# Ontology-Driven Generalization of Cartographic Representations by Aggregation and Dimensional Collapse

Eric B. Wolf

Center of Excellence in GIScience US Geological Survey ebwolf@usgs.gov

**Abstract.** Automatic generalization of cartographic features has been recognized as a goal of Geographic Information Science (GIScience). Many successful algorithms have been introduced for generalization tasks such as point reduction and smoothing of linear features. Such algorithms operate well as a function of change in map scale or resolution. Other generalization tasks have proved considerably more difficult. Two of these operations, aggregation and dimensional collapse, are trivial to implement – replacing a set of points with an area feature or replacing an area feature with a single point – but have proven challenging to make operational. The decision to aggregate or collapse features is as much dependent on the context of the features as they are change in map scale. This dissertation proposes to show how ontologies can be used to inform automated generalization in these operations.

Keywords: ontologies, cartographic generalization, automated generalization.

#### 1 Problem Statement

This dissertation proposes to link two significant areas of research: automated generalization and knowledge engineering with ontologies. Cartographic generalization is often split into two categories: feature modification based on geometry (e.g., sinuosity or area-perimeter ratios) and feature modification based on semantics. This study focuses on two generalization operations: aggregation and dimensional collapse. These two operations occur in response to constraints which can be measured empirically. But as Bertin [2] noted, altering the dimensionality of representation – what he refers to as implantation – is more of a conceptual change. It is the goal of this dissertation to demonstrate how ontologies can be used to constrain or enhance the conceptual changes during aggregation and dimensional collapse.

The intellectual merit of this effort lies in the application of current knowledge engineering methods to automated generalization. Generalization is well-documented as a process that is strongly dependent on the semantics of the information being modified. Ontologies have advanced knowledge in creating a means to encode such semantics in a form that is machine readable and that can guide automated processes. Efforts have been made to encode cartographic rules in ontologies [14] [7]. Other

efforts have involved applying semantic relationships as an input into generalization operations [8]. No known effort has been made to extend existing cartographic task ontologies specifically toward generalization.

The broader impacts of this study lie in connecting the semantics of geographic information with the generalization process in a manner that can be automated. National Mapping Agencies (NMA), such as the United States Geological Survey (USGS), Ordnance Survey of the United Kingdom or Institut Geographique National of France, currently maintain multiple databases representing the same geographic extents at different levels of detail. The ability to derive all levels of detail from a single database would result in significant cost savings and reduction in update errors. The effort will directly involve the research goals of the USGS towards future versions of *The National Map*. Specifically, the effort could bridge existing research projects in the USGS in generalization and ontology development.

### 1.1 Generalization of Cartographic Representations

Data collected for a database with a scale of 1:5,000 would be too detailed for a map at a scale of 1:100,000. A representation appropriate for 1:5,000 scale would need to be generalized to a representation appropriate at 1:100,000 scale. Cartographic generalization is the process in which data is transformed to better represent geographic phenomena that are most significant for the map purpose or scale. Dimensional collapse and aggregation are related in that they alter the geometric dimensionality of the representations. Dimensional collapse involves reducing the geometric dimensionality of a representation: a two-dimensional polygon representing the borders of a city at fine scales may be replaced with a single zero-dimensional point symbol at coarser scales. Aggregation involves replacing collections of similar point symbols with a polygonal representation denoting a field of similar objects (Figure 1) [10]. These operations may be invoked due to constraints or for abstraction. Constraints are the limitations imposed by the communication medium (paper, screen, color palette, etc.) and human perception (semiotics) on how the abstracted data is communicated. Abstraction refers to the process in which the surveyor or data collector gather information and structure it according to demands of the desired end product. For instance, the location of wells, irrigation ditches and land parcels might be collected as locations and areas stored abstractly as points and polygons for the purpose of managing water rights [3].

