

## Spatial Concept Lattices: An Integration Method in Model Generalization

The ability to view and analyze data of different detail and from different perspectives, and to move dynamically from one scale to another requires modeling geographic information at different generalization levels. On this account, in the framework of model generalization, the concept of multi-scale database is adopted to provide a consistent multiple representation of existent mono-scale representations. Spatial Concept Lattices are propounded as a new approach to thematic generalization through the semantic integration of multiple classification schemata and the creation of a multi-scale, multi-context database. The methodology presents in an explicit and systematic manner the integration of classification schemata, which exhibit differences in spatial and thematic resolution. In order to comprehend the stepwise SCL methodology, an actual example is used to demonstrate the integration of three independent land cover/land use classification schemata. The integration process is part of model generalization, since the resulting hierarchical integrated schema supports various levels of thematic resolution and represents geographic space from different application perspectives.

**Keywords:** concept lattices, semantic integration, thematic classification, model generalization.

In contrast to cartographic generalization which focuses on graphic representation issues, model generalization involves modeling geographic information at different levels of spatial and semantic resolution (Müller et al., 1995). From this perspective, model generalization reflects changes in the perception level of geographic information (Ruas & Lagrange, 1995), and hence precedes cartographic generalization (Ruas & Lagrange, 1995; Weibel, 1995; Weibel & Dutton, 1999). The important objective of model generalization is the production of databases at multiple levels of detail, for multiple purposes and applications (Molenaar, 1996; Müller et al., 1995; Uitermark et al., 1998; Voisard & Schweppe, 1998). Multiple representations outstrip static views, expanding users' ability of viewing and analyzing geographic data. The concept of multi-scale (or multi-resolution) databases aims at the representation of the same real-world phenomenon at different resolutions (Weibel, 1995; Devogele et al., 1997; Weibel & Dutton, 1999). This is normally accomplished, either by generalizing a single large-scale database, or by collecting different independent representations and each time utilizing the appropriate representation for the specified level of detail (Buttenfield, 1995; Devogele et al., 1997; Govorov, 1995).

Another approach to building a multi-scale database concentrates on the development of multiple representations by integrating existing databases at different levels of detail, and linking representations that correspond to the same real world phenomenon. This approach allows reuse of data, and interoperability between representations (Devogele et al., 1997). Thereby, users can navigate dynamically and continuously from one level of detail to

Margarita Kokla  
Marinos Kavouras  
Cartography Laboratory  
Department of Rural and  
Surveying Engineering  
National Technical University  
of Athens  
9, H. Polytechniou Str.  
157 80 Zografos Campus  
Athens, Greece  
mkokla/mkav@survey.ntua.gr

### INTRODUCTION

*“The important objective of model generalization is the production of databases at multiple levels of detail, for multiple purposes and applications.”*

*“The intention of the present work is to tackle thematic generalization by putting emphasis on attributes and semantic integration, as a form of model generalization.”*

another, specified by the scale needed for the application. However, if the input databases do not use the same conceptual schema, the process of building a multi-scale database demands a schema integration methodology. The intention of the present work is to tackle thematic generalization by putting emphasis on attributes and semantic integration, as a form of model generalization.

Spatial Concept Lattices (SCL) provide a specific and systematic methodology for the semantic integration of multiple classification schemata. The methodology can be used to build a multi-scale, multi-context database, providing multiple representations of geographic data, not only at various scales and levels of thematic resolution, but also from different application perspectives and thus different semantics. The integration of different classification schemata, apart from providing the means to move along different levels of detail and intelligently change scale, it also allows to move across different contexts and perform a change in the perception of geographic information (Fig. 1).

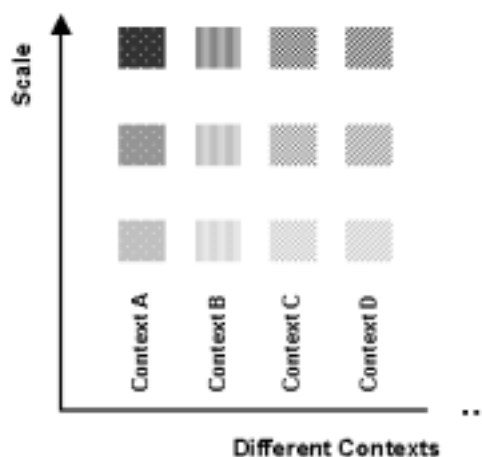


Figure 1. Integration along different scales and across different contexts

Previous work (Kokla & Kavouras, forthcoming) has developed the SCL methodology for the semantic integration of two different classification schemata. This paper exploits the ability of SCL to produce an integrated classification schema as the basis for dynamic model generalization.

#### THE SPATIAL CONCEPT LATTICES METHOD

#### Formal Concept Analysis

*“SCL are founded on Formal Concept Analysis (Wille, 1992), a theory of concept formation and conceptual classification.”*

SCL are founded on *Formal Concept Analysis* (Wille, 1992), a theory of concept formation and conceptual classification. Formal Concept Analysis (FCA) provides a basic analysis of a context and at the same time indicates the implications between attributes. Generally, FCA works in a specific context named *Formal Context*, consisting of a set of objects, a set of attributes and a binary incidence relation connecting objects and attributes. The central notion in FCA is the *formal concept*, or *conceptual class* or *category*, which is defined as a collection of entities or objects exhibiting one or more common characteristics or attributes. Their *extent* and *intent* logically characterize formal concepts. The *extent* is the aggregate of objects or entities belonging to the concept, whereas the *intent* is the sum of attributes (or properties) implying the formal concept.

An important relationship in FCA is the *superconcept/subconcept relation*, which is defined as the order proceeding top-down from more generalized

concepts with larger extent and smaller intent to more specialized concepts with smaller extent and larger intent. Formal Concepts associated to each other with the superconcept/subconcept relation form a class hierarchy, called *Concept Lattice*. In other words, a Concept Lattice is the ordered set of all formal concepts of a formal context.

