the night. The following morning, the section of the global cadastral plan (the result of the update) covering the sector in which the cadastral operation is located are sent to the registry office and municipality concerned. In the near future, when the process of computerizing the land book begins, the results of the update will be sent automatically to the registry offices over the Internet, in computerized form. (At the same time as the reform of the cadastre, the Québec government undertook a reform of the rights publication system, with the aim of making it possible, via the information superhighway, to update, disseminate and obtain secure access to land-related information throughout Québec.)

The same process is currently used to send the results of cadastral updating to the municipalities. The municipalities thus have access to reliable information when updating their property taxation system. They can also use the database to develop a multitude of other land management applications.

**Processing of Cadastral Operations**

The Cadastral Database Management System is a transactional system. This means that the information it contains is updated on the initiative of landowners, via land surveyors, within the framework of transactions known as cadastral operation.

Each proposed cadastral operation is subject to strict monitoring (loading, validating and integrating of data). A standardized procedure is also used, so that processing can be automated, thus reducing the effort involved in updating the cadastral database.

When requesting a cadastral operation, the land surveyor submits documents in a specific digital format. The data are loaded into the system and are then subjected to approximately 400 automatic controls.

Once the controls are complete, the system produces a report indicating:

- the presence of errors: a list of errors is sent by the system to the Department’s representative, who forwards it to the land surveyor client;
- the absence of errors: the system submits the image that would result from the update to the Department’s representative. This is known as an integration proposal. The Department’s representative is then responsible for accepting, modifying or refusing the system’s proposal.

When the integration proposal is complete, the system re-enters the data, updating the cadastral database. The structured process applied by the Cadastral Database Management System ensures that the cadastral database is updated while protecting the integrity of its data.

**Data Exchange**

The process of keeping the cadastral database up to date involves exchanges of data between the Department and its land surveyor clients. To manage these exchanges effectively, the Department has developed an automatic electronic mail application that uses the Internet network as a communication vehicle.

The unit operates as shown in Figure 4. Among other things, it can provide the following services without the need for human intervention:

- extraction of official cadastral data;
• pre-validation of cadastral operation files (service reserved for land surveyors). Pre-validation enables land surveyors to check that their applications for cadastral operations contain all the necessary data, and that the map submitted is a faithful reflection of both the land division and the measurements or other information shown on the cadastral parcel plan. It also ensures that the proposed new data are consistent with those already in the cadastral database. Following pre-validation, a report is sent to the land surveyor, stating whether or not the files are in compliance, and if not, giving the reasons why they have been refused;
• transmission of cadastral operation files for officialization (service reserved for land surveyors).

**Figure 4.** Data Exchange

To access any of these services, clients prepare an e-mail message, inserting a product and service order file. In the file, they enter their name, electronic address, account number, personal identification number and the product or service they wish to obtain.

If they are requesting extraction of cadastral data, they must specify the lots or sector they wish to obtain.

If they are requesting pre-validation of files for officialization, they must attach the files relating to the proposed cadastral operation.

Once the e-mail has been sent, it is opened automatically by the unit. The system then reads the product and service order file and automatically fulfils the order.

When a result is available, an e-mail message is generated. The extracted data files or the results of the pre-validation are attached, and the message is sent automatically to the client’s e-mail address. The cost of the transaction is then billed to the client’s account. All these operations are performed without human intervention.

The entire process, between receipt of the order and transmission of the results, takes 24 hours.
Review and Prospects

Québec’s cadastral system is moving resolutely towards the third millennium. Its operations are based on the management of a continuous, computer-based cadastral map that has legal standing as the original. Obviously, users will have to adjust significantly to the changes resulting from the new methods. However, once the period of adjustment is complete, everyone will benefit from the advantages of the new system. As a result, the people working in the field will have access to a reliable, complete, general and versatile cadastral map, covering the area occupied by 92% of the population. The new map will be a considerable asset in the development of geomatics in Québec.
Implications cartographiques de la rénovation cadastrale au Québec

Yaïves Ferland
Laboratoire de législation foncière, Centre de recherche en géomatique
Université Laval
Québec, Québec, Canada G1K 7P4
courriel : Yaives.Ferland@scg.ulaval.ca

Résumé
Depuis deux décennies, plusieurs États ont entrepris d’implanter leur système cadastral de publicité foncière ou bien de réformer ou de compléter celui qu’ils opéraient déjà. En général, des difficultés d’ordre économique, juridique ou opérationnel leur imposent un si coûteux recours. Ou encore, sous les auspices d’institutions internationales de redressement économique, un nouveau cadastre supporterait mieux une réforme agraire ou un marché immobilier émergent. Alors le cadastre se définit comme une structure formelle d’identification, de description, de positionnement et de sûreté des parcelles et des droits fonciers sur un territoire déterminé. Dans ce contexte, en combinant parfois d’autres usages, l’opportunité de réaliser un cadastre à l’aide d’un système d’information géographique (SIG) s’avère déterminante, quoique sa réalisation soit alors plus complexe.

Un cadre théorique structural, élaboré afin d’analyser cette complexité, considère les aspects juridiques, cartographiques et technologiques d’un plan cadastral, originaire ou rénové, en les mettant en relation avec les réalités vécues par les propriétaires et autres détenteurs de droits fonciers, tout en considérant l’évolution historique concomitante de ces trois aspects dans une même société.

Introduction et contexte des réformes cadastrales
Depuis deux décennies, plusieurs États ont entrepris d’implanter ou de renouveler leur système cadastral ou de publicité foncière. Règle générale, les caractéristiques scientifiques et techniques du cadastre pour la représentation concrète des parcelles en font un instrument réputé fiable pour l’application équitable des politiques immobilières ou foncières. Souvent, la justification pour démarrer cette opération part du constat d’une cartographie inappropriée, à cause de son échelle ou de son contenu, ou d’une documentation trop éparse pour couvrir uniformément le territoire. Là où un cadastre opérait déjà, parfois depuis plus d’un siècle, les autorités prennent parfois la décision administrative de le réformer ou de le rénover, pour le rendre plus fonctionnel et
cohérent envers quelque finalité, comme l’attribution des titres de propriété ou l’imposition foncière. Ou encore, certains États décident de compléter leur cadastre et de le mettre à jour afin que d’autres institutions utilisatrices, comme l’enregistrement des droits ou les municipalités, puissent continuer à remplir effectivement leur mandat de manière uniforme sur l’ensemble du territoire. Mais dresser un cadastre n’est jamais innocent ; ce sont de graves difficultés d’ordre économique, social ou juridique qui imposent un effort institutionnel si coûteux ; quoiqu’il s’agisse aussi d’un grand projet politique, comme une réforme agraire, la colonisation intérieure ou la privatisation des exploitations agricoles collectives.

Il est de plus en plus fréquent qu’un nouveau cadastre soit élaboré sous les auspices d’institutions internationales dédiées au redressement économique. Par exemple, la Banque mondiale peut imposer un nouveau cadastre comme étant une nécessité pour dynamiser un marché immobilier émergent, contrôler l’urbanisme, stimuler l’agriculture, ou fixer l’assiette fiscale. On comprend que le cadastre n’est pas seulement une carte du morcellement foncier : c’est une pièce maîtresse du patrimoine collectif et des institutions d’un État moderne, parce qu’il constitue la meilleure représentation possible de la structure de l’établissement d’un peuple sur son territoire. Le créer ou le réformer nécessitent une grande prudence et une compréhension profonde, en plus des compétences techniques, car il doit adéquatement correspondre aux caractéristiques sociales, culturelles autant que juridiques et économiques de la structure foncière et du parcellaire qu’il est réputé représenter fidèlement. Une réforme cadastrale est donc une opération d’envergure menée dans l’intérêt public, et qu’on ne peut justifier que par les énormes avantages pressentis en régularisant la situation foncière sur un territoire déterminé. Ceci dit, la fonction principale qu’on requiert d’un cadastre s’avère très changeante d’un État à l’autre, selon le régime de droit et le cadre législatif, les diverses institutions concernées, les types de tenure et les formes de morcellement du parcellaire, ainsi que la pénétration des innovations technologiques. En définitive, toutes les expériences qu’on recense à travers le monde s’avèrent originales ou hybrides, tant dans leur spécification et leur modélisation des objets sur lesquels portent un droit foncier à représenter au plan cadastral, que dans les circonstances socio-politiques rencontrées lors de l’établissement du cadastre et de sa réalisation.

L’affaire n’est pas triviale à concevoir, et les organismes spécialisés de plusieurs États développent à la longue une expertise cadastrale, toujours plus liée à la maîtrise des technologies de mesure sur le terrain, de traitement des données et de rendu cartographique. Ces États, qui s’y prennent quasiment toujours à plus d’une fois, entendent bien ensuite mettre à profit cette expertise sur le marché international. Pourtant, les transferts sont ardu à réaliser car souvent les responsables ne saisissent pas toute la complexité d’une situation nationale selon les divers aspects, même dans leur propre pays, pourquoi s’en cacher.

Voilà pourquoi la reformulation d’une théorie cadastrale apparaît nécessaire, afin de couvrir une grande variété de situations passées et présentes, stables ou en réforme, traditionnelles ou technologiquement avancées, tout en dépassant les définitions qui se restreignent aux modèles de certaines typologies ou aux spécifications techniques ou procédurières. Alors pour notre propos ici, on définit le cadastre comme étant “une structure formelle de représentation cartographique pour l’identification, la description, le positionnement et la sûreté des parcelles délimitées et des droits fonciers identifiés sur l’ensemble d’un territoire déterminé, ou au niveau de circonscriptions particulières, constitué et opéré par l’État pour des fins d’ordre public”.

**Trois aspects d’une théorie cadastrale**

Un cadre théorique structural a été patiemment élaboré afin d’analyser tous les genres d’institutions cadastrales, même les plus complexes. Il met en relation les aspects juridique, cartographique et technologique d’un cadastre, originaire ou rénové, dans leurs rapports à la réalité vécue sur le terrain par les propriétaires, d’autres détenteurs de droits fonciers, et l’État. Aussi, il permet d’étudier l’évolution historique et sociale des trois aspects mentionnés dans leurs rapports mutuels avec le territoire. Ce cadre théorique du cadastre est qualifié de “structural” car on
y considère que ces trois aspects sont les modalités nécessaires et suffisantes de toute représentation “construisant” la réalité foncière, eu égard aux définitions usuelles. Ces aspects entretiennent des relations dynamiques et significatives entre eux, lesquelles sont parfois contrariantes et même contradictoires, et cela conditionne l’évolution de la situation qu’ils décrivent à une époque et sous une juridiction données.

L’évolution prend l’apparence d’un événement lors du passage obligé d’un ordre juridique à un autre, pendant une réforme agraire, ou d’une lente dégradation des qualités représentationnelles du plan cadastral sans une tenue à jour, au fil des ans. L’évolution n’est pas tant linéaire mais résulte plutôt des tendances sociales profondes. Selon les rapports entre les modalités, les effets isolés de la structure cadastrale deviennent insatisfaisants pour rendre compte de la réalité, en évolution elle aussi. Puisque les modalités du cadastre interagissent à travers la structure qui les conditionne mais ne les détermine pas, cette évolution ne devient modélisable et explicable que si on parvient à analyser certains effets d’une modification sur la réalité.

L’aspect juridique de la représentation cadastrale

Prenons d’abord le contexte juridique qui définit et règle les relations de droit existantes et licites, dans une juridiction d’État particulière, entre les sujets détenteurs de droit et l’objet foncier sur lequel ce droit s’exerce. La relation de droit foncier la plus courante en Occident est la pleine propriété individuelle, dite “absolue”, qu’une personne détient seule sur un lopin de terre distinctement délimité et identifié, ou au moins reconnaissable en particulier sur le terrain. Cette relation juridique est consacrée dans la loi puisqu’elle émane de la situation de fait admise dans une société civile. Par contre, cela contraint la réalité vécue de chaque sujet de droit en préservant les conditions d’accession et d’exercice de ces droits et en définissant ce qu’est une parcelle ou un lot. La loi établit aussi comment se fait la reconnaissance d’un tel droit individuel et adopte au besoin un mode de publicité foncière ou immobilière, en spécifiant ce qui doit être rapporté à la connaissance du public.

En général, les modes de reconnaissance opèrent par le dépôt, l’enregistrement ou l’inscription d’une preuve de droit, dont les quatre principales sortes sont : la déclaration, l’acte (ou le contrat passé entre deux parties lors d’une transaction), le titre accordé par l’autorité, ou le droit strictement défini et constaté aux registres. Ainsi, la réforme du Code civil du Québec fait passer la publicité foncière d’un mode d’enregistrement des actes à un mode d’inscription des droits réels immobiliers. Selon le système cadastral, le mode induit aussi la fonction qui y est privilégiée, autrement dit la “mission” représentationnelle de la modalité cartographique du plan cadastral. Les missions les plus courantes, qui ne sont d’ailleurs pas toujours exclusives entre elles, sont connues comme étant les types de cadastre “graphique”, “fiscal”, “allocatif” ou “juridique”.

Cette fausse typologie demande quelques éclaircissements, car ces quatre missions ne sont pas de même espèce, ou niveau. En fait, tous les cadastres composés d’une description technique accompagnée d’un plan sont graphiques; or au sens strict, cela signifie que le plan n’a d’autre fonction que de décrire géométriquement et en position relative les parcelles ou les lots de et de les identifier, normalement par un numéro. Dans un cadastre fiscal, on indique la valeur estimée de chaque bien foncier (pas forcément en fonction de la superficie) ainsi que le nom de la personne devant payer l’impôt, normalement en tant que propriétaire; c’est la mission première du cadastre français. En pays de colonisation, lorsqu’un cadastre est confectionné avant qu’on occupe le territoire et qu’il le représente découpage en lots réguliers qu’il faudra ensuite aller arpenter sur le terrain, cela sert à allouer des terres aux colons à qui on accordera un titre de possession ou de propriété; cela fait, le cadastre allocatif demeurera un cadastre graphique à moyené échelle, montrant toute la structure de l’établissement humain et conditionnant par la suite la subdivision des parcelles de terrain. Il s’agit par exemple du cadastre originaire des cantons au Québec et, à une moindre qualité de précision et de fiabilité, des plans d’arpentage usités dans les États américains et australiens qui ont adopté le fameux système Torrens d’enregistrement des titres. Le cadastre juridique porte le plus à conséquence, car les mesures très précises et les positions exactes des limites et des bornes, toutes montrées au plan cadastral à grande échelle, ont préséance juridique sur les autres descriptions et même sur la possession réelle sur le terrain; c’est le type idéal du cadastre helvétique.
Retenons surtout que l’aspect juridique d’un cadastre, dont la finalité est la représentation formelle des droits fonciers, opère de façon contraignante envers l’aspect cartographique, quoique des formes de modalité juridique changeantes ou différentes sont supportables par un modèle cartographique semblable; ces modalités évoluent et s’influencent en une même structure.

