

Main Focus

Making life easier in an XML world

by Denise B. Warzel, Acting
Convenor, ISO/IEC
JTC 1/SC 32/WG 2, Metadata

Extensible Markup Language (XML) is a flexible text format widely used for describing and storing data for interchange on the Web. One of its most important features is its ease of use. Angle brackets <> containing text descriptions – or “tags” – organized into nested structures, encapsulate data to help end users understand its meaning, thus rendering the data within the document “self-descriptive”.

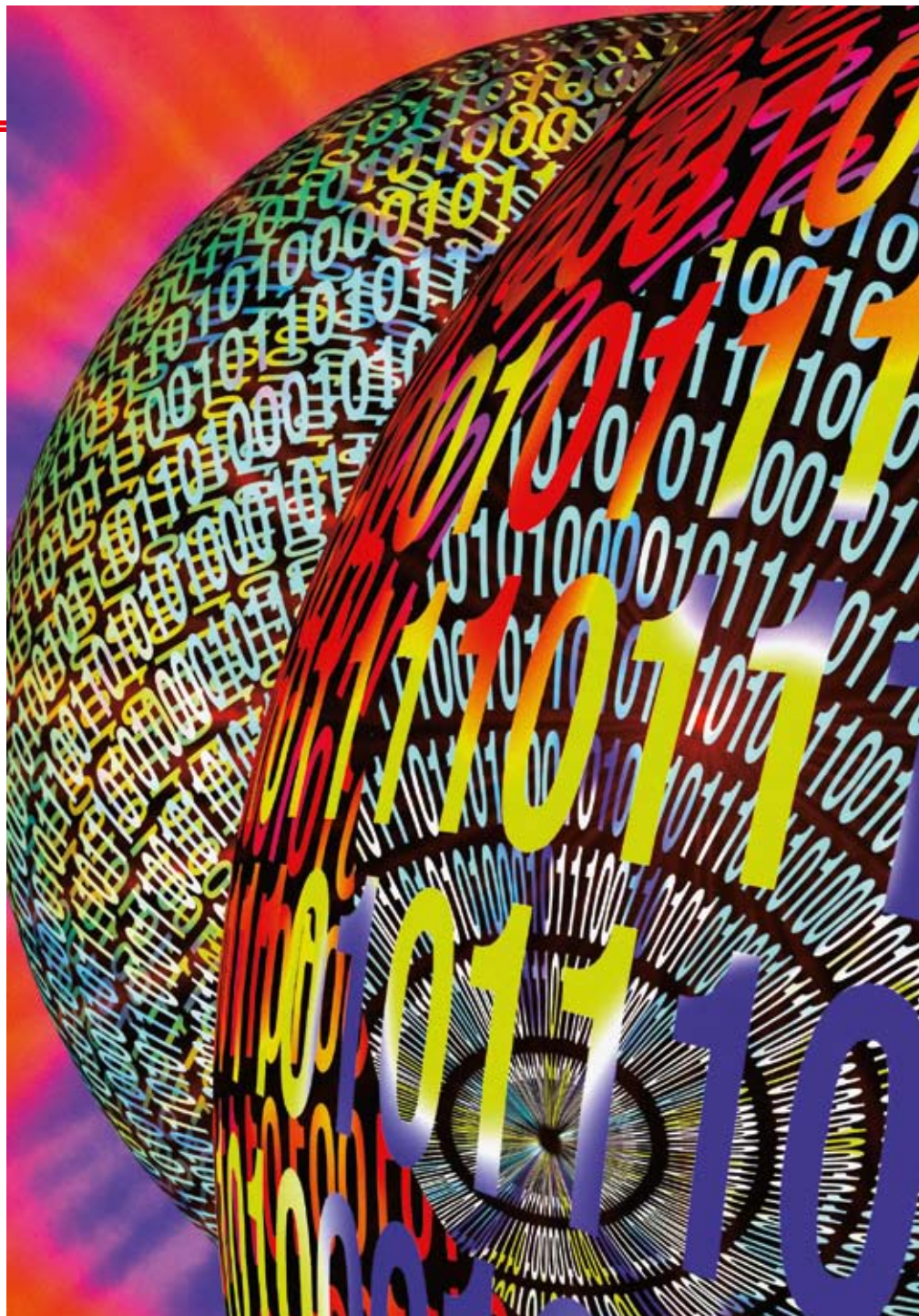
This characteristic, that data and its description are carried together, is largely responsible for XML’s broad appeal and utility across media and other industries. However, the use of text labels to convey meaning has some recognized limitations for both humans and computers.

