Implementation of UML Schema in Relational Databases: A Case of Geographic Information

Maurizio Nagni, Centre for Environmental Data Archival (CEDA), Didcot, UK
Spiros Ventouras, Centre for Environmental Data Archival (CEDA), Didcot, UK

ABSTRACT

Numerous disciplines require information concerning phenomena implicitly or explicitly associated with a location relative to the Earth. Disciplines using Geographic Information (GI) in particular are those within the earth and physical sciences, and increasingly those within social science and medical fields. Therefore geographic datasets are increasingly being shared, exchanged and frequently re-purposed for uses beyond their original intended use. Being part of the ISO 19100 Geographic Information Standard series, the ISO 19136 called Geography Markup Language (GML), defines the rules a data model described using the Unified Modeling Language (UML) has to follow in order to generate from it an XSD schema. However, if GML is essential for exchange data among different organization, it may not be the best option for persisting or searching operations. On the other side, the Relational Database Model (RDBM) has been heavily optimized over the decades to store and search data. This paper does not address “How to store an GML complaint document in an RDBM” but “How to realize an RDBM from an ISO 19100 complaint UML data model” and within this context, it describes the experience and the lessons learnt. The conclusions show how the information contained in such UML is able to produce not only representations as GML schema, but also RDBM or RDF without passing by any intermediary step.

Keywords: Geographic Datasets, Geographic Information (GI), Geography Markup Language (GML), Relational Database Model (RDBM), Unified Modeling Language (UML)

INTRODUCTION

There is a great demand for Geographic Information (GI), i.e., information concerning phenomena implicitly or explicitly associated with a location relative to the Earth, in a wide range of applications within multiple disciplines. These are mainly those within the earth and physical sciences, but also they can be found within social science and medical fields. Furthermore, the widespread application of computers has led to an increased use of geographic data. Thus, geographic datasets are increasingly being shared, exchanged and frequently used for purposes other than those for which they were originally intended.

As a consequence, standardization is required to enable the universal and effective

DOI: 10.4018/ijdst.2013100105
usage of digital geographic information and to facilitate interoperability of geographic information systems, including interoperability in distributed computing environments.

Standardization guidelines for the geographic information are provided by two primary authorities: the ISO Technical Committee 211 (ISO/TC 211) and the Open Geospatial Consortium (OGC). In particular, ISO/TC 211 governs mainly abstract standards for cross-domain applications while OGC concentrates on architectures for distribution of geospatial services and implementation of ISO standards through service interfaces, data models and encodings.

ISO/TC 211 has established the ISO 19100 series (ISO 19100, 2004), a set of structure standards, for a) defining the basic semantics and structure of geographic information for data management and data interchange purposes and b) defining geographic information service components and their behavior for data processing purposes.

Rules for creating a formal description of a data structure and content required by one or more applications in a consistent manner, i.e. an application schema, are defined by the ISO 19109 (ISO 19109, 2005) together with the ISO 19101 (ISO 19101, 2002). This standard specifies a reference model for structuring geographic information standards.

Other standards in the ISO 19100 series define reusable modules of conceptual schemas that may be integrated in an application schema according to the requirements of the application. For example, the ISO 19115 (ISO 19115, 2006) defines general-purpose metadata while the ISO 19108 (ISO 19108, 2002) defines the temporal characteristics of geographic information.

At the conceptual stage, the ISO/TC 211 process of creating an application schema requires a profile of the UML (Object Management Group, 2012) as defined in the ISO 19103 (ISO 19103, 2005). This satisfies the goal of ISO/TC 211, to create a framework to enable syntactic interoperability and to support semantic interoperability, while supporting multiple exchanging formats and multiple service implementations.

As stated in the ISO 19100 series fundamental document “In order to provide more precise definition and understanding, the Domain reference model is described using graphical notation of UML. This is intended for developers of geographic information standards who will use or extend the ISO 19100 series as well as for those who wish to have an in-depth knowledge of this family of standards. […] For the ISO 19100 series, UML is the conceptual schema language whose meta-models conform to the normative object model conceptual formalism” (ISO 19100, 2004). Throughout this paper UML data models (UMLm) are assumed to be compliant with of ISO/TC 211 process.

The ISO 19136 (ISO 19136, 2007) defines the GML as a markup language, like is the XML, and a set of rules that an UMLm has to follow in order to implement the UMLm in a GML document. Using software applications like Fullmoon Golodoniuc and Cox, 2012, ShapeChange ShapeChange, 2012 or Newmoon Newmoon, 2012, it is possible to generate an XML schema (XSD) describing the UMLm’s GML enabling any user to create, validate, import and export documents conforming to the organization UMLm. This approach supports the so called Model-Driven Approach (MDA) (Kleppe et al., 2003) allowing the developer of the UMLm to focus his efforts to describe the domain and business requirements leaving its implementation, that is the XSD, to standard tools, i.e. Newmoon.

Describing geographic information using GML is essential for data exchange between different users; however, it may be not always an exclusive option for persisting or searching operations. When a great number of documents have to be stored or searched a database is often a possible solution. Many types of database are available: from the classical Relational Databases (Postgresql, MySql, etc.) to more exotic flavors like NoSQL (like MongoDB or Cassandra) or native XML (eXist). Nowadays relational databases are extensively used and largely dominate the market over so called non-relational DB Florotou et al., 2012 flavors but even if non-relational databases are expected to significantly grow on the long term, relational
Related Content

A Proposal for Information Systems Security Monitoring Based on Large Datasets
www.igi-global.com/article/a-proposal-for-information-systems-security-monitoring-based-on-large-datasets/202380?camid=4v1a

BAM-Chord: DHT-Based Bandwidth Adaptive Multicast System
www.igi-global.com/article/bam-chord-dht-based-bandwidth/76923?camid=4v1a
Efficient Update Control of Bloom Filter Replicas in Large Scale Distributed Systems
www.igi-global.com/chapter/efficient-update-control-bloom-filter/36434?camid=4v1a

A Grid and Cloud Based System for Data Grouping Computation and Online Service
www.igi-global.com/chapter/grid-cloud-based-system-data/69044?camid=4v1a