**L’aspect de la représentation cartographique**

Toute représentation cartographique est constituée d’une triple médiation de l’espace géographique. Dans le cas du cadastre, il y a d’abord l’identification par l’arpenteur-géomètre des objets perceptibles à représenter, qu’il devra mesurer et situer les uns par rapport aux autres selon un modèle significatif, régularisé et normatif. Vient ensuite la confection cartographique de ces objets sur un support accessible, selon une symbolisation idoine: au plus simple, des lignes correspondent aux limites de parcelles et les polygones formés aux lots eux-mêmes; aussi, il est usuel de numérotter ces polygones selon une séquence qui simule la succession topologique des lots dans l’espace, de proche en proche. Ce double processus de représentation nécessite la maîtrise de moyens technologiques adéquats, et conduit à la confection d’une carte réelle, le plan cadastral. Enfin, c’est la consultation de celui-ci qui permettra à tout sujet de droit de reconnaître et d’identifier sans équivoque l’objet de son droit, son bien foncier. En cas de discordance entre le droit (relatif sur un titre ou dans un acte), ce que le sujet constate au plan et ce qu’il assume être sa possession, le plan cadastral consulté aura une influence concrète dans la reformulation de la réalité. Cette troisième médiation sera encore plus forte et définitive lorsqu’on implantera systématiquement sur le terrain des lignes établies d’abord sur le plan de cadastre allocatif, contraignant la réalité concrète avant même qu’un sujet de droit s’y installe.

La cartographie est l’art, la science et la technique de représenter l’espace. Pour confectionner le plan cadastral, les “règles de l’art” relèvent de la culture juridique de l’arpenteur, qui lui permet d’identifier les objets de droit sur le terrain selon ce que dicte la loi, tandis que la technique pour les représenter est plutôt spécifiée par des normes ou des instructions de l’autorité. Sa science consiste en la modalité cartographique: c’est l’ensemble des méthodes de mensuration, de calcul et de dessin qui lui permettent de réaliser une carte précise et fiable, présentant avec uniformité toute l’information requise, et seulement elle.

Plus il y a d’éléments à représenter afin de servir la mission d’un cadastre, plus les relations seront complexes entre les modalités juridique et cartographique de représentation, plus l’information sera dense tout en comportant des risques d’erreur ou de contradiction, et plus vite la qualité du plan cadastral ira en se dégradant sans une mise à jour ou une compilation tenant compte de la fréquence et de l’intensité réelles des modifications sur le terrain. Cela est d’autant plus vrai lorsqu’on recourt à un même cadastre pour remplir deux ou plusieurs missions, chacune demandant des données propres en supplément. En allant vers une abstraction croissante, il sera crucial de représenter au plan cadastral des objets de droit stables ou des repères remarquables, ainsi que les structures géographiques sur lesquelles s’appuie la mission, en particulier les lignes de base et les limites de circonscriptions foncières, ou encore les lignes de référence d’un système de coordonnées géodésiques. Les chemins et les cours d’eau, même sans constituer des lots ou des limites fixes ni porter de numéro d’identification propre, constituent aussi des réseaux fondamentaux. Ils ont structuré l’établissement humain et ils permettent de montrer avec plus de précision, et d’une façon plus accessible pour la plupart des gens, la position relative entre les parcelles qui les bordent.

**L’aspect instrumental et technologique**

Pour sa part, la technologie est le troisième aspect de la cartographie et par conséquent de la structuration du cadastre. Dans l’ordre, elle consiste en ces instruments de mesure, de positionnement, de gestion des données et de confection du plan qui autorisent la réalisation plus rapide de travaux plus sûrs, donc des médiations plus efficaces. La technologie est donc le passage obligé entre l’ensemble des notions et concepts de la science...
cartographique qu'elle supporte et applique, et la compréhension possible des informations disponibles sur une représentation cartographique qu’on consulte. Les techniques traditionnelles de dessin et d’arpentage, puis les photographies aériennes montées en modèle stéréoscopique, ont longtemps livré des résultats très satisfaisants en fonction des besoins de représentation, bien avant le perfectionnement d’instruments de mesure et de positionnement ayant des composantes informatiques ou satellitaires. Depuis l’Antiquité, des techniciens à l’esprit pratique conçoivent les instruments qu’ils utilisent, améliorant leur maniement et leur précision, ce qui détermine directement l’évolution des moyens et des formes de représentation cartographique.

Or les rapports à l’innovation changent lorsque, la technologie l’informatique est développée depuis l’extérieur de la pratique de la cartographie plutôt que de surgir de son évolution propre, car cela impose alors un système de contraintes techniques auquel il faut savoir s’adapter. Aujourd’hui, divers instruments permettent la cueillette de données concernant tel objet mesurable à la surface de la Terre, son positionnement relatif mais aussi absolu par rapport au début et aux coordonnées géodésiques, et sa représentation à l’écran selon des systèmes de projection bien contrôlés; si au besoin on trace une carte numérique, elle ne sera qu’une représentation de la base de données. Grâce à certains autres instruments, qu’on peut en ce cas appeler des “outils” parce qu’ils modifient la réalité, la médiation inverse d’implantation est aussi facilitée: passer de la représentation cadastrale à la concrétisation sur le terrain des limites et de repères, conçu abstraitement et planifiés d’avance.

La technologie joue donc un rôle différent selon la médiation cartographique à laquelle chaque instrument s’applique, même lorsqu’on considère leur niveau de développement ou de complexité. Le système de positionnement sur le globe (GPS) a une incidence qui diffère complètement de celle d’un système d’information géographique (SIG) utilisé pour la modélisation des données à référence spatiale et la représentation strictement géométrique à l’écran, et tout autant de celle de l’Internet pour l’accès et la transmission des données. Par ailleurs, avec l’efficacité de l’opérateur, l’exactitude des mesures fournies par ces instruments est sans doute la plus importante qualité technique que puisse requérir l’aspect juridique du cadastre. La loi reste stable pendant de longues périodes avant d’élever ses normes exigibles à cause de progrès plus rapides d’une technologie qui l’aura longtemps dépassée. Cependant, sauf préséance légale, la technologie a rarement influencé la modalité juridique pour définir un droit foncier et son objet. Elle ne semble pas plus modifier la réalité vécue par les sujets de droit (sinon par la sûreté qu’elle procure quant à la précision), sauf quand il s’agit d’outils pour implanter des limites calculées sur le terrain.

Nous avons vu à propos de l’aspect juridique que la mission graphique était en fait le cas général pour tous les cadastres comportant un plan cartographique. Quelques explications s’imposent, car avec le progrès scientifique sont apparus des types de cadastres qui ne relèvent pas des missions poursuivies, mais qui ont provoqué un fort impact sur la forme de support des représentations cartographiques et leurs données surtout. En effet, après le type graphique de base qui consiste en un dessin cartographique, sont apparus ces dernières décennies les types de cadastre “numérique” (imaginé au début du XXe siècle), puis aujourd’hui “informatique”. Le cadastre numérique, réfère à l’identification précise ou au calcul exact des sommets, des limites, des bornes ou des centres géométriques (appelés aussi des “centroïdes”) de chaque parcelle, selon les données chiffrées d’un système de coordonnées absolues. Il s’agit d’un changement majeur par rapport à l’habituale description technique d’une parcelle, qui prend une forme littérale dans un registre ou graphique au plan cadastral, et dont les données sont relatives à la structure de disposition et du découpage des parcelles. Cette propriété numérique permet de concevoir un même cadastre à missions multiples (on dit aussi “multi-usage”) qui représenterait des lignes et des formes géométriques se superposant mais pouvant ne pas concorder entre elles, sans que cela cause un problème de représentation ou de confusion: pour un même immeuble, son unité d’évaluation foncière pourrait être géométriquement différente de son lot cadastral.

Quant au cadastre informatique qu’on développe sur un système d’information géographique (SIG), il s’agit d’une base de données à référence spatiale intégrant la forme numérique (qu’on peut même montrer sur une table) avec la représentation géométrique de ces données à l’écran, laquelle peut aussi être imprimée. Une
telle représentation informatique ne montre plus une image du territoire cadastral, mais une image de la base de données, ce qui est fondamentalement nouveau quoiqu’en réalité, pour la personne qui ne fait que consulter, la différence sera accessoire. Pourtant, une nouvelle médiation cartographique est ainsi introduite dans notre modèle théorique, et la précision souffrira de cette manipulation supplémentaire des données du plan. C’est que la constitution d’une telle base de données cadastrales peut s’effectuer de deux façons:
- soit par la numérisation directe d’un plan cadastral existant, à l’aide d’un “scanner”, ce qui donne une structure de données formée d’un grand nombre de “pixels”, au point par point; cette structure est dite “en tessellation”;
- soit par la saisie ligne par ligne ou sommet par sommet de chaque élément géométrique apparaissant sur un plan cadastral existant, ce qu’on appelle une numérisation “vectorielle”.
En ce dernier cas, on pourra au moins reconstituer la géométrie et la topologie de chaque parcelle en liant les lignes adjacentes pour reformer la figure polygonale des objets de droit complets. Mais les données géométriques autant que descriptives d’un cadastre informatique sont enregistrées et conservées sous une forme tabulaire: la base de données. Une autre fonctionnalité informatique permet de disposer ces objets sur quelques couches de données, selon des thèmes représentatifs pouvant concorder, par exemple, avec des diverses missions d’un cadastre “multi-usage”. Enfin, on dotera chaque objet d’un identifiant unique et on pourra lui attribuer d’autres caractéristiques ou attributs, avec leurs propres données.
Pour écarté cette médiation supplémentaire et ses effets sur la précision des données, la technologie informatique s’avérera mieux utilisée si on saisit les données avec de puissants instruments de mensuration, directement sur une base de données, lors d’une opération d’arpentage sur le terrain. Selon l’importance que les autorités et la loi accorderont à une telle entreprise, cette troisième façon pourra aller de quelques mesures de contrôle pour vérifier l’exactitude du plan cadastral antérieur ou sa version déjà numérisée, jusqu’à une véritable réforme de l’institution cadastrale effectuée par un arpentage informatisé. Entre les deux, des opérations menées au besoin sur des territoires restreints mais densément morcelés, comme les villes, seront la seule façon efficace et justifiée qu’adopteront les États actuellement dépourvus pour confectionner leur nouveau cadastre.
En Occident, où les institutions cadastrales sont déjà bien établies, l’avènement du cadastre informatique comporte une transformation surprenante, et d’autant mal contrôlée, des relations entre les modalités de représentation avec la réalité vécue sur le terrain. L’aspect technologique de la rénovation cadastrale prend soudainement plus d’importance socio-politique et cette modernisation de l’institution devient la justification première de réformes entreprises par certains États, sans qu’il y ait de crise foncière urgente ni de grands enjeux à l’égard de l’occupation ou de l’organisation du territoire. Ainsi la technologie informatique, avec ses formidables capacités de traitement et d’enregistrement des données, s’impose à l’aspect juridique. Parfois la mise à jour, l’uniformisation ou la refonte du plan cadastral déclassent même d’autres besoins pourtant identifiés envers une réforme juridique de certaines relations de droit foncier qui étaient nouvelles ou complexes, ou ignorées jusque là. Également, en privilégiant certaines structures de modélisation des données, elle contraint les représentations cartographiques et provoque des problèmes aux conséquences imprévues. Quelques éléments de la mission cadastrale peuvent s’en trouver changés et des usages plutôt officieux du cadastre sont rendus difficiles, sinon empêchés. Souvent, la troisième façon de remplacer un cadastre graphique par un cadastre informatique constitue plus une coupure qu’une évolution, alors que des éléments de représentation disparaissent parce que la nouvelle structure uniformisée et simplifiée n’en tient plus compte. C’est en partie ce que nous observons dans le cas du Québec, qui illustrera le mode opératoire de cette théorie cadastrale qui fut justement développée pour mieux comprendre ce genre d’impacts technologiques sur les représentations cartographiques.
Le cadastre québécois et sa réforme

Le cadastre officiel de la province de Québec a été institué par l’adoption du Code civil du Bas Canada, en 1866, pour servir au système d’enregistrement des hypothèques, lui-même ayant été lentement établi depuis 1830. Cette époque effervescente aux points de vue socio-économique et politique avait connu entre autres l’abolition du régime seigneurial et l’instauration d’un régime municipal, vers 1854-55. Désormais, un seul type de tenure et de droit de propriété foncière était reconnu, avec certains démembrements comme les servitudes et l’usufruit, et la loi veillait à protéger le marché immobilier. La crise foncière trouvait une solution institutionnelle très articulée qui allait fonctionner pendant plus d’un siècle, jusqu’à nos jours.


Brève histoire des cadastres officiels, ou originaires

Pays neuf ouvert à la colonisation à l’aube du XVIème siècle, le sud du Québec avait connu deux méthodes de découpage du territoire pour l’attribution de terres aux nouveaux arrivants: les seigneuries sous la Nouvelle-France, le long des principaux cours d’eau, puis les “townships” vers l’intérieur, après la conquête britannique et l’arrivée des Loyalistes. Bien que donnant des résultats assez différents, les deux méthodes avaient en commun, dans leur principe, de procéder à un arpentage primitif. On traçait des grandes lignes de référence d’un territoire de concession, qui allait devenir plus tard l’unité cadastrale à être représentée par un plan particulier; ensuite, on découpaît dans cette unité de territoire des rangées de grands lots ayant une orientation, une forme et des dimensions semblables, pas toujours régulières d’abord mais au moins cohérentes en pratique, avec peu d’égards pour la topographie ou le couvert végétal. Les premiers cadastres officiels furent confectionnés après 1866 par l’inventaire sur le terrain de tous les lots alors existants, dans leur grandeur d’origine ou en parcelle déjà subdivisée dans les villes et villages. On créa un cadastre officiel pour chaque paroisse (en territoire seigneurial), chaque ville ou village et chaque “township”. Il n’y avait de cartes, déjà dressées et qu’on aurait pu copier à grande échelle, que dans ces derniers cas.

Due à sa logique interne, la structure spatiale de l’arpentage primitif (remontant parfois à deux siècles) était évidente sur la carte comme sur le terrain, quoique altérée par l’état du morcellement foncier. Celui-ci se produisait à l’intérieur du canevas, même en cas de remembrement de lots. Cela se reflétait aussi par la numérotation originale de tous les lots à la confection de chaque cadastre: on identifia chaque lot par un numéro permanent d’une seule séquence, de proche en proche, rangée par rangée. Toute subdivision subséquente provoquait une sous-numérotation, de la forme 101-2 pour désigner le deuxième lot subdivisé à partir du lot originaire 101. Ce système cohérent évoluait avec le morcellement car les sous-numéros indiquaient l’ordre temporel des subdivisions ou des opérations de lotissement. Par malheur, on n’appliqua pas ce système strictement en opérant nombre de morcellements fonciers sans cadastrage ni désignation propre, n’y référant qu’à des “parties de lot”.

Bien des problèmes cadastraux ont surgi puisque le développement urbain intensif s’est produit surtout dans les anciens cadastres du domaine seigneurial, où la structure de division territoriale, les mensurations et la numérotation étaient moins régulières.