Human understanding is limited by the users’ ability to interpret the author’s language, naming convention and information model, which can be a hindrance, particularly when attempting to compare the meaning of data across different XML documents from different sources. Computers are largely limited to comparisons of text strings (exact matches) for searching or aggregating data, which can also require that the document tags be in exactly the same order.

Adding enhancements

Several other specifications by the World Wide Web Consortium (W3C), in particular Resource Description Framework (RDF) and Web Ontology Language (OWL), enhance XML’s expressiveness by providing additional standard mechanisms for describing data.

Schema languages (e.g. XML Schema Definition: XSD, RDFS, OWL-S) can enhance the use and understanding of data contained in these documents by enabling data owners to provide meaningful constraints on the document’s content, which can make it easier for both



humans and computers to interpret and validate the data.

In addition, query languages such as SPARQL (Protocol and RDF Query Language) can be used with these technologies with exciting possibilities for text-based information discovery. However, since schema languages also rely on the use of text strings for naming schema elements and attributes, the limitations of text-based processing remain. Their use for interpreting or finding data that can be combined from different sources or transformed for aggregation is still limited.

“ISO standards can provide a way for data owners to improve upon XML’s self-descriptive characteristics.”

About the author

Denise B. Warzel is Associate Director, CORE Program Manager, Center for Biomedical Informatics and Information Technology, National Cancer Institute, National Institutes of Health, USA, and Acting Convenor of JTC 1/SC 32/WG 2.



Useful for humans and computers

Can ISO standards provide a way for data owners to improve upon XML's self-descriptive characteristics? Combining features of ISO/IEC 11179, *Information technology – Metadata registries (MDR)*, and the W3C specification for semantic annotation for Web Service Definition Language (WSDL) and XML schema (SAWSDL), provides the potential for data owners to mitigate or eliminate the limitations of XML's text-based format. Registered metadata can be shared within and between organizations

to promote common understanding of the meaning and representation of data for management, interchange, harmonization, discovery and reuse.

How can this be achieved? The ISO/IEC 11179 series of standards is intended to enable the creation and registration of metadata descriptions for application data and data standards that are precise, explicit and unambiguous, and useful for humans and computers. ISO/IEC 11179-6:2005, *Information technology – Metadata registries (MDR) – Part 6: Registration, specifies the assignment of International Registration Data Identifiers (IRDI)*, which are a composite of a unique organization identifier (the registrar), in accordance with ISO/IEC 6523, *Information technology – Structure for the identification of organizations and organization parts*, and the MDR item identifier and version.

These unique item identifiers serve to encapsulate the meaning of the data they describe and can be incorporated into a uniform resource identifier (URI) or uniform resource locator (URL) to make it easy to access the information via the Web. While SAWSDL is designed as a way to reference ontology concepts (model reference) and mapping schemas (lifting/lowering schema mapping) for relating elements in an XML schema, or Web service description to a URI, it provides a convenient and useful mechanism for referencing ISO/IEC 11179 MDR IRDIs for use by humans and computers. The registered metadata can be accessed and used to interpret, compare and integrate data in different documents from different sources, helping to overcome the limitations of text descriptions.

Powerful mechanisms

In addition to model references, SAWSDL provides a mechanism to reference stored data transformations – or model mappings – which can be derived from ISO/IEC 11179 metadata and used to address the problems of integrating data between schema elements that are semantically identical but have different data representations, (e.g. “1”=Yes, “Y”=Yes), different structures (e.g. “Name” and “firstName, middleInitial,

lastName”), different languages or different modelling paradigms.

Leveraging the ISO/IEC 11179 metadata standard and W3C SAWSDL specification brings together two powerful mechanisms for describing data to help address some of the challenges facing data owners and users of XML for the purposes of data interchange and integration.

Two areas of work to note in joint technical committee ISO/IEC JTC1, subcommittee SC 32, *Data management and interchange*, working group WG 2, *Metadata*, are:

- The revision of ISO/IEC 11179 to expand and integrate the use of concepts and concept systems so as to anchor the meaning of ISO/IEC 11179 metadata descriptions
- Further development of the ISO/IEC 19763 family of standards on meta-model framework for interoperability (MFI).

“Leveraging ISO 11179 and W3C SAWSDL brings together two powerful mechanisms for describing data.”

The ISO/IEC 19763 series provides mechanisms for registering various types of information technology artefacts to support discovery and sharing of data and services on the Web. These include registration of ontologies (Part 3), model mappings (Part 4) and process models (Part 5). A possible Part 7 for a meta-model on service registration has been proposed to facilitate their reuse across organizations.

For more information

The ISO/IEC JTC 1/SC 32/WG 2 Web site, containing links to its current projects, can be found at <http://metadata-standards.org> ■