Chapter 9
New Dimension in Relational Database Preservation: Using Ontologies

Ricardo André Pereira Freitas
Lusíada University, Portugal

José Carlos Ramalho
University of Minho, Portugal

ABSTRACT
Due to the expansion and growth of information technologies, much of human knowledge is now recorded on digital media. A new problem in the digital universe has arisen: Digital Preservation. This chapter addresses the problems of Digital Preservation and focuses on the conceptual model within a specific class of digital objects: Relational Databases. Previously, a neutral format was adopted to pursue the goal of platform independence and to achieve a standard format in the digital preservation of relational databases, both data and structure (logical model). The authors address the preservation of relational databases by focusing on the conceptual model of the database, considering the database semantics as an important preservation “property.” For the representation of this higher layer of abstraction present in databases, they use an ontology-based approach. At this higher abstraction level exists inherent Knowledge associated to the database semantics that the authors tentatively represent using “Web Ontology Language” (OWL). From the initial prototype, they develop a framework (supported by case studies) and establish a mapping algorithm for the conversion between databases and OWL. The ontology approach is adopted to formalize the knowledge associated to the conceptual model of the database and also a methodology to create an abstract representation of it. The system is based on the functional axes (ingestion, administration, dissemination, and preservation) of the OAIS reference model.

DOI: 10.4018/978-1-4666-2669-0.ch009
INTRODUCTION

In the current paradigm of information society more than one hundred exabytes of data are used to support information systems worldwide (Manson, 2010). The evolution of the hardware and software industry causes that progressively more of the intellectual and business information are stored in computer platforms. The main issue lies exactly within these platforms. If in the past there was no need of mediators to understand the analogical artifacts, today we depend on those mediators (computer platforms) to understand digital objects.

Our work addresses this issue of Digital Preservation and focuses on a specific class of digital objects: Relational Databases (RDBs). These kinds of archives are important to several organizations (they can justify their activities and characterize the organization itself) and are virtually in the base of all dynamic content in the Web.

In previous work (Freitas & Ramalho, 2009) we adopted an approach that combines two strategies and uses a third technique—migration and normalization with refreshment:

- Migration which is carried in order to transform the original database into the new format—Database Markup Language (DBML) (Jacinto, Librelotto, Ramalho, & Henriques, 2002);
- Normalization reduces the preservation spectrum to only one format;
- Refreshment consists on ensuring that the archive is using media appropriate to the hardware in usage throughout preservation (Freitas, 2008).

This previous approach deals with the preservation of the Data and Structure of the database, i.e., the preservation of the database logical model. We developed a prototype that separates the data from its specific database management environment (DBMS). The prototype follows the Open Archival Information System (OAIS) reference model (by the Consultative Committee for Space Data Systems, 2002) and uses DBML neutral format for the representation of both data and structure (schema) of the database.

Conceptual Preservation

In this work, we address the preservation of relational databases by focusing on the conceptual model of the database (the Information System – IS). It is intended to raise the representation level of the database up to the conceptual model and preserve this representation. For the representation of this higher level of abstraction on databases, we use an ontology-based approach. At this level there is an inherent Knowledge associated to the database semantics that we represent using OWL (McGuinness & Harmelen, 2004). We developed a prototype (supported by case study) and established an algorithm that enables the mapping process between the database and OWL.

In the following section we overview the problem of digital preservation, referring to the digital object, preservation strategies and the preservation of relational databases. Section 3 describes our previous work and states the open issue (database semantic representation) the lead us to the current approach. In Section 4, we outline the relation between ontologies and databases establishing the state-of-the-art and referring to related work. The prototype and the mapping process from RDBs to OWL is detailed in section 5. At the end, we draw some conclusions and specify some of the future work.

DIGITAL PRESERVATION

A set of processes or activities that take place in order to preserve a certain object (digital) addressing its relevant properties, is one of the several definitions. Digital objects have several associated aspects (characteristics or properties) that we
Related Content

Negotiating Early Reuse of Components - A Model-Based Analysis
[www.igi-global.com/chapter/negotiating-early-reuse-components-model/30538?camid=4v1a](www.igi-global.com/chapter/negotiating-early-reuse-components-model/30538?camid=4v1a)

Using Semantics in XML Information Access: Application to the Portuguese Emigration Museum
[www.igi-global.com/chapter/using-semantics-xml-information-access/73180?camid=4v1a](www.igi-global.com/chapter/using-semantics-xml-information-access/73180?camid=4v1a)

Graphical Notations for Rule Modeling
[www.igi-global.com/chapter/graphical-notations-rule-modeling/35855?camid=4v1a](www.igi-global.com/chapter/graphical-notations-rule-modeling/35855?camid=4v1a)

A Framework for Cost-Based Query Optimization in Native XML Database Management Systems
[www.igi-global.com/chapter/framework-cost-based-query-optimization/41504?camid=4v1a](www.igi-global.com/chapter/framework-cost-based-query-optimization/41504?camid=4v1a)