Il n’y avait pas encore d’uniformité, car chaque unité de territoire à concéder avait été arpentée indépendamment, en s’ajustant aux limites des concessions antérieures. Mais profitant de l’expérience acquise avec l’inventaire cadastral des premières décennies, et pour assurer les politiques de colonisation intensive, une troisième méthode est apparue: le canton (traduction du terme “township”). Selon des instructions systématiques, on peut arpenter
plusieurs cantons d’une seule venue et les cadastrer à peu près en même temps, en suivant des lignes de base géodésiques dépassant la centaine de kilomètres, tout en corrigeant l’orientation cardinale selon la rotondité de la surface terrestre. Les principes de délimitation planifiée à l’aide de la carte et de numérotation pour identifier chaque lot originaire demeuraient, alors que la qualité technique des instruments s’améliorait pour l’implantation précise des lots selon des coordonnées géodésiques. Le cadastre graphique des cantons eut d’abord une mission allocative, pour l’attribution de titre aux colons s’établissant sur des terres arpentées au préalable.

Dès les débuts, toute les structures cadastrales ont servi de base à la division territoriale des municipalités, ce qui supportait aisément à la nouvelle mission fiscale d’évaluation foncière des lots. Ce régime d’allocation des terres est tombé en désuétude vers les années 1950, alors qu’environ 150 000 km² seulement ont été cadastrés dans le tiers sud du Québec. Cela donne actuellement 1449 cadastres officiels ayant leur propre plan originaire d’un seul tenant, aux dimensions très variées sur le territoire comme sur la carte. De temps à autre, une compilation cartographique à plus grande échelle pouvait être faite pour intégrer tous les plans parcellaires produits lors d’opérations cadastrales, donnant un plan cadastral plus exact et remis à jour.

**La première rénovation cadastraale de 1985**

Un programme de rénovation cadastraale fut lancé en 1985 pour résoudre quelques difficultés qu’on vient d’évoquer, surtout dans les milieux urbanisés. D’abord, l’enregistrement de droits fonciers sur des “parties de lot” s’était gravement compliqué avec la hausse des transactions immobilières. Pour sécuriser les droits sur de petits lots souvent mal déterminés, il devenait nécessaire de remesurer puis de renuméroter selon la même méthode qu’avant: les “parties de lot” non identifiées, des lots nouveaux issus de remembrement ainsi que les lots originaires ayant subi quelques morcellements. De plus, en interdisant désormais l’enregistrement d’un droit réel sur “partie de lot”, on forçait d’opérer une subdivision cadastraale conforme. Ceci dit, la rénovation cadastraale avait peu d’effets juridiques nouveaux, si ce n’est le remembrement des parcelles contiguës d’un même propriétaire acceptant de ne faire qu’un lot avec son immeuble, mais sans unifier des immeubles différents. Aussi, le programme était volontaire en ce sens que les clientes, les municipalités, demandaient la rénovation de leur territoire parce qu’elles escomptaient disposer d’une nouvelle cartographie à jour, à grande échelle et sur support informatique. Le plan rénové, quoique respectant la structure foncière, ne correspondrait pas nécessairement aux divisions du cadastre officiel. Néanmoins, il montrerait le détail de chaque lot grâce à un relevé d’arpentage systématique des limites sur le terrain, avec rattachement au réseau géodésique. Ainsi, tout en recourant à la dernière technologie disponible, cette rénovation du cadastre était bel et bien une compilation cartographique, mais montrée sur des feuilles cadrées selon le découpage officiel de l’État.

**L’actuelle réforme du plan cadastral**

d’autres utilisateurs potentiels des données rénovées, grâce à son support informatique qu’on pourra ainsi tenir à jour quotidiennement. De ce fait, le cadastre ne constituera plus une représentation cartographique de la structure du territoire, il sera déterminé par la structure des données informatiques.

Les objectifs premiers de la réforme de la publicité foncière émanaient d’abord des besoins juridiques de sécuriser les titres de propriété, de simplifier la procédure de tenue des données décrivant les droits fonciers et les lots sur lesquels ils portent, afin de maintenir une information complète et suffisante pour toute transaction, ce qui faciliterait les recherches notariales. Mais les objectifs déclarés de la réforme cadastrale ont suivi leur propre évolution depuis cinq ans. D’après l’évaluation externe menée à l’aide de notre cadre théorique, il s’avère que plusieurs problèmes conceptuels, juridiques et surtout cartographiques suscitent des questionnements chez plusieurs acteurs de l’institution cadastrale quant aux intentions et aux besoins de l’État et des usagers envers le cadastre rénové, sur informatique. Bien qu’il n’y ait pas de crise foncière, à peine une simplification légale de la propriété, et que le marché immobilier soit mûr et requiert moins de lotissements, la rénovation du plan cadastral demeure justifiée en pratique et en droit, mais on constate sans aucun doute qu’elle est d’abord technologique. Finalement, la loi n’en spécifiant pas les conditions techniques et graphiques, on voit que ce cadastre sur informatique fonde de lui-même un nouveau type juridique de lots individualisés, ayant une forme et des limites figées, absolues, sans rapport au contexte spatial, ce qui diffère complètement de la définition antérieure du lot en tant que portion de terre découpée dans le territoire. Désormais, la modification cadastrale d’un ou de plusieurs lopins de terre entraînera la disparition et l’apparition simultanée de lots immatriculés avec des numéros de désignation légale différents, alors que l’immeuble demeure bien en place.

Avec ce renversement de perspective, allant d’une définition juridique vers les contraintes informatiques de la représentation cadastrale, nombre de problèmes d’application cartographiques surgissent, concernant entre autres : l’effacement complet de la structure foncière, donc la perte de lisibilité du territoire et de son évolution; l’immatriculation à numéro unique, dépourvue de signification, de chaque lot délimité au plan; la perte d’information par la mise et la tenue à jour constante du plan; la modélisation du système de projection, du datum et des rattachements géodésiques ne tenant pas compte de l’altitude du lot au sol et dans ses limites concrètes; le découpage en feuilles selon le système de coordonnées géodésiques, générant une redondance systématique le long du cadre, contrairement aux anciennes cartes de cadastre officiel d’un seul tenant, sans aucune coupure la provenance des données sur les limites et les mesures, obtenues par compilation de sources indirectes (bién autorisées) telles que les actes notariés, plutôt qu’avec des relevés d’arpentage contrôlés sur le terrain; le traitement des limites par compensation statistique des erreurs systématiques de mesures et des marges de tolérance, dont le degré de précision dépend de la résolution de l’œil et de l’élément graphique à l’écran d’ordinateur (pixel), et non de l’échelle spatiale.

**Principales difficultés représentationnelles sur la modalité cartographique**

Le cadastre rénové présente de réelles améliorations représentationnelles, mais aux conséquences parfois imprévues. Avec une seule base de données sur une seule couche informatique, il n’y aura plus qu’un seul cadastre virtuel du Québec, couvrant tout le territoire concédé à la propriété privée. Ce concept fantastique repose uniquement sur les coordonnées géodésiques et le système officiel de projection et de découpage cartographiques, autrement dit, sur la position absolue ( quoique compensée) des entités de base individualisées qu’on représente dans la base de données du cadastre rénové, soit les lots immatriculés. Tel que confectionné, le cadastre ne montrera plus la position relative des limites des lots par rapport à la structure cadastrale et foncière. Chaque limite entre deux lots parait une simple ligne dont la position est reconnue dans les limites de tolérance graphique autorisées; cependant, deux entités voisines ont chacune leur limite superposée, avec une mesure pouvant avoir une différence arithmétique! Chaque entité déterminée possède un identifiant univoque et est liée à une certaine géométrie définie pour un lot particulier; toutefois, cette entité est éphémère puisqu’en y opérant une modification cadastrale sur une seule limite, on supprimera deux immatriculations pour en créer.
deux autres alors que les propriétaires voisins demeureront les mêmes, au même endroit. Malgré sa flexibilité, la représentation par entité apparaît comme une contrainte informatique.

Les mandats de rénovation cadastrale sont exécutés par secteurs de quelques centaines de lots à la fois, tenant normalement dans les limites d’une seule municipalité et d’un même cadastre officiel originaire. Mais sans contrôle effectif de l’ensemble des limites au sol, la compensation statistique de toutes les dimensions des lots rénovés mène à une généralisation des erreurs et des mesures plus ou moins précises, voire contradictoires. La discordance devient inévitable entre les dimensions inscrites aux actes ou aux titres de propriété, les mesures levées sur le terrain et les limites montrées au plan, même si ces dernières ont été compilées à partir des premières (mais compensées, donc toutes légèrement déplacées) ou des deuxième (ne serait-ce qu’à cause de l’altitude réelle du lot à la surface terrestre, où les distances sont nécessairement plus grandes que sur le datum géodésique). Lorsqu’on tentera d’assembler le plan des secteurs rénovés, leurs limites ne concorderont pas ensemble ni avec celles des municipalités ou des cadastres officiels. Une seconde compensation à cette échelle ne ferait qu’empirer le résultat.

L’immatriculation est une procédure de numérotation obligatoire de chaque lot individualisé au plan cadastral. Ce nouveau mode de numérotation fournit la seule désignation légale et exclusive pour tous les lots cadastrés du Québec. Le numéro matricule unique, abstrait et non séquentiel n’est rien d’autre que l’identifiant de l’entité dans la base de données informatique. En résolvant ainsi le problème des “parties de lot”, il y a donc abandon complet d’un mode opératoire de désignation et de numérotation séquentielle des lots de proche en proche, puis de sous-numérotation des subdivisions. Suivant la procédure prévue, ce mode donnait un numéro d’identification pérenne à tout lopin de terre déterminé et existant à un certain endroit dans l’espace géographique, c’est-à-dire dans les limites d’un lot originaire, faisant partie d’un cadastre officiel particulier.

La représentation cartographique subit une grave perte de lisibilité en regard des structures topologiques de découpage du territoire et des réseaux viaires et hydrographiques. Cette perte de lisibilité et de signification est à la fois spatiale, temporelle et culturelle. Non seulement la structure foncière disparaît du plan cadastral, entre autres par les remembrements, mais ensuite elle devient difficile à reconstituer pour interpréter la représentation elle-même du plan et de ses rapports à la structure réelle du territoire. Or ce genre de lisibilité doit être directe et autonome, c’est-à-dire que même le spécialiste n’ait pas à recourir aux fonctionnalités informatiques pour saisir la topologie des lieux. L’évolution historique est aussi en cause, par l’abandon d’un système de sous-numérotation significatif localement, montrant la séquence du lotissement et les parties résiduelles de lots originaires, ainsi que par l’élimination de toute référence usuelle, comme la toponymie foncière des noms de cadastres officiels, des rangs, des concessions, des cantons et même des municipalités. Rétablir la correspondance entre les numéros de lot relatifs de tel cadastre officiel et le numéro nouveau, unique et absolu, d’un lot rénové ne sera pas suffisant pour maintenir la lisibilité spatiale et temporelle entre les cadastres originaire et rénové. Cela écarte d’autant plusieurs possibilités d’utilisation externe du plan cadastral à d’autres fins que l’immatriculation et l’enregistrement des droits fonciers.

**Des correctifs cartographiques ou informatiques à proposer**

L’intention principale de proposer des correctifs au plan cadastral rénové vise à réintroduire plusieurs éléments significatifs de la structuration du territoire dans la représentation cartographique du cadastre. L’intérêt de la modalité cartographique est justement de maintenir la continuité de l’information spatiale ou du moins la capacité de la reconstituer aisément, en fonction de la modalité juridique et aussi administrative du cadastre; outre les fins de la publicité des droits fonciers, cela servira surtout aux municipalités, et aussi à la consultation du plan cadastral rénové par des tiers en tant qu’usagers peu spécialisés.

En admettant le principe de l’immatriculation de chaque lot, on conservera ces longs et uniques numéros à titre d’identifiants informatiques des “événements” cadastraux modifiant les mesures ou les limites de la forme.

_Ottawa ICA / ACI 1999 - Proceedings /Actes_
géométrique d’un lot (par exemple la création du lot, sa rénovation ou une opération de bornage ou de lotissement). Cependant, on faudrait réintroduire dans la base de données la numérotation parcellaire séquentielle des lots voisins, avec une méthode efficace et significative de désignation des subdivisions de lots, quitte à corriger ou à consolider les situations contradictoires ou aberrantes qui seraient rencontrées, comme la trame d’îlots urbains qui auraient été orientés différemment de la structure foncière originale.

Parallèlement, pour vraiment montrer toutes les limites en position et en mesures relatives, et non pas seulement absolues, réintroduire les cours d’eau et les plans d’eau, lesquels constituent un réseau structurant dans la délimitation de lots riverains. En fait, les limites naturelles ne peuvent être réduites à une simple ligne définitivement fixée avec une trop de précision, sauf si elle a déjà été légalement démarquée, car le lit d’un cours d’eau risque toujours de subir une déviation naturelle ou artificielle sans que cela porte atteinte à la description de l’objet de droit (un lot bordé à la rivière), qui demeure explicite.

On ne devrait jamais procéder à un remembrement obligatoire et automatique lorsque cela n’est pas justifié (comme ce serait le cas pour un seul et même immeuble clairement identifiable), ou quand cela porte atteinte à l’intérêt du propriétaire, ou que cela efface des limites de lot originaire faisant partie de la structure foncière mais qui ainsi ne seraient plus identifiables. Dans tous les cas de remembrements et de conservation des lignes originales, on devrait procéder au contrôle systématique, sur le terrain, des résultats de la compensation statistique appliquée aux mesures recueillies dans les titres ou les actes de propriété.

Enfin, pour établir une vrai cadastre de l’information foncière, on devrait profiter des capacités informatiques pour créer des couches de données secondaires afin de pouvoir enregistrer la position de démembrements ou des contraintes administratives envers les droits de propriété sur des parcelles, comme des servitudes ou des zones de protection riveraines ou agricoles.

**Conclusion: un cadastre pérenne ou utopique**

Voilà le résultat d’une réflexion critique visant un cadre théorique structural développé à la fois pour l’étude de l’institution cadastrale en général et d’un cas québécois de rénovation en particulier. Il s’agissait d’estimer l’impact des technologiques informatiques sur la représentation cartographique du territoire, tout en intégrant les inter-relations parfois discordantes entre les diverses modalités juridique, cartographique et technologique d’un cadastre. Ce cadre théorique reste applicable à diverses époques et en toutes circonstances où l’organisation foncière serait reformée et l’ordre juridique de la propriété serait modifié. D’ailleurs, le renversement de perspective par rapport au cadastre, qui n’est pas qu’une carte thématique, pose des problèmes de représentation en plaçant la présénce sur une modélisation informatique qui impose des contraintes technologiques sur la qualité représentationnelle. Sans nier les avantages des innovations technologiques, l’éthique commande d’admettre que cette transformation radicale des méthodes cartographiques entraîne des conséquences imprévues sur la réalité des gens et sur leurs droits fonciers. Nous endossons l’argument que le cadastre idéal, significatif et performant ne peut se réduire, par définition, à une collection de lots sans lieux, à des entités dans une base de données. Cette image du territoire et de son évolution doit être et demeurer lisible en représentant la structure d’organisation des relations de droit sur les lots fonciers. Pour paraphraser le slogan du présent congrès cartographique international, on dirait que la vision d’avenir trop informatique des concepteurs du cadastre rénové s’est imposée, au détriment des images du passé qui sont pourtant encore bien réelles et significatives.
Object-Oriented Modeling Approach on Geo-Databases and Cartographic Products

Leena Salo-Merta

GIS Analyst, National Land Survey of Finland, Development Centre
P.O. Box 97, FIN-33101 Tampere, Finland.
leena.salo-merta@nls.fi

Abstract

Analysis of geographical information and map production often share interest in the same objects, although the requirements for information may be different. National Mapping Agencies try to serve both needs with geo-databases that originate from map production. Bringing several scales into the system - the concept of multiple representation - makes the situation more complicated, especially when it comes to the updating procedures. Object-modeling and object-oriented GIS environments are assumed to bring new light to the NMAs seek for improved data management in integrated production environments. Flexibility and interoperability are system qualifications that have growing importance.