Concept Lattices have been applied to a number of different fields, such as medicine (Spangenberg & Wolff, 1999), biology (Ganter & Wille, 1989), sociology (Ganter & Wille, 1989), and information and computer science (Kent & Neuss, 1995; Faid et al., 1997; Schmitt & Saake, 1997; Deogun et al., 1998; Priss, 1999). Schmitt and Saake (1997) have applied the algorithm provided by FCA for deriving concept lattices from context information for schema integration. In order to satisfy the demands of database design and maintenance, they transform the integrated schema with respect to different quality criteria.

In order to deal with the geospatial characteristics of entities and the classification schemata involved in geographic applications, FCA has been employed to derive Lattices of Spatial Concepts, thereafter called Spatial Concept Lattices (SCL). In this research, SCL are applied to manage hierarchical geographic data with overlapping classes. The process of creating the Spatial Concept Lattice, and hence the integrated hierarchical schema, necessitates knowledge of scale transitions and inheritance relationships for the entities involved and their attributes. Entities and attributes are in most cases scale dependent. Moreover, very often, the meaning of spatial entities, as well as their level of detail is inherently based on their attributes. This intrinsic spatial knowledge is necessary for identifying correspondences and resolving conflicts between the input classification schemata. However, this does not influence the mechanics of the method at all, which may be claimed to be independent of the application.

The integration process described in this research is based on the theory of FCA. The conversion of schemata into one merged context, as well as the transformation of the integrated schema relies on an improved approach as described in Schmitt & Saake (1997).

In order to demonstrate in a comprehensive fashion the application of FCA, a running example is used involving the integration of three independent classification schemata:

- The hierarchical CORINE Land Cover nomenclature (CORINE Land Cover-Technical Guide, 1994) for scales 1:100,000–1:1,000,000.
- The DIGEST nomenclature for geographic objects (DIGEST Standards Specification, DGIWG, 1997), addressing a variety of scales.
- The classification used by the Hellenic Mapping and Cadastral Organization (Technical Specifications of the Greek Cadastre, HEMCO, 1996) to record land use characteristics referring to scales 1:1,000–1:5,000.

For reasons of presentation and space limitation, only a small, but complete excerpt of the above case is presented.

The SCL methodology used in this research is formalized by an algorithm, which proceeds in two main steps. At the first step, the different contexts are merged into a single, integrated context. The second step is devoted to the generation of the concept lattice of the integrated context. More specifically, in order to integrate the different contexts into a single one, it is necessary to find and resolve conflicts and identify correspondences between the input classification schemata. Therefore, the classes of the input classification schemata are analyzed (Fig. 2), to specify equivalencies and overlappings between them (extensional decomposition).

Thus, the first column of Matrix 1 lists the original classes of the input

*“... a Concept Lattice is the ordered set of all formal concepts of a formal context.”*

*“In this research, SCL are applied to manage hierarchical geographic data with overlapping classes.”*

*“The SCL methodology used in this research is formalized by an algorithm, which proceeds in two main steps. At the first step, the different contexts are merged into a single, integrated context. The second step is devoted to the generation of the concept lattice of the integrated context.”*

classification schemata and the remaining columns represent the classes derived by the extensional decomposition. In case of overlapping between two or more original classes, these are split to subclasses. A cross (x) for the intersection of a column (subclass) and a row (original class) means that the specific subclass belongs to the corresponding original class. For example, the CORINE Land Cover category "Fruit trees and berry plantations" is decomposed to subclass  $s_7$  and  $s_8$ , corresponding respectively to cadastral categories "Citrus fruits" and "Other trees". Categories "Arable land" (CORINE Land Cover and HEMCO) and "Cropland" (DIGEST) denote the same thing and thus include the same subclasses  $s_1$  ("Non-irrigated arable land"),  $s_2$  ("Permanently irrigated arable land") and  $s_3$  ("Rice fields").

Similarly, specific attributes are assigned to each original class (Fig. 3), in order to determine equivalencies and overlappings between attributes (intensional decomposition). A cross (x) for the intersection of a column (attribute) and a row (original class) means that the specific attribute is ascribed to the corresponding original class. For example, category "Crop-

		Subclasses									
		$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
CORINE Land Cover	<b>2. Agricultural Areas</b>	x	x	x				x	x	x	x
	<b>2.1. Arable land</b>	x	x	x							
	2.1.1. Non-irrigated arable land	x									
	2.1.2. Permanently irrigated land		x								
	2.1.3. Rice fields			x							
	<b>2.2. Permanent Crops</b>							x	x	x	x
	2.2.1. Vineyards									x	
	2.2.2. Fruit trees and berry plantations							x	x		
	2.2.3. Olive groves										x
CADASTRE	<b>1.1. Cultivated areas</b>	x	x					x	x	x	x
	1.1.1. Trees							x	x	x	x
	1.1.1.1. Vineyards									x	
	1.1.1.2. Olive groves										x
	1.1.1.3. Citrus fruits							x			
	1.1.1.4. Other								x		
	1.1.2. Arable land	x	x	x							
	1.1.2.1. Irrigated		x								
1.1.2.2. Non-irrigated	x										
DIGEST	<b>4. Vegetation</b>										
	<b>4.1 Cropland</b>	x	x	x	x	x	x	x	x	x	
	4.1.1. Cropland	x	x	x							
	4.1.2. Hedgerow				x						
	4.1.3. Nursery					x					
	4.1.4. Botanical Garden						x				
	4.1.5. Orchard/Plantation							x	x		
4.1.6. Vineyards									x		

Figure 2. Extensional decomposition-Matrix 1

land” (DIGEST) has the attribute “Crop type” ( $a_{18}$ ). It is important to mention that subclasses inherit the attributes of their superclasses, as for example CORINE Land Cover class “Arable land” inherits attributes  $a_1$  to  $a_3$  from its superclass “Agricultural Areas”.

Then, attributes specified during the intensional decomposition are associated with subclasses resulting from the extensional decomposition. In other words, Matrix 3 (Fig. 4) is created by combining Matrices 1 and 2. Essentially, Matrix 3 constitutes the cross-table of the integrated context, and it is used to build the multi-context Concept Lattice.

At the second step, Formal Concept Analysis is applied, in order to generate the Concept Lattice of the integrated context.

