Research Paper ■

Generating XML Schemas for DICOM Structured Reporting Templates

LUYIN ZHAO, MS, KWOK PUN LEE, PHD, JINGKUN HU, DPS

Abstract In this paper, the authors describe a methodology to transform programmatically structured reporting (SR) templates defined by the Digital Imaging and Communications for Medicine (DICOM) standard into an XML schema representation. Such schemas can be used in the creation and validation of XML-encoded SR documents that use templates. Templates are a means to put additional constraints on an SR document to promote common formats for specific reporting applications or domains. As the use of templates becomes more widespread in the production of SR documents, it is important to ensure validity of such documents. The work described in this paper is an extension of the authors' previous work on XML schema representation for DICOM SR. Therefore, this paper inherits and partially modifies the structure defined in the earlier work.

■ **J Am Med Inform Assoc.** 2005;12:72–83. DOI 10.1197/jamia.M1519.

The Digital Imaging and Communications in Medicine (DICOM) standard[1] aims at creating a public and license-free standard method for the transmission of digital medical images and their associated information. The addition of structured reporting (SR) improves the expressiveness, precision, and comparability of not only images and waveforms, but also documentation of diagnostic observations. More specifically, DICOM SR provides a means to encode structured information to enable unambiguous exchange of clinical information and documentation among systems from different vendors. Structured reporting is defined online,[2] and there is a good book on this topic.[3]

As with any structured data in health care, benefits exist in outcome analysis and point-of-care applications. Patient information and data acquisition are typically performed at a location different from that of interpretation and analysis, necessitating exchange of information. Analysis of historical data is facilitated by an accurate representation of the relevant data as well as the interpretation process. Structured reporting is designed to capture unambiguously structured medical data. Structured reporting in its most general form (referred to throughout as general SR) is very flexible, and the same content can be expressed in different forms and structures, hampering interoperability. Templates are a means to put additional constraints on an SR document to promote common formats for specific reporting applications or domains. For example, DICOM standard Part 16 (DICOM Content Mapping Resource[2,4]) defines reusable SR templates and mammography computer-aided detection SR information object definition (IOD) templates. Additional templates for other specialties are being developed and standardized, such as:

Supplement 26: Ultrasound OB-GYN Procedure Reports
Supplement 66: Catheterization Lab SR SOP (Service-Object Pair) Classes
Supplement 71: Vascular Ultrasound Procedure Reports
Supplement 72: Echocardiography Procedure Reports

Currently, the DICOM standard is maintained as a set of Word documents. The rich structure of SR in general and templates in particular is described using tables, leading to inevitable errors and inconsistencies as more and more templates are developed by different teams of specialists. Such errors and inconsistencies can be typographical errors, inconsistent naming, incorrect structure, and other errors that occur as a consequence of maintaining formal structures using word-processing tables. When the standard is imposed on real-world medical documents, there is no easy way to ensure that the document structure actually conforms to the standard. Therefore, it is highly desirable to capture the structure with a more formal notation that is amenable to machine processing to overcome the above limitations.

In the past few years, we have been focusing our efforts on applying XML technologies to DICOM SR. EXtensible Mark up Language (XML) is a standard format for encoding structured data and has been widely adopted by standards and industries. As the Web now has global acceptance for human information access, XML has become an emerging standard for data exchange. Health care standards organizations including HL7[5] are transitioning to XML. In previous work,[6,7] we developed XML schemas[8] from the DICOM SR specification that can be used to create and validate general SR documents encoded in XML. As an extension of that work and adaptation to the emerging standards, this paper proposes

Affiliation of the authors: Philips Research, Briarcliff Manor, NY.

The authors thank the anonymous reviewers for their useful comments, which improved the presentation of this paper.

Correspondence and reprints: Kwok Pun Lee, PhD, Philips Research, 345 Scarborough Road, Briarcliff Manor, NY 10510; e-mail: <kp.lee@philips.com>.

Received for publication: 12/19/03; accepted for publication: 08/15/04.