Defining the Universe of Discourse for a multi-purpose geographical database for topographic data is problematic because of the unpredictable and changing needs of digital data customers. An objectivistic standpoint is proposed for defining a real-world-model, which does not include real-world-processes. The in-house requirements for producing maps and digital products are achieved by completing the real-world model with required features. The in-house process’s specific behavior and object life-cycles should be regarded for implementation of an intelligent and integrated production environment.

This presentation analyses the cartographic extensions that are needed for completing the real-world-model with cartographic requirements. The focus is on topographic maps and their representational requirements to the real-world-data, although the object categories presented are by nature general. A case study on implementation is described. Presented ideas are based on the experiences of Laser-Scan’s object-oriented GIS products GOTHIC and LAMPS2, which were evaluated in the NLS Finland in 1997. They are not applied on production systems of NLS.

Introduction

Object-orientation has been one of the hot topics in software engineering during the last decade. In many references geographical information systems are mentioned as potential application areas for o-o databases and technology, due to the complex nature of geographical information. The principles of o-o have been discussed in the GIS community and o-o GISs have reached a stage where commercial products are available and in operational use [e.g. Clodoveau et al., 1994; Howard et al., 1999]. O-o approaches in system development can be divided into two categories: building a GIS product and building an organizational GIS [Helokunnas, 1995]. More expertise is probably available in the former than in the latter. A full utilization of o-o benefits requires maturity of o-o approaches in the user organization as well and sharing the experiences is a common interest.

Map-sheet based flowline and database driven approach can be considered as two basic approaches to digital mapping [Woodsford and Hardy, 1997]. The emphasis of this paper is on discussing the modeling and design
issues of a geodatabase which is primarily intended for topographic map production, but other uses emerge as well. The ‘case’ is more or less national mapping agency’s (NMA’s), who is undergoing a transition phase from an old production system into the next generation of o-o GIS. Instead of introducing o-o modeling techniques a basic question is asked: where to reach with an o-o model and what advantages may we achieve by using a complex model. The paper is innovated by experiences of o-o GIS products of Laser-Scan Ltd. [Laser-Scan, 1999]. Laser-Scan’s object-oriented geographical environment GOTHIC and LAMPS2 map production software were evaluated in National Land Survey of Finland (NLS) during the winter of 1996-1997. Currently a new production system is being built on Smallworld [Smallworld, 1999] environment.

**Object-oriented approaches on modeling and design of multi-purpose geodatabase**

Nationwide topographic geo-databases can be considered a part of the basic national information infrastructure. The idea of a multi-product, multi-purpose geo-database is tempting in return for the public investment made in data capture. Obviously the geo-database cannot be considered only as raw material for making a certain map, but it should be able to serve the needs of analysis, various tailored outputs and data products as well. However, more qualities are needed both from the data and the application.

**Multi-purpose geo-database in NMA**

As regards topographic data, the background of the databases of NMA’s rely strongly in history with cartography and automation of map production processes, both in modeling the data and as a source. The start-up of the systems have been map production oriented but on the way new customer needs have emerged and the data capture has gained new features that add value to the basic data, such as referencing by street names, addresses and place names, joining route planning and vehicle navigation specific data to the basic geo-spatial data, digital terrain model capabilities etc.

A multi-purpose geo-database is a source of map production and digital data products. There is one master dataset to be updated, from which derived datasets and products are generated more or less dynamically. It is a basic approach to integrated production with multi-product access to the common database, update propagation, automated generalization and multiple representation. These issues have been studied e.g. by Kilpelainen [1997] and Harrie [1998]. Currently the data architecture is based on storing derived copies into separate datasets with no interconnectivities.

In the National Land Survey of Finland the digital topographic map production project was started in early 1990’s with the idea of a multi-purpose database. The capture of the basic data in Topographic Data System (TDS) was started in 1992, as well as the production of the topographic map 1:20 000 using TDS as a source. In 1994 the production of the topographical map 1:50 000 was based on the TDS basic data. In 1995 the revision of the raster version of 1:20 000 map was based on the TDS basic data and became available in the Map Site in the Internet. A review of the NLS’s work in the 1990’s [Jacobsson, 1999] shows that unpredictable customer needs have emerged and therefore some value adding features have been added to the basic data. However, because of the limited record structure of the existing system, new features have been captured in specific project environments. In 1995 the road dataset covered the whole country and a specific copy called the road database was extracted for serving customers of digital road data. In 1996 the capture of road names and addresses was started as a separate project. In 1998 the capture of turn restrictions for roads was started as a separate project. In 1999-2000 the GIS platform is unified by joining the data from the separate project environments into a new system. The ongoing JAKOMAA project uses Smallworld GIS, which is already utilized in cadastral applications. In the first phase the map publishing and generalization tasks are handled by transferring data back into the old system.
The modeling dilemma - where to set the UoD

Analysis of geographical information and map production have often mutual interests in the same geographical objects. However, the requirements for information may be far from each other. The in-house need for the data is the map production, but the customers of digital data have emerging needs of other kind. A multi-purpose geo-database should be prepared for both, although we cannot predict what the needs of the future will be. Data capture for a nationwide database takes years and during that period the needs of the customers change due to new challenges and evolution in technology. Factoring unpredictable user needs into the system design lowers the stage of definability of the system and makes the modeling process more complicated and risky.

A map product or a map production process is only an external view to the schema and does not give a satisfactory result for multi-purpose aims. It is essential to reach beyond the map and model the real world. Nationwide data capture is time-consuming and even the business processes change. Where to set the Universe of Discourse (UoD) when considering a geodatabase for topographic data? The basic idea behind object orientation is that the phenomena the system deals with are simulated in the system. Objects are considered to be more stable than processes that deal with them, which gives ground to a system design with more flexibility. Should we try to simulate the real world? In the FRISCO report Falkenberg et al. [1996] have introduced three philosophical standpoints for modeling in system analysis: nominalistic, objectivistic and constructivistic. The constructivistic approach is generally considered to be the best while the other two are regarded as too simplifying. As far as topographic information is concerned, a typical method used in data acquisition is basically observing the physical appearance of the environment in a given area at a given time. This is close to the objectivistic philosophy in modeling. We model the environment as we can observe it.

Objects encapsulate data and behavior. It is important to distinguish the real-world-specific behavior and in-house process-specific behavior and concentrate on the latter. Some simple rules like a minimum size or an existence dependency (e.g. an obstruction must have a relationship to a road) may be applied, but in general it is better not set rules to the reality or try to model the processes of the environment. That would make the task far too complicated and the cost of data acquisition unbearable. The processes appear to be quite different from each other in most disciplines and applications and attempts towards conformance would probably lead into an analysis paralysis.

A general topographic map can be considered to be a reasonable compromise of customer needs. Despite of the tailoring capabilities the content of a geodatabase of an NMA remains a compromise as well. The actual data capture is the most expensive part and finally it is the resources of the map agency that set limits to the contents, classifications and the quality of the data.

The abstraction levels and object definition - how to define the objects

To use conceptual, logical and physical levels in database design is common practise. The conceptual model defines the UoD. For geographical information, the UoD can be set to map production or it can reach beyond to the real world as discussed above. A real-world model discusses the concepts of the user by applying his own terms. In case of a geo-database, the concepts we are dealing with are lakes, rivers, roads, buildings etc. The nature of real-world entities and their definability vary: some phenomena are discrete and some continuous. The discrete ones are easily picked and digitized but for the continuous type the definition relates to data quality. The level of ‘rows’ complexity is another issue - is there a need to bundle the basic objects into complex geometries and how complex these should be. What kind of relationships and associations are there, between object classes on the one hand and between object instances on the other? When to construct explicit complex or associated objects in the database and when to use topological relationships or spatial adjacency with standard GIS functionality? There are more alternatives to consider than before.
The reasonable definition of an object is connected with the scale we are working on. For example, the impression of Lake Saimaa may be clear on a scale of 1:1 000 000, but on 1:10 000 it consists of sub-regions, some of which clearly belong to the bundle, but at the edges the case may be really frustrating. The motivation and cost of object’s complexity has to be examined. Even though we do not bundle everything together, we have the GIS facilities like topological and coordinate based analysis for querying and ‘building up’ the higher abstraction level of constructed objects.

In database jargon a real-world-object (rwo) is used in the context of any user defined classes. From the database provider’s point of view that is a current definition in comparison with system classes which are more or less invisible to the end-user. Let us redefine a rwo in a way that suits better the higher abstraction level, i.e. as an object in the real world seen from the objectivistic standpoint.

In NLS the TDS is based on a rwo-oriented data dictionary. It reflects the philosophy of multi-product database and a higher abstraction level of modeling. A special feature in the data dictionary is that it doesn’t include cartography and in some respects the abstraction is beyond the implementation capabilities of those days, e.g. place names and explanatory texts of objects are represented as attributes.

Analysis and design pattern approach

The physical model defines how data is stored in the computer - the tools that the system offers for implementing the geographical, non-geographical and topological entities and their relationships. In an o-o model user defined classes get their geometrical and topological properties and behavior by inheriting from the system base classes. Geometries act as properties of objects. Besides, the system has tools for creating complex objects, for example for turning simple geometries into complex ones. User can define methods for adding functionality and intelligence to objects - actually the methods may be seen as an interface for tailoring the database management system.

Implementing a conceptual model into an o-o database is said to be a straightforward task [Kemppainen, 1991]. Anyhow, we have to translate the conceptual model into the database schema, and also model the transformations from the existing systems’ data structures (and probably back as well). We can design a low level object model with a close resemblance to existing data or a high level object model with added relationships and complexity. Implementing a low level model means that practically no changes take place compared to the previous system. The transfer is straightforward, but what are the true benefits we gain from object orientation?

It is a fascinating idea to put more intelligence to the data and use a flat client. We are used to complex working procedures, but would it be possible to make the procedures simpler, if the database had more rule-based intelligence? However, as we try to create a high level model we run into new design problems, which others in the same application area surely are familiar with. This leads us to the design pattern approach, which is popular in the o-o community. Design patterns are documented solutions to common design problems in (o-o) software engineering. The basic motivation is to share the knowledge of previous designs and use them as elements of reusable o-o software [Gamma et al., 1995]. GIS applications of NMAs’ have domain specific problemacy that deal with e.g. cartographic representation, generalization of geographical objects and multiple representation.

Cartographic extensions

The geometrical entity of a real-world object is in some cases inadequate for representation. Cartographic extensions are needed for showing generalized copies of rwos or for extending the cartographic representation on the same level of generalization. In the following model (fig. 1) geographical objects are analyzed in a way
that combines the the real-world-model and the map production process. The approach genuines from the database-driven philosophy and considers what kind of objects the DBMS deals with in respect to the ‘reality’. The basic object classes are real-world-object, generalized real-world-object, associated cartographic object and non-associated cartographic object. Storing the data in separate datasets is not considered in the model.

Figure 1. Real-world-objects and cartographic extensions.

The real world objects (RWO) refer to geographical objects that are observed or measured with a contribution to the real world. Their position, geometry and attributes have significance in the real world and should not be changed for cartographic reasons. RWO is a primary instance of a class in the conceptual model. A Road is an example of an RWO class. It’s geometrical entity is line geometry and it has attributes like Road class, Road name, Road number and Pavement type. A RWO class has defined topological and association rules. A RWO may also have defined behavior, for instance a minimum size or a list of conditions under which objects are allowed or not allowed to exist in the database. For example, an Obstruction is not allowed to exist without association to a Road.

Most RWO classes have a suitable geometry for presentation on a topographic map. However, some special cases occur when the geometry of the RWO is not sufficient for producing the representation, for example if one needs to emphasize an area or line with certain map symbols or show the attribute of an RWO as a text on the map. In the implementation level these cartographic extension objects need to be defined. These associated cartographic objects (ACOs) have attributes and coordinates related to the map, and they have RWOs they are associated to, but they don’t exist in the real-world. Theoretically, the cartographic extensions could be produced automatically, on-the-fly so to say, especially in contemporary and screen maps. What is essential when considering the art of cartography and the quality of representation is the editability of cartographic extension objects and therefore they need to be stored in the database as well.

Generalized cartographic objects (GRWOs) are transformed from RWOs by generalizing. Their geometry and attributes have a contribution to the level of generalization. Kilpelainen [1997] divides the generalization tasks into model generalization, which is essential for maintaining the correct object level relationships and cartographic generalization, which is needed for the final cartographic output. There may be several levels of
generalizations, which is shown as GRWO’s relationship to itself. In this model generalization is treated very loosely and may be interpreted as a real-time generalization by an invoked object method or a separate process. The former means construction of GRWOs when RWOs are created. Because of this the relationship between RWO and GRWO is not defined more precisely. It is only stated that GRWO derives the RWO’s attributes and there are methods for propagating updates to GRWOs. An experiment of real-time multiple representation of buildings using object reflex methods (e.g. constructor and dependency) is described in Kilpelainen’s thesis.

ACO and GRWO objects may be created and associated to RWOs, with no need to duplicate the attributes of RWOs. Instead the attributes of the extension objects are treated as methods which extract their values from RWOs. Certain methods are needed for creating and maintaining the associations between the RWOs and extensions.

The product-related cartographic objects that have no relationships with objects of other are referred to as non-associated cartographic objects (NACOs). Grid lines, coordinates, legends and compass roses are examples of this category.

**Associated cartographic objects on topographic maps**

The o-o technology enables dynamic representation of objects. The application sends a message to the feature object asking it to draw itself. The methods that are invoked can be user defined, i.e. they may use any combination of the object’s attributes plus other conditions like explicitly or spatially defined associations to other objects to determine how to draw the object. This is a major change compared to the look-up-table type representation based on feature codes, and a promising approach in establishing a multi-product dataset [Hardy and Woodsford, 1997].