If G is the set of subclasses and M the set of attributes, then two sets are defined as follows:

$$Int = \{\{g\}' \mid g \in G\}$$

$$Ext = \{\{m\}' \mid m \in M\}$$

		<b>Attributes</b>																			
		Climate $a_1$	Altitude $a_2$	Productivity $a_3$	Soil type $a_4$	Annual crop $a_5$	Irrigation system $a_6$	Soil humidity $a_7$	Cultivation methods $a_8$	Fruit type $a_9$	Olive type $a_{10}$	Average yield produce $a_{11}$	Percentage of vegetation $a_{12}$	Area $a_{13}$	Number of trees $a_{14}$	Vineyard type $a_{15}$	Citrus type $a_{16}$	Tree type $a_{17}$	Crop type $a_{18}$	Hedgerow width $a_{19}$	Tree canopy levels $a_{20}$
<b>Original Classes</b>																					
<b>CORINE Land Cover</b>	<b>2. Agricultural Areas</b>	x	x	x																	
	<b>2.1. Arable land</b>	x	x	x	x																
	2.1.1. Non-irrigated	x	x	x	x	x															
	2.1.2. Permanently irrigated	x	x	x	x		x														
	2.1.3. Rice fields	x	x	x	x			x													
	<b>2.2. Permanent Crops</b>	x	x	x					x												
	2.2.1. Vineyards	x	x	x					x												
	2.2.2. Fruit trees	x	x	x					x	x											
2.2.3. Olive groves	x	x	x					x		x											
<b>CADASTRE</b>	<b>1.1. Cultivated areas</b>			x							x	x	x								
	1.1.1. Trees			x							x	x	x	x							
	1.1.1.1. Vineyards			x							x	x	x	x	x						
	1.1.1.2. Olive groves			x						x	x	x	x	x							
	1.1.1.3. Citrus fruits			x							x	x	x	x			x				
	1.1.1.4. Other			x							x	x	x	x				x			
	1.1.2. Arable land			x					x		x	x	x								
	1.1.2.1. Irrigated			x			x		x		x	x	x								
	1.1.2.2. Non-irrigated			x					x		x	x	x							x	
<b>DIGEST</b>	<b>4. Vegetation</b>																				
	<b>4.1 Cropland</b>																				
	4.1.1. Cropland																			x	
	4.1.2. Hedgerow																			x	
	4.1.3. Nursery																				
	4.1.4. Botanical Garden																				
	4.1.5. Orchard/Plantation																				x
4.1.6. Vineyards																					

Figure 3. Intensional decomposition-Matrix 2

Subclasses	Attributes	Non-irrigated arable land	Irrigated arable land	Rice fields	Hedgerow	Nursery	Botanical Garden	Citrus fruits	Other trees	Vineyards	Olive groves
		s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>	s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>	s <sub>7</sub>	s <sub>8</sub>	s <sub>9</sub>	s <sub>10</sub>
Climate	a1	x	x	x				x	x	x	x
Altitude	a2	x	x	x				x	x	x	x
Productivity	a3	x	x	x				x	x	x	x
Soil Type	a4	x	x	x							
Annual Crop	a5	x									
Irrigation system	a6		x								
Soil humidity	a7			x							
Cultivation methods	a8	x	x					x	x	x	x
Fruit type	a9							x	x		
Olive type	a10										x
Aver. yield produce	a11	x	x					x	x	x	x
Percentage of vegetation	a12	x	x					x	x	x	x
Area	a13	x	x					x	x	x	x
Number of trees	a14							x	x	x	x
Vineyard type	a15									x	
Citrus type	a16							x			
Tree type	a17								x		
Crop type	a18	x									
Hedgerow width	a19				x						
Tree canopy levels	a20							x	x		

Figure 4. Cross-table of the integrated context-Matrix 3

Set Int includes the attributes that distinguish each subclass, while set Ext represents the subclasses described by each attribute. Specifically for the excerpt of the running example:

$$\begin{aligned}
 \text{Int} &= \{\{s_1\}', \{s_2\}', \{s_3\}', \dots, \{s_{10}\}'\} \\
 &= \{\{a_1, a_2, a_3, a_4, a_5, a_8, a_{11}, a_{12}, a_{13}, a_{18}\} \\
 &\quad \{a_1, a_2, a_3, a_4, a_6, a_8, a_{11}, a_{12}, a_{13}\} \\
 &\quad \{a_1, a_2, a_3, a_4, a_7\} \\
 &\quad \{a_{19}\} \\
 &\quad \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{16}, a_{20}\} \\
 &\quad \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{17}, a_{20}\} \\
 &\quad \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\} \\
 &\quad \{a_1, a_2, a_3, a_8, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}\}\} \\
 \text{Ext} &= \{\{a_1\}', \{a_2\}', \{a_3\}', \dots, \{a_{20}\}'\} \\
 &= \{\{s_1, s_2, s_3, s_7, s_8, s_9, s_{10}\} \\
 &\quad \{s_1, s_2, s_3\} \\
 &\quad \{s_1\} \\
 &\quad \{s_2\} \\
 &\quad \{s_3\} \\
 &\quad \{s_1, s_2, s_7, s_8, s_9, s_{10}\} \\
 &\quad \{s_7, s_8\} \\
 &\quad \{s_{10}\}
 \end{aligned}$$

$$\begin{aligned} & \{s_7, s_8, s_9, s_{10}\} \\ & \{s_9\} \\ & \{s_7\} \\ & \{s_8\} \\ & \{s_4\}. \end{aligned}$$

