

# An Open-Standards Grammar for Outline-Style Radiology Report Templates

Selen Bozkurt · Charles E. Kahn Jr.

Published online: 19 January 2012  
© Society for Imaging Informatics in Medicine 2012

**Abstract** Structured reporting uses consistent ordering of results and standardized terminology to improve the quality and reduce the complexity of radiology reports. We sought to define a generalized approach for radiology reporting that produces flexible outline-style reports, accommodates structured information and named reporting elements, allows reporting terms to be linked to controlled vocabularies, uses existing informatics standards, and allows structured report data to be extracted readily. We applied the Regular Language for XML–Next Generation (RELAX NG) schema language to create templates for 110 reporting templates created as part of the Radiological Society of North America reporting initiative. We evaluated how well this approach addressed the project’s goals. The RELAX NG schema language expressed the cardinality and hierarchical relationships of reporting concepts, and allowed reporting elements to be mapped to terms in controlled medical vocabularies, such as RadLex®, Systematized Nomenclature of Medicine Clinical Terms®, and Logical Observation Identifiers Names and Codes®. The approach provided extensibility and accommodated the addition of new features. Overall, the approach has proven to be useful and will form the basis for a supplement to the Digital Imaging and Communication in Medicine Standard.

**Keywords** Radiology · Structured reporting · Standards · Knowledge representation · Extensible Markup Language (XML) · RELAX NG · Grammar · Regular language

---

S. Bozkurt  
Department of Biostatistics and Medical Informatics, Akdeniz University,  
Antalya, Turkey

C. E. Kahn Jr. (✉)  
Department of Radiology, Medical College of Wisconsin,  
Milwaukee, WI, USA  
e-mail: kahn@mcw.edu

## Abbreviations

CDA	Clinical Document Architecture
DICOM	Digital Imaging and Communication in Medicine
DTD	Document Type Definition
HL7	Health Level Seven
HTML	Hypertext Markup Language
LOINC®	Logical Observation Identifiers Names and Codes
RELAX NG	Regular Language for XML Next Generation
SNOMED CT®	Systematized Nomenclature of Medicine Clinical Terms®
W3C	World Wide Web Consortium
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations

## Background

Radiology reports communicate the observations and interpretations of medical imaging procedures. Reports should include basic elements such as patient identifiers, imaging procedure descriptions, clinical indications, imaging findings, and summary information [1]. In addition to these essential report elements, reports may contain images and multimedia data, record critical results communication, and integrate information for decision support, data mining, quality improvement, and regulatory compliance [2].

The conventional radiology report consists of narrative (“free”) text. To address shortcomings of this approach, such as ambiguous terms, clinically important errors, and the failure to address key clinical questions, the radiology community has begun a transition to structured reporting [3].

Structured reporting can help radiologists synthesize information efficiently, promote the quality and consistency of clinical radiology reports, and enable extraction and analysis of report information. A structured radiology report, ideally, is divided into meaningful, consistently ordered sections and contains standardized language [4]. The Radiological Society of North America (RSNA) has undertaken the creation of a library of “best practices” reporting templates to serve as a foundation for structured reporting [5].

Radiologists and referring physicians have expressed a preference for itemized or structured reports; for example, they prefer reporting templates that contain separate headings for each organ system [6–8]. To facilitate data mining—the extraction of coded information—from radiology reports, it would be ideal for report values to be coded using standardized vocabularies, such as RadLex® [9, 10] or Systematized Nomenclature of Medicine (SNOMED) Clinical Terms® [11]. In addition to free-text (“fill in the blank”) data values, it would be useful to restrict values to specific datatypes, such as integers or dates. Reporting templates could specify optional report sections or items, predefined value sets (“pick lists”), and repeatable blocks of information (for example, to describe multiple gestations on an antenatal ultrasound examination).

A technical framework is needed to store, disseminate, and implement reporting templates in software applications. Open information standards can play an important role in the creation of open, universally comprehensible structured reports [12]. The Extensible Markup Language (XML) offers a standardized and widely used language for information interchange [13], and has been adopted as the interchange format for clinical messaging, such as Health Level 7 (HL7) Clinical Document Architecture (CDA) [14]. We applied and evaluated the use of an XML-based language to define templates for structured reporting of radiology procedures.

## Methods

To define radiology reporting templates, we applied the “Regular Language for XML–Next Generation” (RELAX NG) XML schema language [15–17]. Reporting templates written in RELAX NG served as schemas for XML-based report documents: they specified the elements that could appear in a report, how the report elements were structured, and their allowed values. In this article, we describe the principles for using RELAX NG to define radiology reporting templates, present examples of the application of this approach, and evaluate the approach’s suitability.