Dynamic representation suits well the concept described above. If multiple representation is not considered, the types of cartographic extensions in topographic maps of the NLS are fairly simple. In the following (table 1) the ACOs that appear in the Topographic Map 1:20 000 are studied in detail and some examples are given. The extension objects needed are *additional cartographic text* and *symbol*. They have different behavior depending on the RWO’s geometry. Note that there is no need for an associated symbol for a point-rwo. The extension is needed only when the geometry of the RWO is insufficient for intended representation.
Applying the RWO-ACO design

The aim was to provide additional cartographic objects that associate to real-world objects and act as cartographic extensions to the real-world objects. This kind of functionality is also known as ‘live cartography’ [Woodsford and Hardy, 1997]. The objective was the implementation of general tool objects, which can be inherited by other user-defined classes so that the functionality is defined once and reusable thereafter.

An experiment was made for testing the design in NLS using Laser-Scan products Gothic OODB and LAMPS2 map production client. Laser-Scan user API language LULL and Gothic Reference Module Library were used for programming [Laser-Scan 1996, 1997]. The implemented design had a slight difference to the simplified one that is presented in fig. 2.

Table A. RWO-ACO relationships and characteristics in the Topographic Map 1:20 000.

<table>
<thead>
<tr>
<th>Rwo-type/ Cartographic Extension</th>
<th>General Requirements</th>
<th>Point-rwo</th>
<th>Line-rwo</th>
<th>Area-rwo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional Cartographic Text</td>
<td>- Language support is required in some cases - Typesetting code is needed in some cases</td>
<td>- 1:1 relationship - placed near the rwo</td>
<td>- 1:n relationship - text string is extracted from rwo - orientation is extracted from rwo - placed on the rwo - placed near the rwo</td>
<td>- 1:n relationship - placed inside the rwo - placed near the rwo</td>
</tr>
<tr>
<td>Mast height Elevation value of ground control point</td>
<td></td>
<td></td>
<td>Contour height value Road number</td>
<td>Elevation of a water body Explanatory text for a nature reserve</td>
</tr>
<tr>
<td>Additional Cartographic Symbol</td>
<td>- Symbol may be chosen by extracting an attribute from rwo</td>
<td>(None)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water flow direction Pipeline classification symbol</td>
<td></td>
<td></td>
<td>Water flow direction Pipeline classification symbol</td>
<td>Quarry symbol Mineral resources extraction symbol</td>
</tr>
</tbody>
</table>
Multiple inheritance and object reflex methods were key concepts in the design. New tool classes were created to define properties and methods for required functionality which is applicable thereafter by inheritance. The associations between extensions and RWOs were created and maintained using object reflex methods, which are invoked by the ODBMS. Following classes were used to create the functionality needed for an ACO: ACO_symbol and Has_ACO_symbol and ACO_text and Has_ACO_text. The ACO_* class has the methods required for creating the association to a RWO and validating that the association exists. It also defines the bi-directional references between RWO and ACO classes which are named using a certain naming convention. The user-defined ACO class is defined by inheriting from this class. The Has_ACO_* classes hold the methods for maintaining the association between the ACO and RWO and deletes the ACO if RWO is deleted. This class is inherited by the user defined RWO class, e.g. a quarry. The conflict of multiple inheritance of object reflex methods is solved by invoking all inherited methods by the given name.

**Advantages and disadvantages for applying cartographic extensions**

Associating cartographic extensions with RWOs make the database design elegant because it enables the ODBMS to take care of the real-world entity and its extensions as one integrated unit and guarantees the database’s
integrity in this respect. On the other hand, it adds complexity to the system. This appears in two connections: Firstly, setting up association and validation rules forces us to follow more strict working procedures. Because of the precedence rule adding a dependent object is allowed only if the base object is present in the database. A two-phase policy could be useful in object validation, by applying looser rules within a transaction and tighter ones in the commit phase. Secondly, restructuring the data that is already captured with the current system means check-ups and corrections. We can skip the re-organizing costs on the existing data by setting the validation rules loose but in that case the benefits of setting them at all are minimal.

On the other hand, there are only few ACOs needed on the topographic map, and for some of them annotation-functionality may be applied. The number of the attribute type properties in the different classes is rather low in general, and it is especially low in the instance level. In many cases the ACO related attributes are gathered only when they appear on the map, instead of full coverage. The RWO-ACO approach might be more significant in an application, where more attribute data is available and there are changing needs for cartographic output.

Discussion - Map production model meets RWO model?

The database driven approach to digital mapping becomes essential when we think about tailored and digital products and data services in general. Bringing several scales into the system - the concept of multiple representation - makes the situation more complicated, especially when it comes to the updating procedures. Whether the cost of building up such a fully-automated integrated system ever becomes profitable, some consideration of the organization of the base data and the derived versions is useful.

Previously it has been stated that for a multi-purpose geo-database we should build a real-world oriented model which treats the topographic data in an objectivistic way. The in-house specific processes and object life-cycles should be included in the model in a manner that doesn’t conflict with the real-world model. The model is topologically sound, but does not cover much behavior or rules on the real-world level. Instead the object relationships, behavior and life-cycles in the in-house processes are defined and applied. The presented analysis pattern of RWOs and cartographic extensions emphazises that we do not only have geographical and non-geographical object classes, but we have classes of different properties and behavior regarding to whether they are true RWOs, generalizations or purely cartographic objects. This may be a useful standpoint when considering the process integration and multi-purpose database approach.

Futher research and comparison between intelligent DBMS approach and process oriented approaches would be interesting. Defining and implementing the object life cycles and object identity globally in the system - not only in a subsystem - is another important issue in process integration and data management.

NMAs are seeking for ways to improve data quality and services based on geo-spatial data. So far the customer needs have been served by making the topological quality and features that are included in the data somewhat more than would be necessary for just making maps. The classifications that are used are based on real-world qualities rather than graphical output. Metadata qualities are improved, objects are time-stamped and the source and positional accuracy are recorded. For emerging needs, production is re-directed to certain areas or themes of greater interest.Value adding qualities are joined to the basic geodata in separate projects if necessary.

The actions of NLS Finland in 1990’s show a cycle starting from one source database, dividing into new branches of emerging customer needs, and finally unifying together on a new GIS platform. The cycle may reappear: It may again become necessary to react by creating new branches to the production line - and later on either cut them off or unify them in the basic system in one way or another. GIS is utilized more and more outside the traditional map-user community. Co-operation and data exchange between authorities is (or at least should be) increasing and the techniques for detecting changes in the data and propagating updates from one system to another are becoming actual. Applying the international standards like ISO/TC 211 is inevinent.
Abstract

A description is given of tests made to determine whether map complexity can be quantified by data compression ratios applied to the map image. Simple data compression techniques such as run length encoding have been applied to binary maps with results showing that map intricacy and complexity is directly related to compression ratios achieved. Data compression methods can be divided into ‘lossy’ methods, which lose data when compressing and re-constructing images, and ‘lossless’ methods, which ensure that images can be re-constructed to exactly the same content and specification after compression. An examination of a number of lossless compression methods, which are more appropriate for map data, has revealed some similarities between map complexity indices as previously determined (including entropy measures) and data compression ratios.

Introduction

The purpose of this paper is to describe some experimental work undertaken on simple topographic mapping with the intention of assessing the relative complexity of various maps. It is essential to initially re-iterate the reasons for undertaking such an exercise, attempt to define the meaning of complexity as used in this work and to outline the methods applied in estimating a measure of it.

Complexity impinges on cartographic practice in both map production and map use activities. It is suggested that the factors indicated in Table 1 justify the study of complexity and its measurement, and that an understanding of some important practical cartographic tasks can be achieved by considering complexity.

<table>
<thead>
<tr>
<th>Cartographic task influenced by complexity</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addressing variation in map type</td>
<td>Ensuring ‘progression’ in education by supplying less complex maps to younger map users</td>
</tr>
<tr>
<td>Map generalisation</td>
<td>Determining whether map images should increase, reduce or maintain complexity on scale reduction</td>
</tr>
<tr>
<td>Assessing reality</td>
<td>Using map complexity as a measure of the nature of the ‘real world’ as portrayed by the map</td>
</tr>
<tr>
<td>Using information science in cartography</td>
<td>Calibrating models of cartographic communication by quantifying the amount of ‘information’ on a map</td>
</tr>
<tr>
<td>Map data conversion</td>
<td>Comparatively predicting the resources required (time and labour) to undertake digitising of different map sheets</td>
</tr>
<tr>
<td>Map and dataset revision</td>
<td>Assessing future workloads in revising ‘complex’ and ‘less complex’ areas in a topographic map series</td>
</tr>
<tr>
<td>Optimising map design</td>
<td>Avoiding clutter and optimising amount of detail portrayed</td>
</tr>
<tr>
<td>Comparing maps and images</td>
<td>Comparing outputs from different photo-interpreters to assess ‘best practice’</td>
</tr>
</tbody>
</table>

Table 1. The possible impact of complexity studies on cartographic tasks
The term ‘complexity’ is, unfortunately, difficult to define precisely. From a cartographic viewpoint, one could denote at least two specific meanings, one related to the intrinsic complexity of the subject matter of the map (‘intellectual complexity’), the other describing the graphical complexity of the marks on the piece of paper or on the computer screen (‘structural complexity’). It may be that no clear distinction is possible between these two, although a preliminary consideration might conclude that the complexity of the subject matter is related to the cognitive processes of understanding the map, whilst graphical complexity is more closely related to the visual impact of the map and to the perceptual processes of viewing it. It would be useful to be able to determine the nature of the complexity of a map unambiguously, but the uncertain separation between cognitive and perceptual activities makes this extremely difficult.

In a broader sense, contemporary and populist approaches to the definition of complexity tend to assume that the archetypal complex system is dynamic, and probably connected with life sciences in some way. The perceived complexity of such dynamic systems is dependent on the actions and operations which they exhibit (rather than on any intrinsic and unchanging property which characterises their components), the way in which those components are put together, or the intricacy of their form. A further dichotomy is therefore apparent – complexity could describe the workings, the task, the function of a system; or it could address its organisation, design, and structure.

From a mapping perspective, structural attributes such as organisation and design can be considered as core components in defining complexity, although the tasks and functions of the map also have a role to play. This research examined the structure more closely than the function, mainly because the factors contributing to structure are more concrete, capable of being measured and are less variable.

The cartographic factors which are considered to contribute to the structure of a map, and hence to its complexity are many. Most notably, there are different constituents in raster and vector data which need to be considered in very different ways. Within this sub-division, there are components such as (in vector) nodes, connections and line lengths and (in raster) clumps of pixels, runs of pixels and frequencies of pixels.

After determining what was being measured a number of possible candidate measures were investigated. The relative merits of each of these metrics has already been reported to the ICA [Fairbairn, 1997] and a list of potential indices of complexity has been created (Table 2). The metrics chosen vary in terms of their mathematical rigour (from complex manipulation of probabilities, to assessment of file size), their applicability to different data structures (i.e. raster or vector), and their means of calculation (using various software packages, manually counting occurrences (for example of nodes and arcs), or writing specialist programs). In addition, it is suggested that the structure of a map can be denoted either by the intricacy of patterns and the arrangement of features in the map space, and/or by the nature of line elements and the topological structure of the objects on the map face. Different complexity indices are appropriate for each of these approaches.

<table>
<thead>
<tr>
<th>Index category</th>
<th>‘Connectionist’ approach</th>
<th>Pattern intricacy</th>
<th>Raster data</th>
<th>Vector data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>File size</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Graph theory measures</td>
<td>✓/-</td>
<td>✓/✓-</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Spatial statistics measures</td>
<td>✓/-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fractal dimension</td>
<td>?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data compression</td>
<td>-</td>
<td>✓/✓-</td>
<td>✓</td>
<td>?</td>
</tr>
</tbody>
</table>

✓ strong relationship; - some relationship; ? uncertain relationship

Table 2 Potential complexity metrics
Because the range of complexity metrics applicable to raster data appears to be wider, the sample maps which were chosen for testing were scanned as raster images, Table 3 indicating their characteristics. In addition, Table 3 shows the results of using the IDRISI software and a public domain package called Fragstats to measure the parameters chosen.

After each metric had been calculated for the range of maps indicated in Table 3, a subsequent principal components analysis was applied, using the SPSS statistical package (results in Table 4). Some highly correlated metrics were omitted from this analysis, but of the more independent variables, the one-dimensional run length encoding measure was revealed as an important parameter in describing the variance within the metrics. As this is an elementary form of data compression, it was decided to more closely examine the potential of other data compression methods to note their utility, not in efficiently compacting data, but, as a proxy, to measure map complexity. The meaning of the principal components revealed is difficult to judge, as it is drawn from an extremely small sample and is subject to considerable personal interpretation. In this case it is suggested that component 1 can be interpreted as relating to the ‘uniformity’ of elements on the map face: it is strongly positively linked to the Moran autocorrelation measure and negatively related to run length encoding and the fractal dimension. Component 2 is connected with the ‘graphical arrangement’, being positively linked to the black/white ratio, to the entropy index and to run length encoding reduction. Component 3 on the other hand appears to summarise the ‘real world arrangement’ of the pixels on the map face: it is positively related to the average size of pixel clumps and their standard deviation, whilst being negatively linked with the actual size of clumps, and their standard deviation, in hectares in reality.

