Model-Based Data Engineering for Web Services

Andreas Tolk and Saikou Y. Diallo
Old Dominion University, Norfolk, VA 23529, USA
atolk@odu.edu, sdiallo@odu.edu

The application of the Extensible Mark-up Language (XML) and Web services enabled a new level of interoperability for heterogeneous IT systems. However, although XML enables separation of data definition and data content, it doesn’t ensure that data exchanged are interpreted correctly by the receiving system. This motivates data management to support unambiguous definition of data elements for information exchange. Using a common reference model improves this process leading to Model-based Data Engineering (MBDE). The results can be used immediately to configure mediation layers integrating services into an overall service-oriented architecture. Ultimately, the objective must be to describe the information exchange requirements, contexts, and constraints of web services in metadata allowing intelligent agents to conduct these engineering steps without human support. This chapter describes the current state of the art of MBDE and how it relates to Service Oriented Architectures (SOA) in general and Web Services in particular.

6.1 Introduction

Independently developed and distributed applications each have internal representations of their data. Therefore a transformation layer translating the internal representations into each other has to be created to make information exchange possible between these systems. The traditional approach is to utilize the Extensible Mark-up Language (XML) to enable data exchange between any two systems; nonetheless XML doesn’t ensure that data exchanged are interpreted correctly by the receiving system. Furthermore, XML does not cope with the problem of semantic information exchange. This motivates data engineering to support the unambiguous definition of data elements for information exchange and the definition of a standard approach to mapping heterogeneous data models. This chapter presents an algorithm that can be applied to the data engineering process to ensure correctness in the exchange of bits and bytes but more importantly correctness in the conceptual and semantic exchange.

Practical applications have shown that using a common reference model improves this process leading to Model-based Data Engineering (MBDE). In addition, in order to support operations with rapidly changing requirements, service-oriented architectures are needed instead of the traditional solutions, which are often too inflexible. As an alternative to having a system fulfilling a set of predefined requirements, services fulfilling requirements are identified, composed and orchestrated to meet the current users’ needs in an ongoing operation.
The ideas of MBDE are rooted in federated databases Spaccapietra and colleagues identify the following four classes of conflicts to be solved by data engineering [1]:

- **Semantic Conflicts** occur when concepts of the different local schemata do not match exactly, but have to be aggregated or disaggregated. They may only overlap or be subsets of each other, etc.
- **Descriptive Conflicts** describe homonyms, synonyms, and different names for the same concept, different attributes or slot values for the same concept, etc.
- **Heterogeneous Conflicts** result from substantially different methodologies being used to describe the concepts.
- **Structural Conflicts** results from the use of different structures describing the same concept.

Spaccapietra et al. concluded that a generic meta data model comprising only objects and attributes for values and references is needed to support efficient data management. This generic data model would describe all information exchange requirements and constraints for participating systems, allowing to construct a federated database schema in which (1) all information elements to be exchanged between the systems are modelled in an unambiguous way, and (2) the mapping of these information elements to representing entities in the participating systems is captured unambiguously as well. In order to enable composable services, these ideas must be captured in a framework that consistently captures the required information in the form of metadata.

MBDE was developed in support of integrating Modeling and Simulation (M&S) applications and operational Information Technology (IT) systems to enable decision support – such as alternative simulation and evaluation –, training applications – such as using simulation as a synthetic environment for the trainee using the operational system he is used to –, and testing applications. The second part of this chapter will focus on some special challenges of M&S web services that may go beyond the general integration challenges of web services. Most of this chapter, however, will deal with general challenges of composing web services based on MBDE.

### 6.2 Model-Based Data Engineering for Composable Services

Most Information Technology (IT) specialists consider the ability to merge heterogeneous data models as the single most important challenge in the business world today [1, 2]. The proliferation of independently developed data models and the diversity of data sources and potential consumers, motivate the need for a clearly defined Data Engineering process. Some of the most immediate challenges are:

- **Multiple sources with different formats**: Data are no longer only located in databases or text files. The advent of XML, Web services and Real Simple Syndication (RSS) Streams has added to the variety of formats that need to be handled.
- **Structured, Semi Structured, Unstructured data**: Based on the definitions given in [3], data are considered unstructured when they can be of any type, do not necessarily follow any sequence, format or rule and are not predictable. Data are semi-structured if they are organized in semantic entities and similar entities are grouped together. Entities in the same group may not have the same attributes however, and the order