Specifying sets of classes with common subclasses and attributes, derives the formal concepts of the integrated context. Therefore, from Int and Ext two sets of concepts are generated:

$$\text{Con}_I = \{(I, I) \mid I \in \text{Int}\}$$

$$\text{Con}_E = \{(E, E') \mid E \in \text{Ext}\}$$

For the specific application,  $\text{Con}_I$  and  $\text{Con}_E$  are formed as follows:

$$\begin{aligned} \text{Con}_I = & \{(\{s_1\}, \{a_1, a_2, a_3, a_4, a_5, a_8, a_{11}, a_{12}, a_{13}, a_{18}\}) \\ & (\{s_2\}, \{a_1, a_2, a_3, a_4, a_6, a_8, a_{11}, a_{12}, a_{13}\}) \\ & (\{s_3\}, \{a_1, a_2, a_3, a_4, a_7\}) \\ & (\{s_4\}, \{a_{19}\}) \\ & (\{s_7\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{16}, a_{20}\}) \\ & (\{s_8\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{17}, a_{20}\}) \\ & (\{s_9\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\}) \\ & (\{s_{10}\}, \{a_1, a_2, a_3, a_8, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}\})\}. \\ \text{Con}_E = & \{(\{s_1, s_2, s_3, s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3\}) \\ & (\{s_1, s_2, s_3\}, \{a_1, a_2, a_3, a_4\}) \\ & (\{s_1\}, \{a_1, a_2, a_3, a_4, a_5, a_8, a_{11}, a_{12}, a_{13}, a_{18}\}) \\ & (\{s_2\}, \{a_1, a_2, a_3, a_4, a_6, a_8, a_{11}, a_{12}, a_{13}\}) \\ & (\{s_3\}, \{a_1, a_2, a_3, a_4, a_7\}) \\ & (\{s_1, s_2, s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}\}) \\ & (\{s_7, s_8\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{20}\}) \\ & (\{s_{10}\}, \{a_1, a_2, a_3, a_8, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}\}) \\ & (\{s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}\}) \\ & (\{s_9\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\}) \\ & (\{s_7\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{16}, a_{20}\}) \\ & (\{s_8\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{17}, a_{20}\}) \\ & (\{s_4\}, \{a_{19}\})\}. \end{aligned}$$

From the union of sets of concepts, the set of final classes is derived:

$$\text{Con} = \text{Con}_I \cup \text{Con}_E$$

$$\begin{aligned} C_1 = & (\{s_1, s_2, s_3, s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3\}) \\ C_2 = & (\{s_1, s_2, s_3\}, \{a_1, a_2, a_3, a_4\}) \\ C_3 = & (\{s_1, s_2, s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}\}) \\ C_4 = & (\{s_7, s_8, s_9, s_{10}\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}\}) \\ C_5 = & (\{s_7, s_8\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{20}\}) \\ C_6 = & (\{s_1\}, \{a_1, a_2, a_3, a_4, a_5, a_8, a_{11}, a_{12}, a_{13}, a_{18}\}) \\ C_7 = & (\{s_2\}, \{a_1, a_2, a_3, a_4, a_6, a_8, a_{11}, a_{12}, a_{13}\}) \\ C_8 = & (\{s_3\}, \{a_1, a_2, a_3, a_4, a_7\}) \\ C_9 = & (\{s_4\}, \{a_{19}\}). \\ C_{10} = & (\{s_7\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{16}, a_{20}\}) \\ C_{11} = & (\{s_8\}, \{a_1, a_2, a_3, a_8, a_9, a_{11}, a_{12}, a_{13}, a_{14}, a_{17}, a_{20}\}) \\ C_{12} = & (\{s_9\}, \{a_1, a_2, a_3, a_8, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}\}) \\ C_{13} = & (\{s_{10}\}, \{a_1, a_2, a_3, a_8, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}\}) \end{aligned}$$

Matrix M (Fig. 5) is created, in order to represent the superconcept/ subconcept relation defined for the generated classes by comparing each class to all other classes. A value “1” in the binary matrix M at the intersection of row i and column j means that class  $C_i$  is a subclass of class  $C_j$ .

Then, Matrix  $M_1 = M - M * M$  (Fig. 6) is computed, in order to remove transitive specializations and thus, preserve only the direct subclasses of each class. Consequently, Matrix  $M_1$  reveals the hierarchical structure of the integrated context.

Subconcept-superconcept relation

Final Classes	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	C <sub>11</sub>	C <sub>12</sub>	C <sub>13</sub>
C <sub>1</sub>													
C <sub>2</sub>	X												
C <sub>3</sub>	X												
C <sub>4</sub>	X		X										
C <sub>5</sub>	X		X	X									
C <sub>6</sub>	X	X											
C <sub>7</sub>	X	X											
C <sub>8</sub>	X	X											
C <sub>9</sub>													
C <sub>10</sub>	X		X	X	X								
C <sub>11</sub>	X		X	X	X								
C <sub>12</sub>	X		X	X									
C <sub>13</sub>	X		X	X									

Figure 5. Subconcept-superconcept relation Matrix M

Subconcept-superconcept relation

Final Classes	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	C <sub>11</sub>	C <sub>12</sub>	C <sub>13</sub>
C <sub>1</sub>													
C <sub>2</sub>	X												
C <sub>3</sub>	X												
C <sub>4</sub>			X										
C <sub>5</sub>				X									
C <sub>6</sub>		X	X										
C <sub>7</sub>		X	X										
C <sub>8</sub>		X											
C <sub>9</sub>													
C <sub>10</sub>					X								
C <sub>11</sub>					X								
C <sub>12</sub>				X									
C <sub>13</sub>				X									

Figure 6. Subconcept-superconcept relation Matrix M<sub>1</sub> without transitive specializations



At last, the hierarchical, integrated schema (Fig. 7) can be optimized with respect to different criteria (Schmitt & Saake, 1997). The optimization process includes operations (e.g., removing classes, vertically or horizontally merging classes, splitting classes and removing multiple specializations), which aim at the improvement of the hierarchical schema. Figure 7 shows the corresponding excerpt of the final hierarchical schema generated from the original schemata. In the same figure, the shadowed block represents the class “Cultivated Areas” which can be removed by the transformation procedure.

*“An optimization process can be applied, in order to improve the final integrated schema.”*

**Discussion of SCL**

The SCL methodology and the algorithm presented above for the integration of classification schemata are based on the mathematical theory of Formal Concept Analysis. SCL can be used to formally describe the objects and their attributes at all levels of detail, as well as the relationships between object classes.

The methodology can be successfully applied independently of the spatial and thematic resolution represented by the input classification schemata. Therefore, it is possible to associate classifications created for similar purposes dealing with many overlappings between the input classes or, to integrate classification schemata of different thematic resolutions.