Table 3. Sample binary maps and their complexity metrics

<table>
<thead>
<tr>
<th>File name</th>
<th>Reduced size (%) using RLE</th>
<th>Black/white ratio</th>
<th>Mean patch size (pixels)</th>
<th>Mean patch size (hectares)</th>
<th>Patch size standard deviation (pixels)</th>
<th>Patch size standard deviation (hectares)</th>
<th>Landscape shape index</th>
<th>Double log fractal dimension</th>
<th>Shannon's diversity index</th>
<th>Simpson's diversity index</th>
<th>Shannon's evenness index</th>
<th>Simpson's evenness index</th>
<th>Moran's I (King case)</th>
<th>Moran's I (Rook case)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERNE1</td>
<td>30.40</td>
<td>0.59</td>
<td>252</td>
<td>0.525</td>
<td>5659</td>
<td>10.141</td>
<td>46.909</td>
<td>1.452</td>
<td>0.66</td>
<td>0.467</td>
<td>0.952</td>
<td>0.934</td>
<td>0.61</td>
<td>0.667</td>
</tr>
<tr>
<td>BERNE2</td>
<td>33.11</td>
<td>0.517</td>
<td>115</td>
<td>0.827</td>
<td>2071</td>
<td>14.844</td>
<td>25.41</td>
<td>1.42</td>
<td>0.641</td>
<td>0.449</td>
<td>0.925</td>
<td>0.898</td>
<td>0.575</td>
<td>0.629</td>
</tr>
<tr>
<td>BERNE3</td>
<td>35.15</td>
<td>0.723</td>
<td>146</td>
<td>4.185</td>
<td>1019</td>
<td>29.212</td>
<td>14.064</td>
<td>1.567</td>
<td>0.68</td>
<td>0.487</td>
<td>0.981</td>
<td>0.974</td>
<td>0.575</td>
<td>0.633</td>
</tr>
<tr>
<td>DELTA1</td>
<td>16.53</td>
<td>0.197</td>
<td>2969</td>
<td>1.33</td>
<td>55019</td>
<td>24.65</td>
<td>59.613</td>
<td>1.684</td>
<td>0.447</td>
<td>0.275</td>
<td>0.645</td>
<td>0.55</td>
<td>0.624</td>
<td>0.66</td>
</tr>
<tr>
<td>DELTA2</td>
<td>16.24</td>
<td>0.226</td>
<td>7397</td>
<td>13.256</td>
<td>42452</td>
<td>76.079</td>
<td>30.274</td>
<td>1.681</td>
<td>0.478</td>
<td>0.301</td>
<td>0.69</td>
<td>0.601</td>
<td>0.621</td>
<td>0.682</td>
</tr>
<tr>
<td>DELTA3</td>
<td>13.02</td>
<td>0.197</td>
<td>3853</td>
<td>27.619</td>
<td>15642</td>
<td>112.129</td>
<td>12.764</td>
<td>1.576</td>
<td>0.447</td>
<td>0.275</td>
<td>0.645</td>
<td>0.55</td>
<td>0.675</td>
<td>0.726</td>
</tr>
<tr>
<td>GEN1</td>
<td>2.87</td>
<td>0.342</td>
<td>30385</td>
<td>0.134</td>
<td>50793</td>
<td>0.224</td>
<td>6.987</td>
<td>1.44</td>
<td>0.567</td>
<td>0.38</td>
<td>0.816</td>
<td>0.759</td>
<td>0.931</td>
<td>0.961</td>
</tr>
<tr>
<td>GEN2</td>
<td>2.43</td>
<td>0.433</td>
<td>31973</td>
<td>0.141</td>
<td>63946</td>
<td>0.282</td>
<td>5.25</td>
<td>1.357</td>
<td>0.613</td>
<td>0.422</td>
<td>0.884</td>
<td>0.844</td>
<td>0.958</td>
<td>0.977</td>
</tr>
<tr>
<td>GEN3</td>
<td>1.54</td>
<td>0.589</td>
<td>53329</td>
<td>0.244</td>
<td>76458</td>
<td>0.346</td>
<td>3.756</td>
<td>1.403</td>
<td>0.659</td>
<td>0.467</td>
<td>0.951</td>
<td>0.933</td>
<td>0.969</td>
<td>0.981</td>
</tr>
<tr>
<td>WOOD1A</td>
<td>3.08</td>
<td>0.599</td>
<td>11749</td>
<td>5.264</td>
<td>114442</td>
<td>51.273</td>
<td>14.171</td>
<td>1.276</td>
<td>0.661</td>
<td>0.469</td>
<td>0.954</td>
<td>0.937</td>
<td>0.999</td>
<td>0.949</td>
</tr>
<tr>
<td>WOOD2A</td>
<td>4.79</td>
<td>0.651</td>
<td>3376</td>
<td>6.05</td>
<td>28605</td>
<td>51.264</td>
<td>11.887</td>
<td>1.291</td>
<td>0.671</td>
<td>0.478</td>
<td>0.967</td>
<td>0.955</td>
<td>0.934</td>
<td>0.953</td>
</tr>
<tr>
<td>WOOD3A</td>
<td>6.55</td>
<td>0.662</td>
<td>2608</td>
<td>18.693</td>
<td>12442</td>
<td>89.192</td>
<td>8.379</td>
<td>1.27</td>
<td>0.672</td>
<td>0.479</td>
<td>0.976</td>
<td>0.955</td>
<td>0.915</td>
<td>0.929</td>
</tr>
<tr>
<td>WOOD4A</td>
<td>8.00</td>
<td>0.82</td>
<td>2119</td>
<td>60.754</td>
<td>4980</td>
<td>142.804</td>
<td>5.404</td>
<td>1.296</td>
<td>0.688</td>
<td>0.495</td>
<td>0.993</td>
<td>0.993</td>
<td>0.907</td>
<td>0.929</td>
</tr>
</tbody>
</table>

BERNE1: 1:50,000 urban; BERNE2: 1:100,000 urban; BERNE3: 1:200,000 urban; DELTA1: 1:25,000 mountain contours; DELTA2: 1:50,000 mountain contours; DELTA3: 1:100,000 mountain contours. GEN1: 1:2500 urban; GEN2: 1:2500 weakly generalised GEN1; GEN3: 1:2500 strongly generalised GEN1. WOOD1A: 1:25,000 woodland areas; WOOD2A: 1:50,000 woodland areas; WOOD3A: 1:100,000 woodland areas; WOOD4A: 1:200,000 woodland areas.
### Table 4  Principal components analysis

<table>
<thead>
<tr>
<th>Complexity metric</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced size (%) using 1D RLE</td>
<td>-.544</td>
<td>.582</td>
<td>.265</td>
</tr>
<tr>
<td>Black/white ratio</td>
<td>.600</td>
<td>.627</td>
<td>.257</td>
</tr>
<tr>
<td>Mean patch size (pixels)</td>
<td>.554</td>
<td>-.508</td>
<td>.481</td>
</tr>
<tr>
<td>Mean patch size (hares)</td>
<td>.469</td>
<td>.489</td>
<td>-.627</td>
</tr>
<tr>
<td>Patch size standard deviation (pixels)</td>
<td>.582</td>
<td>-.455</td>
<td>.435</td>
</tr>
<tr>
<td>Patch size standard deviation (hares)</td>
<td>.485</td>
<td>.476</td>
<td>-.614</td>
</tr>
<tr>
<td>Landscape shape index</td>
<td>-.648</td>
<td>.399</td>
<td>.442</td>
</tr>
<tr>
<td>Double log fractal dimension</td>
<td>-.518</td>
<td>.370</td>
<td>.449</td>
</tr>
<tr>
<td>Shannon’s diversity index</td>
<td>.537</td>
<td>.617</td>
<td>.534</td>
</tr>
<tr>
<td>Morar’s (King case) spatial autocorrelation</td>
<td>.890</td>
<td>.163</td>
<td>.331</td>
</tr>
<tr>
<td></td>
<td>35.26%</td>
<td>23.67%</td>
<td>21.18%</td>
</tr>
</tbody>
</table>

3 components revealed

### Data compression

The study of data compression has been longstanding in the discipline of computer science. In many cases when handling digital data, storage space is at a premium or there is a requirement to most efficiently transmit large files of computer compatible data. Data compression techniques are familiar to many as methods of overcoming these problems. The use of ‘file-zipping’ software, for example PKZip and its derivative WinZip, is standard within the PC environment, and there are equivalent standard tools in other operating environments. In the disciplines of cartography, mapping, and satellite and aerial remote sensing, it is the contention of the author that more voluminous material is handled than in virtually any other subject area (with the possible exception of meteorology). Data file compression techniques are necessary and are already in widespread use in GIS and image handling software, and the notable incorporation of the much-publicised MrSID image compression software into ArcView, Intergraph, and Cities Revealed packages and datafiles has led to considerable interest.

### Methods of data compression

The most significant aspect of the many data compression algorithms which exist is the fact that they minimise the redundancy within a dataset such that extraneous data is discarded, whilst the data which describes the essential character of the dataset is retained. Redundant data is that which can be predicted, involves repetition, or is pre-known. Additionally, data which occurs infrequently within a dataset (i.e. it has a low probability of being encountered) might also be discarded. The assessment of redundancy is a statistical process, which has links to the entropy of the datafile. The initial methods of data compression were developed by Shannon, the prime investigator of data communication, information content and entropy measurement. The Shannon-Fano coding method relied on a preliminary calculation of the probability of the symbols in a message to be communicated, and a subsequent encoding of those symbols by giving short codes (economic in terms of file space) for the most common symbol and longer codes for the infrequent element. The subsequent development of Huffman coding followed similar lines, and although a radically different technique called arithmetic coding became more popular soon after its introduction in 1979 [Witten, et al., 1987], it too relies on knowledge of the probabilities of the component symbols in a message.

Alternatively to such ‘statistical’ methods, ‘structural’ methods can be considered, which rely on the order or distribution of the symbols in a message. One-dimensional run length encoding takes a run of repeating
(uniform) symbols in a message and summarises them to an indication of the position of starting the run and the length of that run. Clearly, uniform messages will reveal considerable potential for such compression. ‘Dictionary-based’ methods develop the idea of strings of symbols further by giving shorter codes for commonly occurring sequential combinations of (non-uniform) symbols.

All the methods described so far are ‘lossless’, as information additional to the coded and compressed data is held and attached to the compressed file to allow for decoding to exactly the original form. Such information includes probabilities, codes and dictionary items which are necessary to ensure a matching of the encoding and decoding process. For short data files, such an overhead will ensure that the compression is likely to be inefficient (i.e. the ‘compressed’ file may actually be larger than the original file). However, for larger data files, including image files, such as scanned maps, it is suggested that comparative assessments of the rate of data compression can be profitably made.

Improvements in compression rates can be made by applying adaptive techniques which progressively scan the datafile and alter the probabilities, codes and dictionary entries as new data is encountered. Adaptive techniques are commonly applied to the statistical and dictionary based procedures. Further, it is possible to combine methods such that, for example, a dictionary based compression is then followed by arithmetic coding. Clearly, such routines are more complex to encode but, although the overhead in sending decoding data with the compressed file may be increased, most adaptive techniques are capable of higher compression rates.

Most current research into data compression is directed towards lossy methods. These are most appropriate for graphics (especially photographic) and sound files which can accept the slight loss of quality of data which such methods entail on decoding. The inherent data degradation may not be noticeable to a human recipient which is why lossy compression can be widely applied to data sensed visually or aurally. It is particularly appropriate to use lossy methods, such as JFIF (the JPEG File Image Format), to encode pictures transmitted over the Internet: the resolution of an original photograph scanned at 300 dots per inch can be safely reduced to allow for rendering on a standard computer monitor at approximately 90 pixels per inch without the human viewer noticing the degradation. Contemporary wavelet methods give similar results, although the MrSID routines already mentioned do hold graduated levels of the image which are displayed at an appropriate resolution dependent on the zoom factor. Along with text-files, databases and programs, however, machine analysis of data, which will pick up the most minimal data variation, is likely to require losslessly compressed data. This suggests that the research described here, which deals with manipulating and assessing digitally encoded data, should address only lossless algorithms.

**Applying data compression to sample mapping**

In addition to the map extracts already considered, some further examples were chosen, exemplifying different types of topographic area and different methods of holding the scanned raster data. A series of simple demonstration programs were prepared to test the sample maps chosen. Statistical methods (e.g. Huffman coding), structural methods (e.g. one-dimensional run length encoding and two-dimensional run length encoding) and dictionary based methods (e.g. LZ77) were applied using a mixture of program code sourced from commercial software, from published literature and individually written. The intention was to determine both the extent to which compression could be undertaken, reflected by the various compression ratios in Table 5, and the relationship between such ratios and the other complexity measures already obtained.
Conclusions

Images with a simple structure (for example, the binary images) clearly benefit from statistically based coding methods. Run-length encoding methods, however, also exhibit good results for binary data whilst the compression of the grey scale photographic-type images and coloured mapping is less efficient. This is to be expected as grey scale data is recognised as being more effectively encoded bit-plane by bit-plane, rather than pixel by pixel.

<table>
<thead>
<tr>
<th></th>
<th>Huffman Program HUFF.CPP [statistical]</th>
<th>LZ coding Program LZ77_SS.CPP [dictionary]</th>
<th>1-D RLE (implemented in IDRISI) [structural]</th>
<th>2-D RLE THISONE.CPP [structural]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERNE1</td>
<td>12.72</td>
<td>12.77</td>
<td>30.40</td>
<td>40.69</td>
</tr>
<tr>
<td>BERNE2</td>
<td>13.39</td>
<td>12.92</td>
<td>33.11</td>
<td>43.33</td>
</tr>
<tr>
<td>BERNE3</td>
<td>16.08</td>
<td>13.10</td>
<td>35.15</td>
<td>46.26</td>
</tr>
<tr>
<td>DELTA1</td>
<td>12.55</td>
<td>12.23</td>
<td>16.53</td>
<td>25.04</td>
</tr>
<tr>
<td>DELTA2</td>
<td>12.70</td>
<td>12.21</td>
<td>16.24</td>
<td>25.66</td>
</tr>
<tr>
<td>DELTA3</td>
<td>13.30</td>
<td>12.10</td>
<td>13.02</td>
<td>21.54</td>
</tr>
<tr>
<td>GEN1</td>
<td>12.63</td>
<td>11.83</td>
<td>2.87</td>
<td>5.57</td>
</tr>
<tr>
<td>GEN2</td>
<td>12.63</td>
<td>11.82</td>
<td>2.43</td>
<td>4.56</td>
</tr>
<tr>
<td>GEN3</td>
<td>12.63</td>
<td>11.82</td>
<td>1.54</td>
<td>3.06</td>
</tr>
<tr>
<td>WOOD1A</td>
<td>12.52</td>
<td>11.86</td>
<td>3.08</td>
<td>4.98</td>
</tr>
<tr>
<td>WOOD2A</td>
<td>12.60</td>
<td>11.90</td>
<td>4.79</td>
<td>7.41</td>
</tr>
<tr>
<td>WOOD3A</td>
<td>12.88</td>
<td>11.93</td>
<td>6.55</td>
<td>9.65</td>
</tr>
<tr>
<td>WOOD4A</td>
<td>14.02</td>
<td>11.98</td>
<td>8.00</td>
<td>11.55</td>
</tr>
<tr>
<td>OGLE1</td>
<td>12.71</td>
<td>12.01</td>
<td>6.79</td>
<td>10.28</td>
</tr>
<tr>
<td>OGLE2</td>
<td>12.71</td>
<td>12.05</td>
<td>7.61</td>
<td>10.69</td>
</tr>
<tr>
<td>OGLE3</td>
<td>12.71</td>
<td>12.07</td>
<td>8.27</td>
<td>12.30</td>
</tr>
<tr>
<td>OGLECOL</td>
<td>18.62</td>
<td>20.33</td>
<td>35.23</td>
<td>34.58</td>
</tr>
<tr>
<td>GATES1</td>
<td>12.71</td>
<td>12.37</td>
<td>27.46</td>
<td>36.41</td>
</tr>
<tr>
<td>GATES2</td>
<td>12.71</td>
<td>12.61</td>
<td>69.92</td>
<td>55.05</td>
</tr>
<tr>
<td>GATES3</td>
<td>12.71</td>
<td>12.43</td>
<td>28.46</td>
<td>38.48</td>
</tr>
<tr>
<td>GATES4</td>
<td>87.82</td>
<td>102.10</td>
<td>186.34</td>
<td>188.61</td>
</tr>
<tr>
<td>GATECOL</td>
<td>28.92</td>
<td>26.05</td>
<td>69.68</td>
<td>76.35</td>
</tr>
<tr>
<td>OS_CUMB</td>
<td>17.61</td>
<td>13.82</td>
<td>18.37</td>
<td>24.55</td>
</tr>
<tr>
<td>BARTS_CUMB</td>
<td>16.76</td>
<td>13.07</td>
<td>14.68</td>
<td>20.46</td>
</tr>
</tbody>
</table>

BERNE1, BERNE2, BERNE3, DELTA1, DELTA2, DELTA3, GENT, GEN2, GEN3, WOOD1A, WOOD2A, WOOD3A, WOOD4A: all as above. OGLE1: rural 1:50,000 binary; OGLE2: rural 1:50,000 binary with screen tints; OGLE3: rural 1:50,000 binary with text; OGLECOL: rural 1:50,000 4-bit colour with screen tints and text; GATES1: urban 1:50,000 binary; GATES2: urban 1:50,000 binary with screen tints; GATES3: urban 1:50,000 binary with text; GATES4: urban 1:50,000 8-bit grey scale with screen tints and text; GATECOL: urban 1:50,000 4-bit colour with screen tints and text. OS_CUMB: 1:250,000 4 bit spot-colour rasterised line map. BARTS_CUMB: 1:250,000 4 bit spot-colour rasterised line map (same area as OS_CUMB, different mapping agency).