Table 1 ■ Comprehensive SR IOD Definition (Table A.35.3-1[2])

IE	Module	Reference	Usage
Patient	Patient	C.7.1.1	M
	Specimen identification	C.7.1.2	C–Required if the observation subject is a specimen
	Clinical trial subject	C.7.1.3	U
Study	General study	C.7.2.1	M
	Patient study	C.7.2.2	U
	Clinical trial study	C.7.2.3	U
Series	SR document series	C.17.1	M
	Clinical trial series	C.7.3.2	U
Equipment	General equipment	C.7.5.1	M
Document	SR document general	C.17.2	M
	SR document content	C.17.3	M
	SOP common	C.12.1	M

C = conditional; IE = information entity; IOD = information object definition; M = mandatory; SOP = service-object pair; SR = structured reporting; U = user option.

new XML schemas for DICOM SR template specifications and an approach to generate such schemas automatically. These schemas can be used to guide the creation and validation of XML-encoded SR reports created with templates and to promote interoperability of such reports.

The organization of this paper is as follows. In the first section, we discuss the relationship between general SR and templates. In the next section, we briefly describe DICOM SR and its XML encoding using approaches described in our previous work. The following section describes the process of generating XML schemas for SR templates. In the fifth section, we discuss some of the issues encountered when adapting our previous approach to include additional template constraints and our solutions. We then provide our conclusions and suggestions for further work.

The reader is assumed to have some knowledge of the DICOM standard, various SR template supplements, and XML technologies, specifically XML schema. The results described in this paper are extensions of previous work.[6,7]

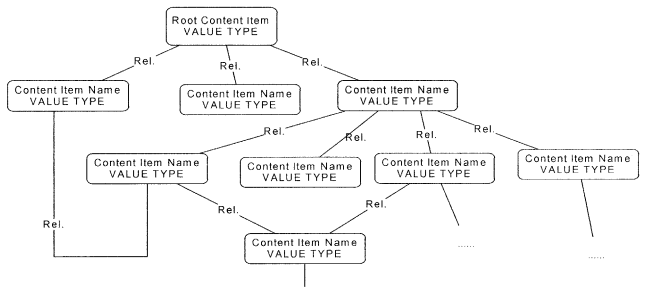

Figure 1. A sample SR content tree.

General SR and Templates

General SR specifies three IODs: basic text, enhanced, and comprehensive. They have the same basic data structures but with more advanced features added to the enhanced and comprehensive IODs. Each IOD contains several normalized or composite object classes called information entities (IEs). An IE consists of a number of modules. For instance, the first column in Table 1 (from the DICOM standard) specifies five IEs within a comprehensive IOD. For each IE, the second column lists modules that belong to the specific IE. Each module is defined in another section of the standard. The last column indicates whether the use of the module is mandatory, a user option, or conditional. The subtle difference between user option and conditional is that the presence of optional elements is dependent on user preference, but the presence of conditional elements is dictated by the presence or absence of other elements. These constraints can be used in combination on a single element.

The document IE is SR specific. Within this IE, the SOP common module is common to all SR IODs. The SR document general module contains general information of an SR document. The SR document content module establishes the information model of observation contexts; it is defined in Section C.17.3 of the DICOM standard but is more easily visualized as a tree, as shown in Figure 1. Such a tree is known as an SR content tree, and each node is known as a content item.

General SR basically specifies the data structures and the allowed types of content items. It does have some semantic constraints such as enumerated values for a few attributes. However, different reporting applications have their own specific observation contexts or vocabularies, which are beyond the general SR constraint scope. This issue is handled through the use of SR templates that apply content constraints on the SR document content module and its components. SR templates are used to put additional constraints on an SR content tree to promote common formats and vocabularies for specific reporting applications (e.g., OB-GYN ultrasound[9]). They are patterns that specify the concept names, relationship with parent, value type, value multiplicity, requirement type, and value set constraint attributes of content items for a particular application. Currently, these templates are represented as tables. For examples, see Tables 2 through 5.[9]

Figure 2 shows part of the top-level template identifier (TID) 5000 OB-GYN Ultrasound Procedure Report template (each template is identified by a unique TID). To make the relationship between general SR and template clear, compare Figures 1 and 2. Figure 1 indicates that besides the hierarchical relationship,* general SR imposes no constraints on, for example, how a content item should be named, what the relationship should be between content items. However, in a template such as the one shown on Figure 2, specific constraints have been imposed: the tree root must be named "OB-GYN Report" and its value type is CONTAINER, the relationship

*Relationships are directional. Thus, the lower left Rel. (Figure 1) indicates a relationship from the lower content item to the upper content item.