*“The methodology can be successfully applied independently of the spatial and thematic resolution represented by the input classification schemata.”*

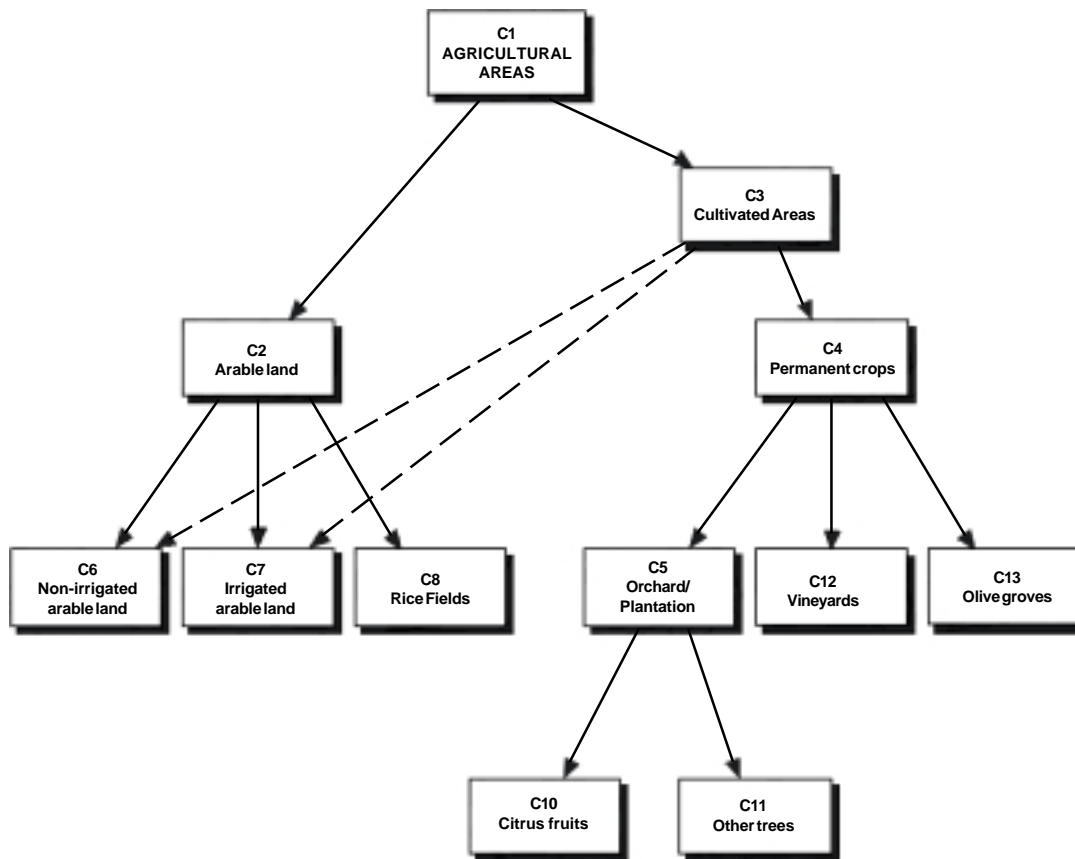


Figure 7. Excerpt of the final hierarchical schema.

*“... in SCL, hierarchies are used as a conceptual tool and not as a restriction of the method.”*

The integrated schema is hierarchical, but not strictly tree-structured, thought it can be transformed to a tree. This means that in the final schema, certain classes may have more than one superclass, and it depends on the user to select the appropriate link for the application at hand. For example, class “Cemetery” (DIGEST) may as well belong to “Green urban areas” (CORINE Land Cover), or to “Religious sites” (Cadastre). Therefore, in SCL, hierarchies are used as a conceptual tool and not as a restriction of the method.

Moreover, during the integration process, possible conflicts (Batini et al., 1986; Bishr et al., 1997; Pitoura et al., 1995; Reddy et al., 1994; Spaccapietra et al., 1992) between different schemata are resolved. These relate mainly to (Bishr et al., 1997): (a) naming conflicts, (b) semantic conflicts, and (c) schematic conflicts. Naming conflicts contain homonyms, where the same name is ascribed to different concepts and synonyms, where different names describe the same concept. Semantic conflicts occur due to different interpretations of the same concept. Schematic conflicts rise due to differences in schema elements (i.e., objects, attributes and relationships) of different schemata.

Finally, the integration process converts the input classification schemata to a single schema corresponding to an integrated but also uncompromising conception of space. Namely, the original classes and attributes are not altered, but semantically related to each other to form the final hierarchical schema. Therefore, the integration process identifies similarities and reconciles differences without preventing the independent and autonomous use of the original schemata.

#### USAGE OF THE INTEGRATED SCHEMA

The integration of different classification schemata provides a flexible and effective means to build a multi-scale, multi-context database. The integration can proceed both to the “vertical” and the “horizontal” direction (Fig. 1). “Vertical” integration refers to the association of classification schemata created for different scales, whereas “horizontal” integration refers to the fusion of classification schemata created by different agencies or for different applications. Thus, users can navigate along different scales, but also across different classification schemata and hence different conceptualizations of geographic data.

*“The present research focuses primarily on class-driven generalization...”*

The present research focuses primarily on class-driven generalization, separating the conceptual problem from the visualization context (Ruas & Lagrange, 1995; Kilpeläinen & Sarjakoski, 1995). From this perspective, generalization is considered as a process induced by variations in the conceptualization of space. Space is represented differently as scale changes. Correspondingly, variations in the perception and semantics of geographic information alter the level of detail. Class-driven generalization deals with changes in the perception and semantics of geographic information at a given level of detail. In this process, and contrary to geometric generalization, changes in the geometric aspects of features succeed changes in the conceptualization of geographic phenomena.

*“... the final hierarchical schema operates as a guide for determining the appropriate classification for a specific map scale.”*

SCL methodology serves as a basis for the development of a dynamic generalization/specialization process, which can subsequently be automated. Namely, the final hierarchical schema operates as a guide for determining the appropriate classification for a specific map scale. Given a scale and a context, the generalization hierarchy makes it possible to determine the appropriate “band” and derive the classes to be used (Fig. 8). Figure 9 shows an excerpt of the final hierarchical schema.