Table 5 Compression rates of sample maps (all numbers are reduced sizes in %)

Clearly, some compression methods use common data constructs – for example, Huffman codes or dictionary strings – and thus give similar results for maps which vary in scale, but, in general exhibit similar binary structures (the figures for Huffman coding and LZ coding are remarkably similar for all images) and structural intricacy. The fact that many of the compression ratios are identical is indicative of the ways in which symbols (in the case of many of these map extracts merely black or white pixels) are handled by Huffman methods, or in which strings of variable data (but in the case of this program, of standard length) are handled by the LZ algorithm.

The Huffman coding procedure used is not adaptive, so a binary representation will always yield similar results, no matter what the size or intricacy of the image. The major variation is in the associated ‘header’
information which holds information needed for the decompression. The notable difference in the results of Huffman coding of GATES4, GATECOL and OGELCOL results from handling images with 16 or 256 colour levels, rather than binary data.

The RLE method reveals more variable results, reflecting as it does the arrangement of pixels rather than summarising their frequency: this in itself may indicate that it is potentially the more useful compression method to study in order to order map complexity. The most compressible images of the original binary maps chosen (GENx) are somewhat contrived maps (each the same scale but with different generalisation levels), but high levels of compressibility were also obtained for standard topographic products exemplified by OGLEx and WOODxA. The latter is the least intricate of the maps and gives the most efficient RLE. The figures for two-dimensional run length encoding are disappointing, indicating that it does not contribute to more efficient data compression, as had been initially thought. In fact, such a method works more effectively on more regular raster images with overall pixel dimensions related to powers of two and internal constructs with similarly measured pixel dimensions.

For the structural methods, the impact of grey scale and coloured representations is once again clear. In addition, the complex urban areas of Berne (Switzerland) and Gateshead (England) exhibit the least efficient RLE. By further analysing Table 3, it appears that the complexity metric with the closest correlation to RLE is, unsurprisingly, that which directly measures the clumping and arrangement of pixels – mean patch size in pixels. Also of interest is the close link with rook’s case spatial autocorrelation, which also summarises uniformity in the direction of the run length encodings. The links to other measures, notably those characterising entropy – the Shannon and Simpson indices – are not so clear, and for the real world raster maps (BERNEx, WOODxA, OGLEx and GATESx) it seems that there is no direct link between the entropy value and any of the file compression measures.

In general, Table 5 illustrates that the sample maps chosen, even of different scales and intricacy, exhibit considerable uniformity when subjected to compression using standard statistical and dictionary based techniques. It is only when images of a different type (grey scaled, coloured) are considered that significant variation can be revealed. Structural methods, however, do reflect the variance within the map extracts to a greater degree.

References


Modelling Objects for Screen Map Design

Mats Dunkars
VBBVIAK
100 26 Stockholm, Sweden
mats.dunkars@sweco.se

Anna Bergman
University of Gävle
Kungsbäcksvägen 47, 801 76 Gävle, Sweden
anna.bergman@hig.se

Abstract
In object oriented modelling of geographical information the modelling process is often seen as being separate from the map design process. This paper proposes a method to model geographical information where the map design is considered to be an inseparable part of the modelling process. This viewpoint is motivated through a discussion on how the human brain stores and expresses knowledge and how categories in geographical information are formed and expressed. The method is described with focus on how the data model for a new screen map is defined and implemented in an object oriented database. The object classes of the new map are equipped with functions that create objects in the new map through retrieving information from other parts of the database and performing an automatic generalisation to create objects suitable for this particular screen map. The creation process of the new screen map tries to simulate how a cartographer works when creating a new small scale map from maps at larger scales.

Introduction
Maps in a geographical information system are usually presented on a computer screen and can be seen as a part of a user-interface that gives access to information stored in a geographical database. This user-interface provides: a visualisation of the geographical database - the map, ability to retrieve attributes from the database and functionality to perform analysis of the information stored in the database. To a certain extent the user interface might also provide functionality to modify the visualisation of the geographical information, e.g. through changing content or colours. The work presented in this paper concerns how such a user interface might be designed and focus on providing tools for the design of the user interface. A major issue is how the geographical information is to be generalised to give the most appropriate presentation in a user interface designed for a particular application. In this paper a data model is defined for each user interface since the design of the user interface does not only concern the map but also analysis and access of attribute information. The generalisation of information to suit a particular user interface is then a matter of transformation of information from one data model to another.
Assumptions

The design of the user interface is to a large extent done through object oriented modelling of the geographical information using the Unified Modelling Language, UML. The model is then implemented in an object oriented database. A central concept in object oriented modelling is abstraction. According to Rumbaugh et. al. [1991] "Abstraction is a selective examination of certain aspects of a problem." An abstraction is performed to isolate the most important aspects of a problem and neglecting the unimportant. Rumbaugh et. al. [1991] states that: “All abstractions are inaccurate and incomplete.” How geographical information is abstracted into an object oriented data model depends on more or less well defined assumptions about the characteristics of the information. Some assumptions are based on the Aristotelian theory on categories which states that categories exists independently of human beings and all we have to do is to discover and define them. According to MacEachren [1995] the Aristotelian theory has two basic tenets:

1. Categories are like containers with individual objects being either inside or outside the container
2. Individual objects are assumed to be in the same category if and only if they have certain properties in common.

The Aristotelian view on categories was not questioned until the mid 20th century and does still have a significant impact in science. During the last decades a different view has developed which states that categorisation is a tool used by the human brain to store and express knowledge [Lakoff 1987]. Object oriented modelling of geographical information is influenced by the Aristotelian view in a sense that we assume that we can discover the perfect or true categories to describe the geographical information and model these as object classes. These object classes can be defined without any concern about the visual appearance of object classes in maps. Nyerges cited in Buttenfield [1995] for instance defined …the non-graphical digital data model that contains spatial and attribute relationships as ‘deep structure’ and distinguished this from the graphic image which he calls ‘surface structure.’ According to the more modern theories about categorisation, categories evolve through experience and since the categories used in geographical information has evolved through the use of maps as well as real world experience it seems appropriate to assume that the graphical image would provide important knowledge when the geographical data model is defined. For some reason this knowledge is considered to be of minor importance and we assume that this depends on the Aristotelian influence. This will be described further below.

When using object oriented modelling to design a user interface for a particular application the information presented through the user interface has to be abstracted in a way that is most suitable for this application. This abstraction should be based on theories about how the human brain stores and expresses knowledge and the general theories on object oriented modelling as described by e.g. Rumbaugh et al. [1992].

Human Knowledge

Humans has the ability to recognise 7 million different shades of colours [MacEachren 1995]. To be able to reason about the colours the human brain groups them into categories. Lakoff [1987] gives a detailed description of how the human brain forms categories which includes for instance such topics as fuzzy categories where the distinction between different categories is not crisp. An example in geographical information is whether a building that has been used as a farm until the 1950’s but is used as a dwelling house today shall be mapped in the category farm or dwelling house. For a tourist map it might be appropriate to model it as a farm but in a database for taxation it is more suitable to see it as member of the dwelling house class. Generally it can be said that entities built by a human institution such as the National Road Administration are easy to map into explicit categories while natural objects such as streams are more difficult. To define a motorway as
having separate roadways for meeting traffic is a much more sufficient definition than to define a stream mapped as a line as flowing water with a maximum width of seven meters. There will always be short stretches on the stream that are wider than seven meters that will be mapped as a line.

Another characteristic of how categories are formed is that they are ordered in hierarchies. There is a basic level category from which we can generalise or specialise to other categories. Chair for instance is a basic level category while furniture is a generalisation. The basic level categories are, among other things, recognised through a similar overall perceived shape that can form a single mental image. The mental image is a form of knowledge representation used by the brain to store information. There are several theories on how such knowledge representations are organised and stored. MacEachren [1995] describes a theory offered by Rumelhart and Norman [1985] where knowledge representations or schematas are divided into three kinds:

**Propositional schemata** - are used to organise declarative knowledge, or knowledge about objects, attributes and places. Schemata can be seen as “packets of information that contain variables.” The variables can be fixed or variable where the fixed parts correspond to such knowledge as is usually true and the variable corresponds to such knowledge as may vary. In a propositional schemata for a dog the fixed part would be such knowledge as four legged animal while the variable parts correspond to size, colour etc.

**Image schemata** - are used to organise visual patterns. Stephen Hall [1992] speculate that “it may be that the human brain not only perceives but stores the essentials of a visual scene using the same geometrical, quasi-symbolic, minimalist vocabulary found in maps”. Tied to the propositional schemata of the dog described above there are image schemata of several different kinds of dogs in our head that gives us the ability to classify an animal we meet in the street as a dog.

**Procedural schemata** - concern knowledge on how to perform activities in a sequence to achieve a certain result, such as the sequence of steps that need to be performed to get from one step to another.

A category can be formed through combinations of different schematas. Furthermore the different schemata are tied together into more complex structures called “idealized cognitive models, ICM” to represent different kinds of knowledge at different abstraction levels [Lakoff 1987]. A skilled map reader for instance can recognise all kinds of characteristics of a landscape from a map using ICMSs that consists of different image and propositional schemata that has been developed through experience. A characteristic of the different knowledge schemata however is that it is very difficult to translate knowledge that is stored in one type of schemata to another kind of schemata, e.g. from image schemata to propositional schemata. The ability to recognise the breed of a certain dog for instance has developed as image knowledge through having been shown several different samples of that breed. It might be possible to describe a breed of a dog sufficiently well using propositional knowledge or words to someone so that he acquires the ability to recognise the breed of that dog. However, it is quite difficult and it is much more efficient to give the person the ability to develop an image schemata through pointing at a dog. When a cartographer performs a generalisation he relies on knowledge stored as image schemata but when we ask him to describe what he does we expect him to translate the knowledge into propositional form which is very difficult.

In object oriented modelling of geographical information we mainly rely on knowledge stored in propositional schemata when defining object classes and attributes. Cartography however rely on knowledge stored in both propositional and image form. The categories used in geographical information are formed both through map reading and real world experience and hence the categories used in a geographical database to a certain extent has to be based on maps. The characteristics an entity should have to belong to a certain category in a map is defined through both propositional and image knowledge. The propositional knowledge that describe a built up area for instance is that it consists of dense buildings, streets, etc. How a certain built up area shall be delineated however depends on an image knowledge that recognises patterns of these objects and relations to neighbouring objects. The propositional knowledge in this case is relatively independent of resolution while
the delineation of the built up area changes with scale. The propositional knowledge that define forest is relatively independent of resolution: trees, blueberries etc. However the forest has to cover a certain area to be considered as a forest. This area changes with resolution, and it seems reasonable to argue that this knowledge is propositional. Which patterns in relation to other area objects that form a forest at a certain resolution however is represented in the image schemata.

**Design of the Object oriented data-model**

As was stated above an object oriented data-model is always a simplification of reality that should focus on the most important aspects of the problem at hand. In this paper the most important aspect is to abstract information to design a user interface for a particular application. When abstracting the information the following aspects are considered:

**Object Oriented Modelling** - General theories on object oriented modelling as described by e.g. Rumbaugh et al [1992] should be adhered to.

**Human Cognition** - The user interface has to provide functionality for two different forms of analysis of the geographical information: visual and analytical. The analytical consist of such questions as “How many dwelling houses are located within three hundred meters from this highway?” and are independent of how the geographical information is presented in a map. To be able to perform such analysis the object classes ought to be defined according to the Aristotelian tenets described above (an exception is the fuzzy set theory developed by Zadeh [1965] which is not considered here).

Visual analysis is a human ability that relay on knowledge stored in image schematas. How the information is generalised and presented in a map gives us the ability to recognise different landscape characteristics. Consider for instance how a group of buildings are generalised and presented at different scales. At a large scale every building is presented and the category building can be defined using the Aristotelian tenets. At a smaller scale a subset of the buildings are selected to represent the pattern of buildings. The buildings now represent themselves in a sense that it is possible to look at the attributes of a certain building. But simultaneously they are selected to represent the pattern of buildings using image knowledge of the cartographer. At a yet smaller scale a typification is made and the buildings that know are displayed only represent the pattern of buildings but there is no 1:1 correspondence to a real world object. We believe it is reasonable to model this as three different object classes since they correspond to three different but related mental concepts. At the large scale such information as geometrical accuracy are of major importance and is kept when an entity is abstracted to form a member of the object class. At the small scale the pattern of buildings, topological relations, shape etc are considered to be more important than geometrical accuracy and when the real world entities are abstracted to form objects the important parameters will be given as true values as possible on the expense of geometrical accuracy. Each object class representing buildings need to be well defined and it is only the object class designed to be represented at a large scale that is suitable to use in analytical questions. To provide functionality for visual analysis the geographical information has to be modelled in a manner that correspond to human image schemata.

During the modelling it has become apparent that object classes in geographical information does not exist independently of other object classes. Some relations between different object classes has to be modelled explicitly. The following concepts are introduced:

**The Level** - How geographical objects are looked upon depend to a significant extent on the neighbouring objects. Three different aspects of the relation between objects has been noted: How objects are delineated depends on which other object classes are to be displayed in the same map. Some object classes have topologi-
cal relations e.g. roads and lakes while other object classes have not, e.g. roads and geological information. Finally object classes that are to be presented together has to have the geometries modified due to symbolisation. Based on these observations the concept "level" is introduced into the data model. The level is similar to a layer or a map but has the following characteristics:

- The level is defined at a certain resolution.
- An object class belongs to one and only one level.
- All object classes that belong to a level has topological relations.
- The presentation of the objects are defined in the level and all conflicts due to symbolisation are solved when the level is created.

**The Network** - Crossings between linear objects are important to preserve since they provide reference points in the landscape. Therefore a node-link structure is created between all linear objects (streams, roads, streets, railways etc).