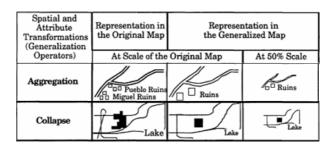


Fig. 1. Aggregation and Collapse generalization operations [10]

Bertin [2] divides cartographic generalization into two categories: structural generalization and conceptual generalization. Structural generalization involves geometric or symbolic changes to feature representation without impacting the semantic relationship with other features. Conceptual generalization requires not just symbolic or geometric changes, but a change in "implantation" [2]. A change in implantation means a change in representation from one primitive type, point, line or area, to another. Collapsing an area to a point or aggregating a collection of points into an area are examples of change in implantation [12]. Bertin provides the example of 'mines' symbolized by points and then aggregated into an area symbol for a 'coal field'. Structural generalization may alter the number of point symbols for mines but a conceptual generalization replaces the points entirely with an area symbol representing the coal field.

Similarly, Ratajski [11] divides generalization into the quantitative and the qualitative. Quantitative generalization is defined as "a gradual reduction of map contents depending on the reduction of the map scale" whereas qualitative generalization is given as "turning from the more elementary ideas to the more general ones". Continuing with Bertin's example, the application of quantitative generalization would result in a reduction of the number of mine symbols whereas the application of qualitative generalization would replace the concept of individual mines with a new concept of a 'coal field'.

Weibel [16] also creates a binary division describing generalization operators as either *context-independent* or *context-dependent*. Specifically, the generalization operators selection/elimination, simplification and smoothing are termed context-independent because they can be applied to individual feature representations without respect to surrounding features. Aggregation and displacement are examples of context-dependent operators that are only used when the surrounding spatial context necessitates it.

Mackaness [9] states "ultimately the results of generalisation do manifest themselves through the manipulation of geometric primitives. But current thinking argues that the reasoning behind those manipulations needs to be based on the analysis of the context." Thus, structural or quantitative generalization can be automated through the application of computational geometry but conceptual or qualitative generalization requires a context-oriented, semantic approach.

Dutton and Edwardes [4] declare "semantic properties and relations needed for generalization that cannot be inferred from geometry must be coded explicitly." In other words, the semantic relationships that govern generalization cannot always be determined through mathematical manipulation of the representations. These semantic relationships must be explicitly established. Further, in order to automate the generalization process the relationships must be encoded in a machine-readable format.

#### 1.2 Ontologies for Cartographic Generalization

According to Smith [13], in the philosophical sense, ontology is the branch of metaphysics that "seeks to provide a definitive and exhaustive classification of entities in all spheres of being." In contrast to Smith, Gruber [5] defines an ontology as a formal, explicit specification of a conceptualization. In this definition, the term *conceptualization* refers to a model of some phenomenon constructed through the identification and enumeration of relevant concepts. The ontology should be *explicit* 

in that each concept and constraints on its use should be clearly delineated. This explicit conceptualization should be encoded *formally* – that is, the ontology should be machine-readable. The term *conceptualization* appears similar to both a map and a spatial database. Both provide an "abstract model of some phenomenon in the world" and both, through cartographic or model generalization identify "the relevant concepts of that phenomenon". In the case of a map, the conceptualization, or abstract model of some phenomenon in the world, is encoded in geometric shapes, patterns of color and texture. The kinds and categories in this abstract model are explicitly defined in the map legend but the constraints on their use are only loosely defined as visual variables and cartographic guidelines. But the conceptualization encoded as a map is not formal or machine-readable In the case of the spatial database, the data model nicely equates to the abstract model of some phenomenon in the world. The data model of a spatial database is also explicit and formal. The data model explicitly defines the concepts used and the constraints on their use. These data models are also machine readable but generally machine readable by specific database systems.

Gruber's definition of ontology has been made operational through formal languages like the Web Ontology Language OWL. By encoding conceptual relationships for a domain ontology in OWL, semantic relationships can be evaluated during automated processing. Automated generalization methods can use these linkages to maintain the semantic character of representations while reducing detail [15].

The application of ontologies to generalization and even the use of ontologies is not a new concept. Andrienko and Andrienko [1] develop what they term "the first system on automated mapping considering semantic relationships among data components." Their Descartes system automatically creates thematic maps based on semantic relationships of demographic statistics. Kulik et al. [8] present a novel algorithm for simplification of linear features that takes into account map purpose as encoded in the semantic weights. Other efforts focus on developing domain ontologies to facilitate cartographic generalization and representation in spatial databases. Dutton and Edwardes [4] state that "... without accounting for the *roles* of features in a landscape and in a map, it is difficult to select, simplify, displace and re-symbolize features appropriately." They develop an ontology based on the place names of features common between USGS topographic maps and NOAA navigational charts.