This particular case study addresses and associates two parameters of model generalization, scale and context. Apart from these parameters, the entity definition and its spatial characteristics also affect the model gener-

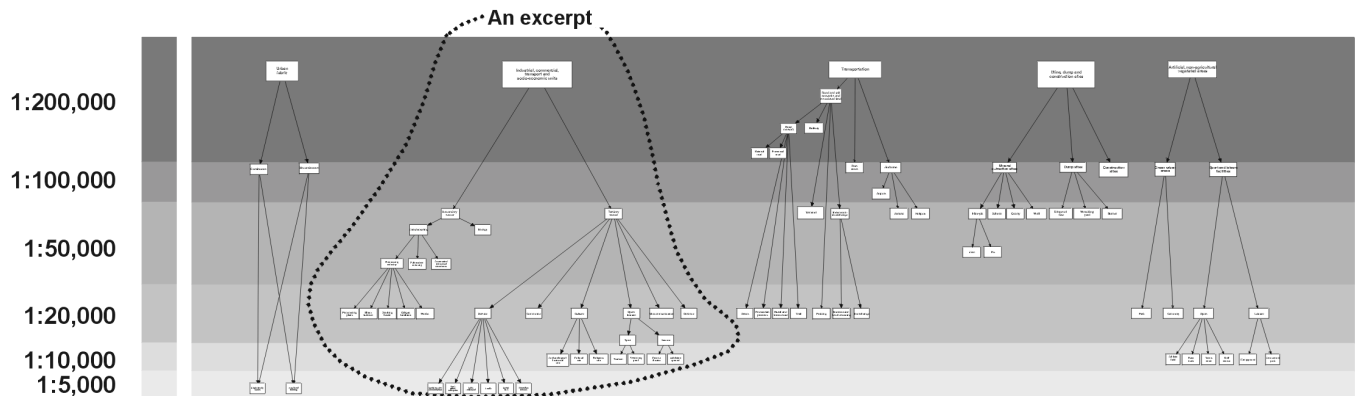


Figure 8. The final generalization hierarchy.

alization process, but they were not dealt with in this paper. Thus, classes are defined according to the scale and context specified by the user. These classes refer to specific geographic entities. For example, classes “Apartment building” and “Single-family house” appear at the lowest level of the hierarchical schema (large scale) referring to the geographical entity “building”. If at this large scale, other more general entities must appear (e.g., “building blocks”), their classification shall be determined from the hierarchical level where they are defined.

Moreover, the final integrated schema, due to its hierarchical structure, represents links between similar classes at different levels of detail. This ability can be utilized to transfer from one level of detail to another and thus provide continuous on-the-fly generalization on the screen, depending on the zoom factor. Zooming in or out prompts a change to the classification schema and its corresponding legend. Large scales permit the inclusion of more classes. As scale decreases, thematic classes tend to collapse (Fig. 10). Users who zoom in or out in a dynamic environment perceive only those classes that are appropriate for the specific level of detail. Furthermore, they can navigate dynamically and continuously from any scale to any other and from any classification scheme to any other by alternating the parameters of scale and context.

SCL constitute a methodology for integrating classification schemata corresponding not only to different scales, but also to different contexts. The resulting schema, due to its hierarchical structure, models the transition between similar classes at different levels of detail and different application perspectives. Therefore, the methodology constitutes a form of model generalization, based on attributes and semantic integration of different classifications. It provides the basis for determining the appropriate representation for a user-specified scale range or context, thus supporting dynamic thematic generalization.

A notable advantage of SCL is that the input classification schemata preserve their autonomy, because the developed methodology performs a complete integration and not a conversion between them. Consequently, each classification schema can still be used independently, as it is not altered, but it is semantically related to the others.

Future steps include further refinement of the methodology to address several kinds of attribute equivalencies and resolve conflicts occurring due to different domains or values of semantically similar attributes. In this case, the methodology would identify and distinguish data classes having different values of the same attributes.

*“The hierarchical, integrated schema serves as a basis for the development of a dynamic thematic generalization process.”*

