Chapter LII
Current Approaches and Future Trends of Ontology-Driven Geographic Integration[1]

Agustina Buccella
Universidad Nacional del Comahue, Argentina

Alejandra Cechich
Universidad Nacional del Comahue, Argentina

INTRODUCTION

Currently there are many domain areas in Computer Science interested in the integration of various information sources. Federated Databases, Semantic Web, and Automated Web Services are some of them. Particularly in the geographic information area, newer and better technologies and devices are being created in order to capture a large amount of information about Earth. All of this geographic information is analyzed and stored at various levels of detail in Geographic Information Systems (GISs), possibly distributed on the Web. Then a fast search for geographic information on the Web will return several links representing parts of our world. But what happens when someone needs information that is divided into more than one system? For example, information about rivers in a country can be obtained by querying two or more systems. Although distribution of information is one of the main problems, there are some others; these systems have been developed by various entities with different points of view and vocabularies, and here is when face heterogeneity problems arise. They are encountered in every communication between interoperating systems where interoperability refers to interaction between information from various sources involving the task of *data integration* to combine data.

Generally speaking, two essential tasks are involved in the data integration process: *semantic enrichment* and *mapping discovery* (Sotnykova, Vangenot, Cullot, Bennacer & Aufaure, 2005). The main goal of the first one is to reconcile semantic heterogeneity so it involves adding more semantic information about the data. Various proposed approaches add extra semantic information through the use of metadata or ontologies. For example, proposals extending common data models such as Entity-Relationship diagrams (Parent, Spaccapietra & Zimányi, 1999, 2005) and object-oriented ones (Borges, Davis & Laender, 2001) have been presented in order to add geographic features. We are particularly interested in those using ontologies because, by definition, they provide a vocabulary to represent and communicate knowledge about the domain and a set of relationships containing the terms of the vocabulary at a conceptual level. Ontologies are currently extensively proposed as tools to face heterogeneity problems; for example, different proposals are using formal ontologies to enrich the conceptual schema and thus improve the integration (Fonseca, Davis & Camara, 2003; Fonseca, Egenhofer, Agouris & Camara, 2002; Hakimpour, 2003; Hakimpour & Geppert 2002) and the query process (Zhang, 2005).

The semantic enrichment task is essential to reach the second task, *mapping discovery.* Several surveys (Kalfoglou & Schorlemmer, 2003; Klein, 2001) have been presented that analyze various proposals related to semantic matching (i.e., building of mappings); however, they do not take into account geographic information. Considering the hypothesis tested by Mark, Skupin, and Smith (2001) in which "geographic and non-geographic entities are ontologically distinct in a number of ways," a different analysis must be performed when geographic elements are included.

In this chapter, we analyze several proposals that consider geographic information as sources to be integrated. First, we briefly describe basic concepts and conventions that will be used throughout this chapter. Following, an analysis is performed according to the use of ontologies in an integration process of geographic sources. Finally, future trends and conclusions are revealed as a consequence of our analysis.

GIS INTEGRATION: BASIC CONCEPTS

The concept of *data integration* is concerned with unifying data that share some common semantics but originate from unrelated sources. In every integration process, heterogeneity is one of the most common problems. Let us consider two systems sharing data representing rivers, which can be an example to clarify different types of heterogeneity problems (Hakimpour, 2003):

- **Heterogeneity in the conceptual model.** One system represents a river as an object class and the other as a relationship.
- **Heterogeneity in the spatial model.** Rivers can be represented by polygons (or a segment of pixels) in one system, while they are represented by lines in the second system.
- **Structure or schema heterogeneity.** Both systems hold the name of a river, but one of them also keeps information about the border.
- **Semantic heterogeneity.** One system may consider a river as a natural stream of water larger than a creek with a border, and the other defines a river as any natural stream of water reaching from the sea, a lake, and so forth, into the land.

In order to solve several of these heterogeneities, we briefly introduce the concept of *ontologies* as "a formal explicit specification of a shared conceptualization" (Gruber, 1993). A *conceptualization* refers to an abstract model of