Torres et al. [14] describe domain ontologies for hydrological and topographic maps. The goal of their exercise is to explore representations of these maps as ontologies. Iosifesci-Enescu and Hurni [7] introduce ontologies for the formalization of cartographic knowledge and rules. They demonstrate both a cartographic domain ontology "centered on concepts of map, graphic element, visual variable and symbol" and point towards a task ontology that formalizes "the storage, access and exchange cartographic rules."

This dissertation will attempt tasks to help inform how semantic relationships drive the generalization operations of aggregation and dimensional collapse. While these operations are typically employed to maintain the presence of a feature representations in response to reduction in map scale, they are sometimes employed in response to changes in context according to the purpose of the map [4].

## 2 Research Questions

The questions this dissertation addresses are:

- 1. Can the semantic relationships encoded as ontologies be used to automate dimensional collapse and aggregation?
- 2. Separate domain ontologies must be created for each separate map series. Can a single task ontology for generalization be created to manage the semantics for all map scales? How is generalization modeled inside this ontology?
- 3. If multiple ontologies must be created, how can the ontologies be aligned to represent continuity across levels of detail?

## 3 Workplan

As a demonstration of the concepts in this dissertation, application ontologies will be developed for air navigation charts similar to the domain ontologies created by Torres et al. [14]. Air navigation charts are ideal for this case study because the database and scale remain fixed while representation is varied. The charts represent essentially the same information but with two unique perspectives. Visual Flight Rules (VFR) charts are designed for conditions in which pilots can navigate by visually surveying the landscape. Instrument Flight Rules (IFR) charts are designed for conditions in which pilots must rely solely on instrument readouts (Figure 2).



Fig. 2. Section of IFR and VFR Charts over Jefferson County Airport. Two cartographic ontologies based on one spatial database model.

Once the demonstration ontologies are completed, cartographic application ontologies will be created for multi-scale map series produced by the USGS (Figure 3). To limit the scope of the study, focus will be on feature complexes in and around selected international airports. A single ontology will be created for each multi-scale map series by examining the official data dictionaries. The availability of the official data dictionaries may further limit the scope of this task. A significant question arises from this process: Is it better to have one ontology that manages generalization and change in scale internally? Or is it better to have multiple ontologies with the generalization process becoming a process of creating semantic agreements between the ontologies?



**Fig. 3.** San Francisco International Airport in USGS Topographic multi-scale map series. Left to right 1:24,000 7.5" Quandrangle, 1:100,000 30' x 60' Quadrangle, and 1:250,000 1 degree by 2 degree Quadrangle, presented as uniform-scale graphics.

Finally, task ontologies will be constructed for cartographic generalization operations used in each application ontology. Relationships among the domain ontologies developed in this study will be established. For instance, a domain ontology for the spatial databases underlying the USGS *National Map* could be aligned with different cartographic domain ontologies. The same will be done for the generalization task ontologies. Theoretically, such a system will allow for the generation of ontologically-dissimilar maps like the IFR and VFR charts as well as guiding the generalization of data for multi-scale map series.

In the first half of this research project, a series of domain and task ontologies are to be developed for creating similar maps. Several iterations of each ontology may be necessary to balance the requirements of each criterion. Gruber [6] gives objective criteria for evaluating the design of formal ontologies. The ontologies developed in this dissertation will be evaluated against these criteria

The demonstration task involves the flight navigation charts controls for scale and underlying database while varying map purpose. It is anticipated that a single domain ontology may be constructed for both VFR and IFR perspectives. Also, a task ontology for cartographic representation would be created based on a subset of the cartographic rules used on the flight navigation charts similar to Iosifescu-Enescu and Hurni's ontology [7]. The differences in the perspectives would be encoded in two application ontologies based on these domain and task ontologies.

The main research effort involves creating ontologies for multi-scale map series. This effort will be considerably more challenging than the demonstration, not just because of the number of maps involved, but because at least two different models are possible. In both models, it is anticipated that a single domain ontology will be needed for each map series. The cartographic task ontology from the demonstration would be expanded to include generalization operators. Task ontologies encode relationships among tasks but not specific sequences. Application ontologies rely on these relationships by the tasks to concepts in the domain ontology. Thus, application ontologies would be created for either each map or each map series depending on whether or not a single application ontology can be created for an entire map series.