CONCLUSION

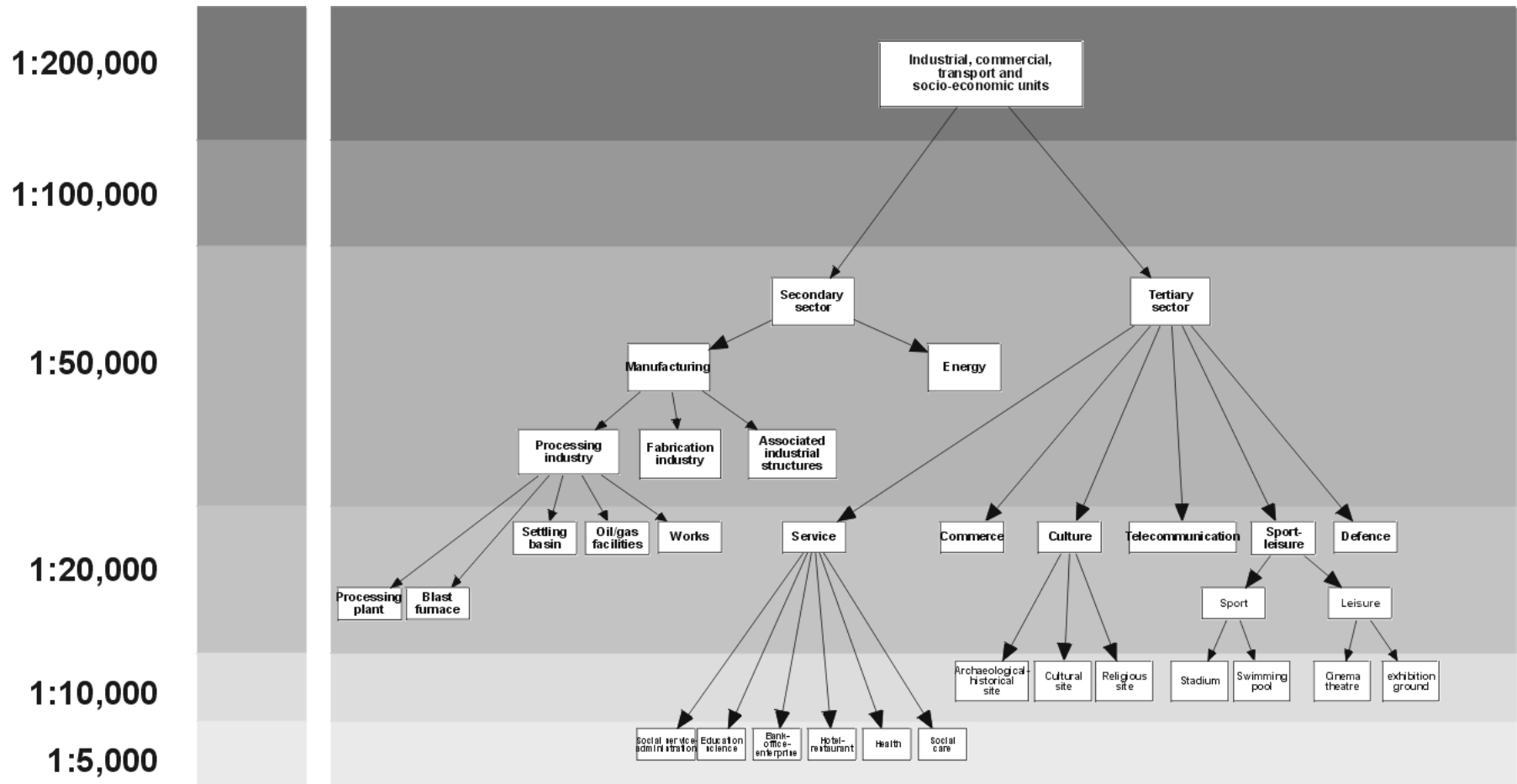
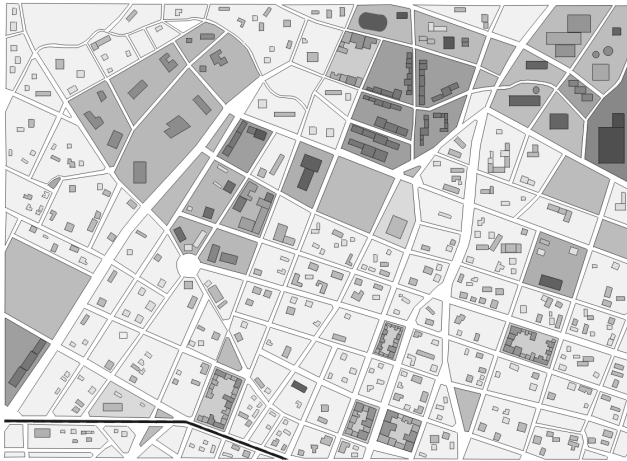


Figure 9. An excerpt of the final generalization hierarchy.

For the 1: 5,000 scale



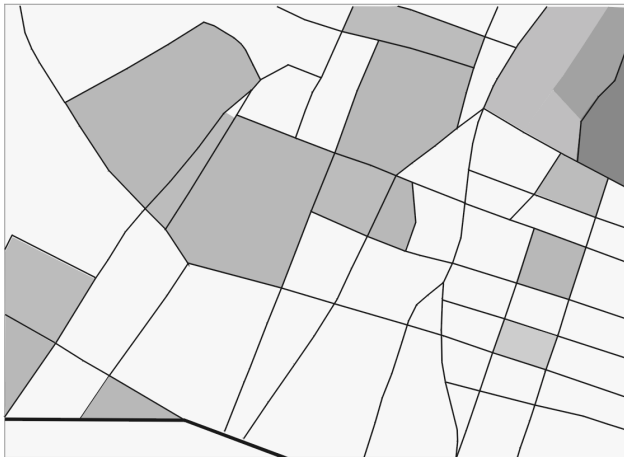
- Urban Fabric**
    - Continuous Urban Fabric
    - Discontinuous Urban Fabric
  - Secondary Sector**
    - Processing industry
    - Fabrication industry
    - Associated industrial structures
  - Tertiary Sector**
    - Commerce
    - Service
    - Culture
    - Sport and Leisure Facilities
  - Artificial Vegetated Areas**
    - Green, urban areas
- Single-family house
  - Apartment building
  - Single-family house
  - Apartment building
  - Processing plant
  - Settling basin
  - Oil/Gas facilities
  - Works
  - Blast furnace
  - Social service-Administration
  - Bank-Office-Enterprise
  - Hotel-Restaurant
  - Health
  - Education-Science
  - Sport
  - Leisure

For the 1: 10,000 scale



- Urban Fabric**
    - Continuous Urban Fabric
    - Discontinuous Urban Fabric
  - Secondary Sector**
    - Processing industry
    - Fabrication industry
    - Associated industrial structures
  - Tertiary Sector**
    - Commerce
    - Service
    - Culture
    - Sport and Leisure Facilities
  - Artificial Vegetated Areas**
    - Green, urban areas
- Processing plant
  - Settling basin
  - Oil/Gas facilities
  - Works
  - Blast furnace
  - Sport

For the 1: 50,000 scale



- Urban Fabric**
  - Continuous Urban Fabric
  - Discontinuous Urban Fabric
- Secondary Sector**
  - Processing industry
  - Fabrication industry
  - Associated industrial structures
- Tertiary Sector**
  - Tertiary Sector
- Artificial Vegetated Areas**
  - Green, urban areas

Figure 10. Thematic classes collapsing across scales.

## ACKNOWLEDGEMENTS

The authors express their gratitude to Barbara P. Battenfield from the University of Colorado for her insightful and constructive comments on an earlier draft of this work. In addition, the comments by William A. Mackaness from the University of Edinburgh are kindly acknowledged.

## REFERENCES

Batini, C., M. Lenzerini, & S.B. Navathe. "A Comparative Analysis of Methodologies for Database Schema Integration," *ACM Computing Surveys*, Vol.18-4:323-364, 1986.