**The Surface layer** - The object classes defined as areas should form a complete surface-cover.

An object class is defined with attributes and geometry type. Furthermore attributes define characteristics an entity should have to belong to the object class. A lake for instance should be of a certain size, a cul-de-sac road should be of a certain length etc. The data model is implemented in an object oriented database where the object classes are equipped with methods. One such method is the constructor of an object. The constructor is defined for an object class and analyses information in one or several more detailed levels defined at larger scales to obtain entities that can instantiate an object in this class. The constructor analyses if the obtained entities are relevant to form objects in this class, performs a generalisation, and locates the new object in the map-level where there is enough available space. If not enough space is available neighbouring objects can be requested to move.

These different concepts leads to a database design with a rather unusual structure for a GIS. The database consists of several different levels defined at different resolutions and for different purposes. A real world entity such as a building can be represented by several different objects defined in different levels. In some cases it might be appropriate to define pointers between objects in different levels that represent the same real world entities such as buildings or lakes but for several objects hierarchical relations are considered to be very difficult to define. Theoretically it would be possible to define which buildings belongs to a built up area. How a built up area is delineated however depends not only on buildings but on such things as street patterns, map purpose and resolution and which buildings that belong to a built up area depends on how the level is defined and will be different for different built up areas defined in different levels. To define such hierarchies is considered to be relatively unimportant.

The information in a level can be obtained from several different sources. In this article we only consider how a level obtains information from one or more levels defined at larger resolutions.

**Design of The User Interface**

The user interface consists of a set of levels defined at different resolutions. As the user zooms the level that corresponds to the chosen resolution is displayed. The set of levels has to be designed so that the changes in display appear to be logical as the user zooms. When a new user interface is to be designed a set of new levels is created. Object classes in each level are defined, symbolisation is chosen and the constructors and methods to obtain and analyse information are written. When everything is defined a process in the level-object initiates the constructors of the different object classes. Since it isn’t possible to construct all objects at once some
consideration has to be taken as to in which order the different objects shall be handled. Our intent is that the process should simulate how a cartographer works when creating a new small scale map using large scale maps. Before generalising the information in a certain area the cartographer forms a mental image of how the area shall appear in the small scale map. Then he starts drawing the most important features and generalises them to leave enough space for less important features that will be drawn later. As he proceeds the new map is gradually filled with more and more details while the cartographer simultaneously analyses information in two levels: the new small scale map and the old large scale map. In the object oriented database the constructors and other methods are written in a similar manner that analyses information in both the original and the new level simultaneously while the information is created.

As an example the creation of a lake object consists of the following steps, where lakes are the first objects to be created in this level:

1. The level object calls the constructor of the lake class.
2. The constructor searches the large scale level to locate any objects that can be suitable to form new lake objects. The objects has to be of a certain size. When an object is found it is collected.
3. Depending on the geometrical accuracy requirements defined for the objects in the lake class an area where the new lake may be located is calculated through forming a corridor around the line segments of the lake object. This area is of course larger than the lake area.
4. The area where the new lake may be located performs an overlay to find any objects in the new level that are located within this area. The area where the new lake may be located is diminished to avoid conflicts with any objects located through this overlay. However it may not be diminished to such an extent that the geometrical accuracy constraints of the lake object are violated.
5. The geometry of the original lake object and the geometry that defines where the new lake object may be located to avoid conflicts are used as input to an algorithm that creates the simplified geometry of the new lake object. With this geometry the new lake object is instantiated.
6. Since the geometry that defines where the new lake object may be located was not allowed to move to such an extent that the geometrical accuracy constraints could be violated conflicts may appear when the new object is instantiated. These are now located through an overlay and a vector is calculated for each object that needs to be moved. Furthermore the part of the geometry of an object that needs to be moved is determined.
7. The move function of each of the objects that are in conflict with the current object is called and cause the particular geometry of that object to move. Only the part of the geometry of the object that is in conflict is moved. After all objects has been moved a smooth function for each object is called to smooth the geometry of each object that has not been moved but is effected by the move.
8. For each object that has been moved an analysis is performed to locate any conflicts created through the move and to solve these conflicts in a similar manner as to the one described above.

The example is designed to be as simple as possible and there are several difficulties in the model. For instance in a complex database there is a risk that the move of objects to solve conflicts may cause a never-ending chain of objects requesting each other to move. However the aim is to develop a method to design user interfaces for particular applications and initially these applications ought to be rather simple. Furthermore the method provides an interesting framework where different generalisation operators can be implemented.
Concluding remarks

According to Rumbaugh et al. [1992] “There is no single ‘correct’ model of a situation, only adequate and inadequate ones.” The model proposed here focus on generalisation as a design process and aims at giving the cartographer tools to design new user interfaces through setting up a definition of a new set of object classes and equip these with methods that fills the newly designed map with information.

Booch et al. [1999] states that “No single model is sufficient. Every nontrivial system is best approached through a small set of nearly independent models.” The level concept introduced above implies that several levels can be defined at the same, or near the same, resolution and contain object classes that represent the same real-world entities. However the definition of the object classes might be slightly different concerning geometrical accuracy, attributes, symbolisation etc and thus represent different abstractions of reality. The different levels represent different models that gives us the ability to focus on different aspects of the geographical information. The models are independent in a sense that there is no need to communicate information between different levels at the same resolution even though individual object classes in different levels might have quite a few things in common.

Acknowledgements

This paper is the result of a joint research project between University of Gävle and VBB Viak Company, sponsored by the Foundation of Knowledge and Competence Development, the Swedish Armed Forces, the Swedish Road Administration, the Swedish Railroad Administration, the Swedish Maritime Administration, the Geological Survey of Sweden, ESRI, Swedish Space Co-operation and the Swedish National Land Survey.

References


A Spline Function is Applied in Map-Projection Transformation

Yinxiang Yuan
Dept of Hydrography
1 Xiaolong Street Zhongshan District Dalian 116018

When conducting numerical transformation between different map-projection, if the included area is too wide and the given known-points are inadequate, the creation of a mathematical model based on these points will not achieve a satisfactory result. In order to guarantee a precise transformation, the whole process may be run in two steps. The first is densifying, the second is interpolating. It would be suitable to employ the third-order spline model as a densifying mathematical model when the spline model is expressed in terms of the first-order derivative and when its first edge condition is given. And it would be better to use the double-second-order polynomial as an interpolating model.

Because the high-order interpolating method is hampered by such faults as slow calculation and instability, the densifying method with spline function is put forward. Since the power of this method is low, its calculation is simple, stable and precise. In addition, it has some degree of smoothness (the first and second derivative continuity).

1. Expression of the third-power spline function

This paper uses the first derivative expression. On basis of Hermite interpolating formula we may write out the first derivative expression of the third-power spline function, which is as follows:

\[ s_3(x) = m_{i-1} \frac{(x_i - x)^2(x - x_{i-1})}{h_i^2} - m_i \frac{(x - x_{i-1})^2(x_i - x)}{h_i^2} \]

\[ + y_{i-1} \frac{(x_i - x)^2[2(x - x_{i-1}) + h_i]}{h_i^3} + y_i \frac{(x - x_{i-1})^2[2(x_i - x) + h_i]}{h_i^3} \]

where \( h_i = x_i - x_{i-1} \)

On the basis of the continuity of its second derivatives we can obtain a relation expression (continuity-equation) of all its factors:

\[ (1 - \alpha_i)m_{i+1} + 2m_i + \alpha_i m_{i+1} = \beta_i \]

where \( \alpha_i = \frac{h_i}{h_i + h_{i+1}} \), \( h_i = x_i - x_{i-1} \), \( \beta_i = 3[(1 - \alpha_i) \frac{y_i - y_{i+1}}{h_i} + \alpha_i \frac{y_{i+1} - y_i}{h_{i+1}}] \quad (1 < i < n-1) \)

This expression is a group of \( n-1 \) equations which contains \( n+1 \) unknown variables. If you want to obtain the unique group of solutions, you must add two equations to it, and define two edge conditions.
2. The method for the definition of edge condition

The ways to define edge conditions consist of three kinds of spline function programs.

a. Giving the first derivatives $m_0, m_n$ at both terminal points;

b. Giving the second derivatives $M_n = M_0 = 0$ at both terminals;

c. Making all the different order derivatives equal at the different order derivatives, that is:

$$S^{(k)}(x^+) = S^{(k)}(x^-), \quad k = 0, 1, 2$$

In order to densify the meridian-parallel network we choose the first kind, giving the first derivatives $m_0, m_n$ at both the terminals, and letting $s'(x_0) = m_0, s'(x_n) = m_n$.

Thus two unknown variables are cut off from equations (2). Therefore, equations (2) become a group of equations with the unique series of solutions, which is then changed into matrix-form:

$$\begin{bmatrix}
2 & \alpha_1 & 0 & \ldots & 0
\
1 - \alpha_1 & 2 & \alpha_2 & \ldots & \ldots
\vdots & \vdots & \vdots & \ddots & \vdots
0 & \ldots & 1 - \alpha_{n-2} & 2 & \alpha_{n-1}
0 & \ldots & 0 & 1 - \alpha_{n-1} & 2
\end{bmatrix}
\begin{bmatrix}
m_1 \\
m_2 \\
\vdots \\
m_{n-1} \\
m_n
\end{bmatrix} = 
\begin{bmatrix}
\beta_1 - (1 - \alpha_1)m_0 \\
\beta_2 \\
\vdots \\
\beta_{n-2} \\
\beta_{n-1} - \alpha_{n-1}m_n
\end{bmatrix}$$

To solve it, we use “catch-up” method:

$$M_j = a_j m_{j+1} + b_j$$

where $a_j = \frac{\alpha_j}{2 + (1 - \alpha_j)a_{j-1}}, \quad b_j = \frac{\beta_j - (1 - \alpha_j)b_{j-1}}{2 + (1 - \alpha_j)a_{j-1}}$

Initial values $m_0, m_n$ may be obtained from the Lagrange formula, but we must derive the proper numerical derivative formula. The Lagrange interpolation formula:

$$L(x) = \sum_{i=1}^{n} y_i \prod_{j=1}^{n} \frac{x - x_j}{x_i - x_j}$$

We can change a bit this formula into the following forms:

$$L_1(x) = \sum_{i=k+0}^{k+1} y_i \prod_{j=k+0}^{k+1} \frac{x - x_j}{x_i - x_j}$$

$$L_2(x) = \sum_{i=k+0}^{k+2} y_i \prod_{j=k+0}^{k+2} \frac{x - x_j}{x_i - x_j}$$

$$\ldots$$

$$L_n(x) = \sum_{i=k+0}^{k+n} y_i \prod_{j=k+0}^{k+n} \frac{x - x_j}{x_i - x_j}$$
After taking the derivative, we obtain the following numerical derivative formulae:

\[ L_2'(x_0) = \frac{-\frac{3}{2}y_0 + 2y_1 - \frac{1}{2}y_2}{h} \]
\[ L_2'(x_n) = \frac{1}{2}y_{n-2} - 2y_{n-1} + \frac{3}{2}y_n}{h} \]
\[ L_3'(x_0) = \frac{-\frac{11}{6}y_0 + 3y_1 - \frac{3}{2}y_2 + \frac{1}{3}y_3}{h} \]
\[ L_3'(x_n) = \frac{-\frac{1}{3}y_{n-3} + \frac{3}{2}y_{n-2} - 3y_{n-1} + \frac{11}{6}y_n}{h} \]
\[ L_4'(x_0) = \frac{-\frac{25}{12}y_0 + 4y_1 - 3y_2 + \frac{4}{3}y_3 - \frac{1}{4}y_4}{h} \]
\[ L_4'(x_n) = \frac{\frac{1}{4}y_{n-4} - \frac{4}{3}y_{n-3} + 3y_{n-2} - 4y_{n-1} + \frac{25}{12}y_n}{h} \]
\[ L_5'(x_0) = \frac{-\frac{137}{60}y_0 + 5y_1 - 5y_2 + \frac{10}{3}y_3 - \frac{5}{4}y_4 + \frac{1}{5}y_5}{h} \]
\[ L_5'(x_n) = \frac{-\frac{1}{5}y_{n-5} + \frac{5}{4}y_{n-4} - 10y_{n-3} + \frac{5}{3}y_{n-2} - 5y_{n-1} + \frac{137}{60}y_n}{h} \]

On basis of different precision requirements we can choose one from upper formulae.

### 3. Densifying

First we take the longitudes as constants and densify on the meridians, then take the latitudes as constants and densify on the parallels, whereupon two double-variable functions \( x = f_1(F, ?) \) and \( y = f_2(F, ?) \) become:

\[ x = ?_1(F, ?_i), \quad \text{the function on the } i\text{-th meridian} \]
\[ x = ?_2(F, ?_j), \quad \text{that on the } j\text{-th parallel} \]
\[ y = ?_3(F, ?_i), \quad \text{that on the } i\text{-th meridian} \]
\[ y = ?_4(F, ?_j), \quad \text{that on the } j\text{-th parallel} \]

\( i = 1, 2, 3, ..., m_1; \ j = 1, 2, 3, ..., m_2 \)

\( m_1 \) is the number of the meridians, \( m_2 \) is the number of the parallels.

### 4. The Precision of Densifying

If \( f \in \mathcal{C}^4(a, b) \), then

\[ |f - s_3(x)| < \frac{5}{384}M_4 h^4 \]

If \( f \in \mathcal{C}^5(a, b) \), then

\[ |f - s_5(x)| < 1/384 M_4 h^4 + 1/60 M_5 h^5 \]

where \( f \in \mathcal{C}l(a, b) \) denotes that \( f \) has continuous derivatives up to the \( l \)-th order.

\[ M_4 = \max \left| \frac{d^4 f}{dx^4} \right|, \quad M_5 = \max \left| \frac{d^5 f}{dx^5} \right|, \quad (a < x < b) \]
5. Interpolation of the Numerical Transformation of Map Projections

To interpolate, we use the double-second-order polynomial. The formula of its factors are as follows:

\[ A_{ki} = x_k^{i-1} y_k^{j-1} \]

Where \( k=1,2,3,\ldots,9; \ i=1,2,3; \ j=1,2,3; \ l=3(i-1)+j \)

Solving the following group of equations:

\[
\begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{19} \\
    a_{21} & a_{22} & \cdots & a_{29} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{91} & a_{92} & \cdots & a_{99}
\end{bmatrix}
\begin{bmatrix}
    u_1 \\
    u_2 \\
    \vdots \\
    u_9
\end{bmatrix}
=
\begin{bmatrix}
    x'_1 \\
    x'_2 \\
    \vdots \\
    x'_9
\end{bmatrix}
\]

\[
\begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{19} \\
    a_{21} & a_{22} & \cdots & a_{29} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{91} & a_{92} & \cdots & a_{99}
\end{bmatrix}
\begin{bmatrix}
    v_1 \\
    v_2 \\
    \vdots \\
    v_9
\end{bmatrix}
=
\begin{bmatrix}
    y'_1 \\
    y'_2 \\
    \vdots \\
    y'_9
\end{bmatrix}
\]

We obtain factors \( u_l, v_l \) (\( l=1,2,\ldots,9 \)). Finally, with factors \( u_l \) and \( v_l \) work out the coordinates of the point on the surface \( A' \), which corresponds to the point \( (x,y) \) on the projection-surface \( A \).