Preface

This book is a result of choosing the best papers from a conference held in Vila do Conde, Portugal, in June 2011, named XATA (XML, Applications, and Associated Technologies) together with some other original work.

XATA was a conference on XML, covering a wide range of topics, from the low-level details of the language and all its related standards, up to its usage in Digital Libraries or e-Learning. The 2011 edition of XATA was the last in a series of nine editions, giving place to a new project, named SLATE: Symposium on Languages, Applications, and Technologies, which covers not just XML and related technologies, but all other markup languages, as well as programming languages and human (natural) languages. The first edition of this event took place on the 21st and 22nd of June, 2012.

XML has been in the last years the de facto language for interoperability between systems, serializing data structures, construct e-learning materials and repositories, and in a lot other areas of application. In this book, we cover a broad range of subjects related to these aspects. We think this book is interesting mainly because of the wide range of covered topics, and the different aspects and applications of XML than by each one of the articles by itself. It can be seen as a compendium of how XML is being used in the world, widening the reader’s knowledge on XML, its applications, and its associated technologies.

The XML, as we know it today, appeared in 1998, with the first World Wide Web Consortium (W3C) Recommendation. It was the result of making the Standard Generalized Markup Language (SGML) more strict, making it easier to parse, and less ambiguous. This recommendation has evolved during the last years, with the second version of this recommendation in 2000, the third in 2003, the fourth in 2006, and currently, the fifth, that has been stable since 2008. The bases behind the XML philosophy did not change during these years. Just a few enhancements were defined, like the support for namespaces, and the specification on how parsers should report and deal with errors.

Although the XML recommendation has been stable during these years, a lot has been defined and specified related to the XML technologies. Some examples include the XML Path language, which defines a syntax to specify nodes and views in an XML document (second recommendation from 2011), the XML Query language, a powerful query language for XML documents and databases (first W3C recommendation from 2011), the Extensible Stylesheet Language (XSL) for expressing stylesheets (whose version 1.1 recommendation is from 2006), and the XSL Transformations (XSLT), a language to transform XML documents into other XML documents (whose second recommendation is from 2007).

These are just a few examples. Even the way to define the structure or grammar of XML documents has evolved. The first approach, derived from the SGML world, was the use of Document Type Definitions (DTD) that is still widely used, but new approaches have appeared, not just to define the grammar but also the semantics, like the XML Schema Definition (XSD version 1.1 was defined in 2012), the Relax-NG (an ISO standard developed within OASIS group [Advancing Open Standards for the Information Society]).
Note that most of these languages (like XQuery, XSL, XSLT, XSD) are themselves XML dialects. In the dialects world, there is a lot of work as well, ranging from the attempts to transform the HyperText Markup Language (HTML) into an XML dialect (with XHTML), to the specifications on how to define vector graphics with SVG (Scalar Vector Graphics format), to define the structure and the presentation of mathematical formulae with MathML (Math Markup Language), to define music scores with MusicXML, to define how multimedia events are synchronized with SMIL (Synchronized Multimedia Integration Language), and the more recent approaches to encode ontologies into XML defining Resource Definition Framework (RDF), Web Ontology Language (OWL), or even TopicMaps and SKOS (Simple Knowledge Organization System).

The usage of XML to serialize computational objects and foster the interoperability between systems has been growing in an exponential way. XML has been using the text composition, with two relevant standards, DocBook and the Text Encoding Initiative (TEI), in the e-Learning field, with IEEE Learning Object Metadata (LOM) and IMS Question and Test Interoperability (IMS QTI), in the computer assisted translation, with the Translation Memory Interchange format (TMX) and the Term Base Interchange format (TBX), in the health institutions with half a dozen standards from Clinical Data Interchange Standards Consortium, in the e-Publishing industry with the ePub format, etc. Even the major companies like Microsoft switched their proprietary formats for counterpart XML versions.

Nevertheless, XATA and this book are not devoted to the definition of new standards, or just their use, but it, instead, points out problems or challenges on their use and extension and answers or identifies solution paths to these problems. In fact, this book is divided in three main sections, and only the first one is directly related to the usage and definition of standards.

SECTION 1: XML STANDARDS AND DIALECTS

The first section of the book includes four chapters. Two of them present systems that allow the edition of XML dialects such as XSLT and XQuery without the need to understand its syntax and with enough clarity to be easily understood by people who are not familiar with the XML languages. Both cases are based on examples, fostering the learning of these XML dialects.

XML is also used to describe Programming Exercises using an XML Interoperable Language called PExIL. This language covers the entire lifecycle of a programming exercise from its creation to its evaluation. In this lifecycle, several e-learning systems intervene, such as automatic evaluators, repositories of learning objects, and learning management systems.

Finally, an XML language, specifically the Extensible Messaging and Presence Protocol (XMPP) language, is used as the basis for an identity management framework that allows users to control asynchronous and dynamic sharing sensitive data, ensuring the security and privacy in a simple and transparent way.

SECTION 2: LEARNING ENVIRONMENTS

This second section of the book focuses on the assessment of the quality of learning environments and standards. The first chapter describes a framework for quality assessment of learning technology specifications and how it was used to evaluate and improve a case in point: the Learning Path Specification.
The following chapter analyses the quality of XML learning object repositories, more precisely, on the models and methods to evaluate the quality of learning repositories. Multiple criteria decision analysis and optimization methods are explored to be applied for evaluating the quality of learning repositories.

Finally, a chapter on mobile learning is presented where the authors compare the most used approaches for mobile device detection. Based on this study, an architecture for detecting and delivering uniform m-Learning content to students in a high school is presented. This system is based on an XML device capabilities repository and on a REST API Web service for dealing with device data.

SECTION 3: DATABASES AND REPOSITORIES

The final section is dedicated to databases and repositories. The first two chapters focus on relational databases. The former presents the definition of the XML structure, which extends the Software Independent Archiving of Relational Databases (SIARD) format used for the description and archive of relational databases, enriching it with a layer of metadata for the data warehouse components. Data Warehouse Extensible Markup Language (DWXML) is the XML language proposed to describe the data warehouse. The latter addresses the problematic Digital Preservation and focuses on the conceptual model within a specific class of digital objects: Relational Databases. The preservation of relational databases focuses 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, the authors use an ontology-based approach.

The third chapter focuses on the management of research data in a higher education and research institution. The work is based on a small-scale data audit study, which included contacts with researchers and yielded some preliminary requirements and use cases. These requirements led to the design of a data curation workflow involving the researcher, the curator and a data repository based on a DSpace repository.

The last chapter of the book relates an ongoing work, in the context of the Portuguese Emigration Museum, about information access in XML collections associated with semantic information. The museum asset is made up of documents of more than eight kinds, ranging from passport records to photos/cards or building-drawings. The authors discuss the approach used to create the exhibition rooms of the virtual Web-based museum. Each room consists of a view over the information contained in those single or interrelated resources. The information exhibited in each room is described by an ontology written in Web Ontology Language (OWL).

In the end, we think this book encompasses a set of interesting chapters from different XML application areas, that can help researchers and professionals to read and learn from previous experiences. However, this book would not be a reality if it was not for the work and dedication from our editorial advisory board, who reviewed and commented on each of the chapters here published, and for the work and interest by the authors who took the time to expand their chapters and include the editorial advisory board comments into them.