Bishr, Y. A., H. Pundt, W. Kuhn, M. Molenaar, & M. Radwan. "Probing the Concept of Information Communities- A Road Towards Semantic Interoperability," paper presented in INTEROP'97 Conference, Santa Barbara, December 1997. URL: <http://www.ncgia.ucsb.edu/conf/interop97/program/>

Battenfield, B. P. "Object-oriented map generalization: modelling and cartographic considerations," in *GISDATA I: GIS AND GENERALIZATION Methodology and Practice*, J.C. Müller, J.P. Lagrange and R. Weibel (eds.), Taylor & Francis: London, 1995.

CORINE Land Cover-Technical Guide, Published by the European Commission, EUR 12585 EN, Luxembourg, 1994.

Deogun J.S., V.V. Raghavan, & H. Sever. "Association Queries and Formal Concept Analysis," The Sixth International Workshop on Rough Sets, Data Mining and Granular Computing (in conjunction with JCIS'98), 23-28 October, Research Triangle Park, NC, USA, 1998.

Devogele T., J. Trevisan, and L. Raynal. "Building a multi-scale database with scale-transition relationships," in *Advances in GIS research II (Proceedings of the Seventh International Symposium on Spatial Data Handling)*, M.J. Kraak, M. Molenaar (eds.), Taylor & Francis: London, 1997.

Digital Geographic Information Working Group (DGIWG). "Digital Geographic Information Exchange Standard (DIGEST) Standards Specification", Part 4, Edition 2.0, NIMA, June 1997.

Faid M., R. Missaoui & R. Godin. "Mining Complex Structures Using Context Concatenation in Formal Concept Analysis." *Second International KRUSE Symposium (KRUSE'97)*, Vancouver, British Columbia, August 11-13, 1997.

Ganter B. & R. Wille. "Conceptual Scaling," in *Applications of Combinatorics and Graph Theory in the Biological and Social Sciences*, F. Roberts (ed.) Springer, New York, 1989.

Govorov M. O. "Representation of the generalized data structures for multi-scale GIS," in *Proc. of the 17<sup>th</sup> ICA/ACI*, Barcelona, 1995.

Hellenic Mapping and Cadastral Organization. Ministry of the Environment, Planning and Planning Works. "Technical Specifications for the Greek National Cadastre-Land Use Classification," 1996 (in Greek).

- Kilpeläinen T., & T. Sarjakoski. "Incremental generalization for multiple representations of geographical objects," in *GISDATA I: GIS AND GENERALIZATION Methodology and Practice*, J.C. Müller, J.P. Lagrange and R. Weibel (eds.), Taylor & Francis: London, 1995.
- Kokla, M., & M. Kavouras. "A Formal Method for the Semantic Integration of Geospatial Classification Schemata," submitted for publication, 1999.
- Kent R.E. & C. Neuss. "Creating a 3D Web Analysis and Visualization Environment," *Computer Networks and ISDN Systems*, Vol. 28: 109-117, 1995.
- Molenaar, M. "The role of topologic and hierarchical spatial object models in database generalization," in *Methods for the Generalization of Geo-Databases*, Number 43, M. Molenaar (ed.), Netherlands Geodetic Commission: the Netherlands, 1996.
- Müller, J.C., R. Weibel, J.P. Lagrange, & F. Salgé. "Generalization: state of the art and issues," in *GISDATA I: GIS AND GENERALIZATION Methodology and Practice*, J.C. Müller, J.P. Lagrange and R. Weibel (eds.), Taylor & Francis: London, 1995.
- Pitoura, E., O. Bukhres, and A. Elmagarmid. "Object Orientation in Multidatabase Systems," *ACM Computing Surveys*, Vol.27-2:141-195, 1995.
- Priss U. "Efficient Implementation of Semantic Relations in Lexical Databases," *Computational Intelligence*, Vol. 15-1, 1999.
- Reddy, M.P., B.E. Prasad, P.G. Reddy, and A. Gupta. "A Methodology for Integration of Heterogeneous Databases," *IEEE Transactions on Knowledge and Data Engineering*, Vol.6-6:920-933, 1994.
- Ruas A. & J.P. Lagrange. "Data and knowledge modelling for generalization," in *GISDATA I: GIS AND GENERALIZATION Methodology and Practice*, J.C. Müller, J.P. Lagrange and R. Weibel (eds.), Taylor & Francis: London, 1995.
- Schmitt, I., & G. Saake. "Merging Inheritance Hierarchies for Schema Integration based on Concept Lattices," *Technical Report*, Faculty of Information, University of Magdeburg, 1997.
- Spaccapietra, S., C. Parent, and Y. Dupont. "Model Independent Assertions for Integration of Heterogeneous Schemas," *VLDB Journal*, Vol.1-1:81-126, 1992.
- Spangenberg, N., & K.E. Wolff. "Concept lattices as indicators of change in the therapeutic process: does formal concept analysis of repertory grids represent a paradigm change of data evaluation?" in *Psychoanalytic research by means of formal concept analysis*, N.Spangenberg, K.E. Wolff (eds.), Sigmund-Freud-Institut, Lit Verlag, Munster, 1999 (to appear).
- Uitermark, H., P. Van Oosterom, N. Mars & M. Molenaar. "Propagating Updates: Finding Corresponding Objects in a Multi-source Environment," in *Proc. International Symposium for Spatial Data Handling (SDH '98)*, Vancouver, Canada, July 1998.

Voisard, A., & H. Schweppe. "Abstraction and decomposition in interoperable GIS," *Int. J. of Geographical Information Science*, Vol.12-4:315-333, June 1998.

Weibel R. "Three essential building blocks for automated generalization," in *GISDATA I: GIS AND GENERALIZATION Methodology and Practice*, J.C. Müller, J.P. Lagrange and R. Weibel (eds.), Taylor & Francis: London, 1995.

Weibel R. & G. Dutton. "Generalising spatial data and dealing with multiple representations," in *GEOGRAPHICAL INFORMATION SYSTEMS Principles and Technical Issues*, P.A. Longley, M. F. Goodchild, D.J. Maguire and D.W. Rhind (eds.), John Wiley & Sons, Inc., 1999.

Wille, R. "Concept Lattices and Conceptual Knowledge Systems," *Computers and Mathematics with Applications*, Vol.23-6-9:493-515, 1992.