Spatial datacubes extend the datacube concept underlying the field of Business Intelligence (BI) into the realm of spatial analysis, geographic knowledge discovery, and spatial decision-support. The traditional computer science community has defined spatial datacubes and their fundamental components (e.g., spatial dimension and spatial measure) through formal models limiting spatial data as only those data that has a geometric representation. The geomatics community has pursued spatial datacube models with a much richer view of spatial data. However, the proposed models by the geomatics community have not yet been formalized using precise mathematical languages. This paper, for the first time, integrates the rigor of mathematical languages with the richer view of spatial data to provide a formal model and precise definitions of spatial datacubes and their fundamental components. The proposed definitions provide the scientific community with a common and precise terminology for the concepts involved in spatial decision-support databases.

1. Introduction

Strategic decision makers (analysts, executives, and managers) need to analyze and compare summarized data extracted from very large volumes of data. Indeed, it is more efficient to use aggregated and consolidated data covering a certain period of time rather than detailed individual records of transactional databases for strategic decision making. The difficulty in supporting both daily transactions and decision-support needs within a single database requires using a dual-database approach. This forms the typical backbone of data warehouses [Bédard and Han 2008]. A data warehouse is a subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organisational decision making [Chaudhuri and Dayal 1997]. Data warehouses are typically modeled using the datacube (or multidimensional, in the sense of business intelligence) paradigm [Gray et al. 1997; Abelló et al. 2006]. In the datacube structure, analysis is performed along a combination of axes of analysis called dimensions (e.g., categories of products, administrative regions, periods), and hence the structure is termed multidimensional. Each dimension includes one or several hierarchies, each composed of different analysis levels (e.g., city-province-country hierarchy and city-county-region-country hierarchy which may compose a spatial dimension labelled “administrative regions”). The hierarchical structure allows users to view and analyze data at different levels of detail. An instance of a level is a member (e.g., “Montreal” is a member of the level “city”). Measures (e.g., population) are measurable quantities; these are analyzed against the members of different levels of dimensions. For instance, one may be interested in analyzing the measure “population” with respect to different levels of “administrative regions” and “time” dimensions.