Once completed, the final effort would be to utilize the hierarchy of domain and task ontologies to explore relationships between scale and ontology in order to answer the research questions. One method of exploring these relationships would be to map the successful ontologies onto heterogeneous database ontologies for the purpose of

guiding the generalization process. The overall goal of the project is to guide the conceptual generalization of geographic representations through the semantic relationships encoded in the ontologies. In order to evaluate this goal, it will be necessary to develop programs that utilize ontologies in automating the generalization process.

## 4 Summary

This dissertation is designed to demonstrate how ontologies can be used to constrain or enhance the conceptual changes during aggregation and dimensional collapse. These two generalization operations modify the conceptual nature of representations and must be guided by semantics, not geometry [11] [2] [16] [9].

The expected outcomes of this research are the creation of a series of domain ontologies from data dictionaries published by NMAs as well as one or more task ontologies which model generalization within the context of cartography. These ontologies will be aligned as application ontologies that will be coupled with scripts to drive automated generalization for dimensional collapse and aggregation.

**Acknowledgments.** This research was supported by the United States Geological Survey as part of the project titled "Building an Ontology for *The National Map*".

#### References

- Andrienko, G., Andrienko, N.: Knowledge Engineering for Automated Map Design in Descarates. In: ACM GIS 1999, Kansas City, MO USA (1999)
- 2. Bertin, J.: Semiology of Graphics, p. 415. The University of Wisconsin Press, Madison (1983)
- Buttenfield, B.P., Mark, D.M.: Expert Systems in Cartographic Design. In: Geographic Information Systems: The Microcomputer and Modern Cartography, pp. 129–150. Pergamon Press, Oxford (1990)
- 4. Dutton, G., Edwardes, A.: Ontological Modeling of Geographical Relationships for Map Generalization. In: The 9th ICA Workshop on Generalization and Multiple Representations, Vancouver, Washington (2006)
- Gruber, T.R.: A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5, 199–220 (1993)
- Gruber, T.R.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies 43(5/6), 22 (1995)
- 7. Iosifescu-Enescu, I., Hurni, L.: Towards Cartographic Ontologies Or How Computers Learn Cartography. In: 23rd International Cartographic Conference, Moscow, Russia (2007)
- 8. Kulik, L., Duckham, M., et al.: Ontology-Driven Map Generalization. Journal of Visual Languages and Computing 16(2), 245–267 (2005)
- 9. Mackaness, W.A., Ruas, A., et al. (eds.): Generalisation of Geographic Information: Cartographic Modelling and Applications, p. 370. Elsevier, Oxford (2007)
- 10. McMaster, R.B., Shea, K.S.: Cartographic Generalization in a Digital Environment: When and How to Generalize. AutoCarto 9, Baltimore, MD (1989)

- Ratajski, L.: Phenomenes Des Points De Generalization. International Yearbook of Cartography 7, 143–151 (1967)
- 12. Slocum, T.A., McMaster, R.B., et al.: Scale and Generalization. In: Thematic Cartography and Geographic Visualization, pp. 103–120. Prentice Hall, Upper Saddle River (2005)
- 13. Smith, B.: Ontology. In: Blackwell Guide to the Philosophy of Computing and Information, pp. 155–166. Blackwell, Malden (2003)
- 14. Torres, M., Quintero, R., et al.: Ontology-Driven Description of Spatial Data for Their Semantic Processing. In: GeoSpatial Semantics 2005. Springer, Heidelberg (2005)
- 15. Tran, V.X., Tsuji, H.: Owl-T: A Task Ontology Language for Automatic Service Composition. In: IEEE International Conference on Web Services. IEEE, Los Alamitos (2007)
- 16. Weibel, R.: Generalization of Spatial Data: Principles and Selected Algorithms. In: van Kreveld, M., Roos, T., Nievergelt, J., Widmayer, P. (eds.) CISM School 1996. LNCS, vol. 1340, pp. 99–152. Springer, Heidelberg (1997)