The top-level element of a RELAX NG schema, the `<grammar>` tag, was used to specify namespaces for the reporting templates. The `ns` attribute defined the namespace

of the reporting elements, namely, the template library Uniform Resource Locator (URL; <http://www.radreport.org/>); the `xmlns` attribute pointed to the URL for RELAX NG (<http://relaxng.org/ns/structure/1.0>). XML namespaces were specified for the Dublin Core Metadata (<http://www.dublincore.org/>), the RSNA RadLex® ontology (<http://www.radlex.org/>) [9, 10], SNOMED CT® (<http://www.ihstdo.org/>) [11, 18], and LOINC® (<http://www.loinc.org/>) [19, 20]. For example, the `<grammar>` tag’s `xmlns:radlex` attribute defined the `radlex` namespace; the `radlex:id` attribute of an `<element>` tag was then used to specify that reporting element’s corresponding RadLex identifier(s).

We limited our approach to modeling the information that radiologists describe when producing the report of an imaging procedure. We excluded nonclinical information from the reporting templates, such as patient identifiers, facility identifiers, and report authentication, as such data typically are managed by the radiology information system for each examination. Of the high-level radiology report components defined at a consensus conference by RSNA Reporting Committee [4], we focused on five clinical sections of the radiology report: imaging procedure (including clinically relevant information about technique), clinical information, comparison examinations, imaging observations, and the report summary or impression.

In each reporting template, `<element>` tags were used to define the named reporting elements, such as the top-level report container (“Report”), report sections (e.g., “Clinical information”), report subheadings (e.g., “Right kidney”), numerical-value data elements (e.g., “Ejection fraction”), discrete value data elements (e.g., “Hydronephrosis”), and coded data values (e.g., “Severe”). For the “Report” element, the `template` attribute specified the URL of the template’s Web page.

Regular-language expressions were defined using appropriate RELAX NG tags. The `<choice>` tag was used to limit an element’s content to a predefined set of values, each of which can be linked to a RadLex term. Other RELAX NG tags, such as `<optional>`, `<zeroOrMore>`, and `<oneOrMore>` were used to indicate the cardinality of individual reporting elements and groups of elements (Table 1). We used `<text/>` tags to indicate where user-provided narrative (free text) content could appear. World Wide Web Consortium (W3C) standard data types [21], such as `decimal` and `positiveInteger`, were used to specify value fields.

The RELAX NG template for a generic radiology report (Fig. 1a) included sections that described the imaging procedure, clinical information, comparison examinations, imaging findings, and impression. A graphical rendering of the template is shown in Fig. 1b. An example XML report instance based on this template is shown (Fig. 1c). Report instances do not contain patient identifier information. The purpose of the “report instance,” as defined here, is to convey the *clinical*

**Table 1** The use of RELAX NG entities to define radiology reporting templates

Entity	Meaning
attribute	Defines an attribute of an element. In the present approach, only the “Report” element has a defined attribute.
choice	Exactly one of the enclosed sub-entities must appear in the report. Analogous to the “radio button” construct in an HTML data entry form.
data	Indicates that the user can enter data values that conform to a specified type such as integer, decimal, or date. Parameters can further limit the range of allowed values.
define	Defines a “block” of grammar, which can be referenced using the “ref” tag.
element	Defines a report element, such as a section heading (“Findings”), a feature to be described (“Heart_size”), or a coded value (“Normal”).
empty	Specifies an “empty” element, that is, one with no sub-elements.
group	Groups sub-elements.
interleave	Indicates that the enclosed entities may appear in any order; otherwise, entities must appear in the order specified by the template.
oneOrMore	The enclosed entities appear one or more times.
optional	Indicates that the enclosed entities may or may not appear in a report. The analogous construct in an HTML data entry form is the checkbox.
ref	Allows one to include a block of grammar by reference; the block of grammar is defined using the “define” tag.
text	Indicates that the user can enter an arbitrary text string (except angle brackets), or nothing at all.
value	Specifies fixed content of an element or attribute.
zeroOrMore	The enclosed entities appear zero or more times.

*content* that needs to be expressed by the radiologist. A more detailed example is shown schematically for part of a renal stone CT reporting template (Fig. 2).

## Evaluation

The RELAX NG schema language was applied successfully to encode 110 outline-style reporting templates developed by the RSNA Reporting Committee. These templates represented a wide variety of imaging modalities, organ systems, and radiology subspecialties, and included both diagnostic examinations and interventional radiology procedures. The approach was able to correctly express features of the reports, such as selection of predefined options (“pick lists”), optional appearance of report elements or sections, repeated report elements or sections, constrained data types, and the inclusion of unstructured text. Final schemas were reviewed and approved by an editorial committee consisting of radiologists and imaging informatics experts.

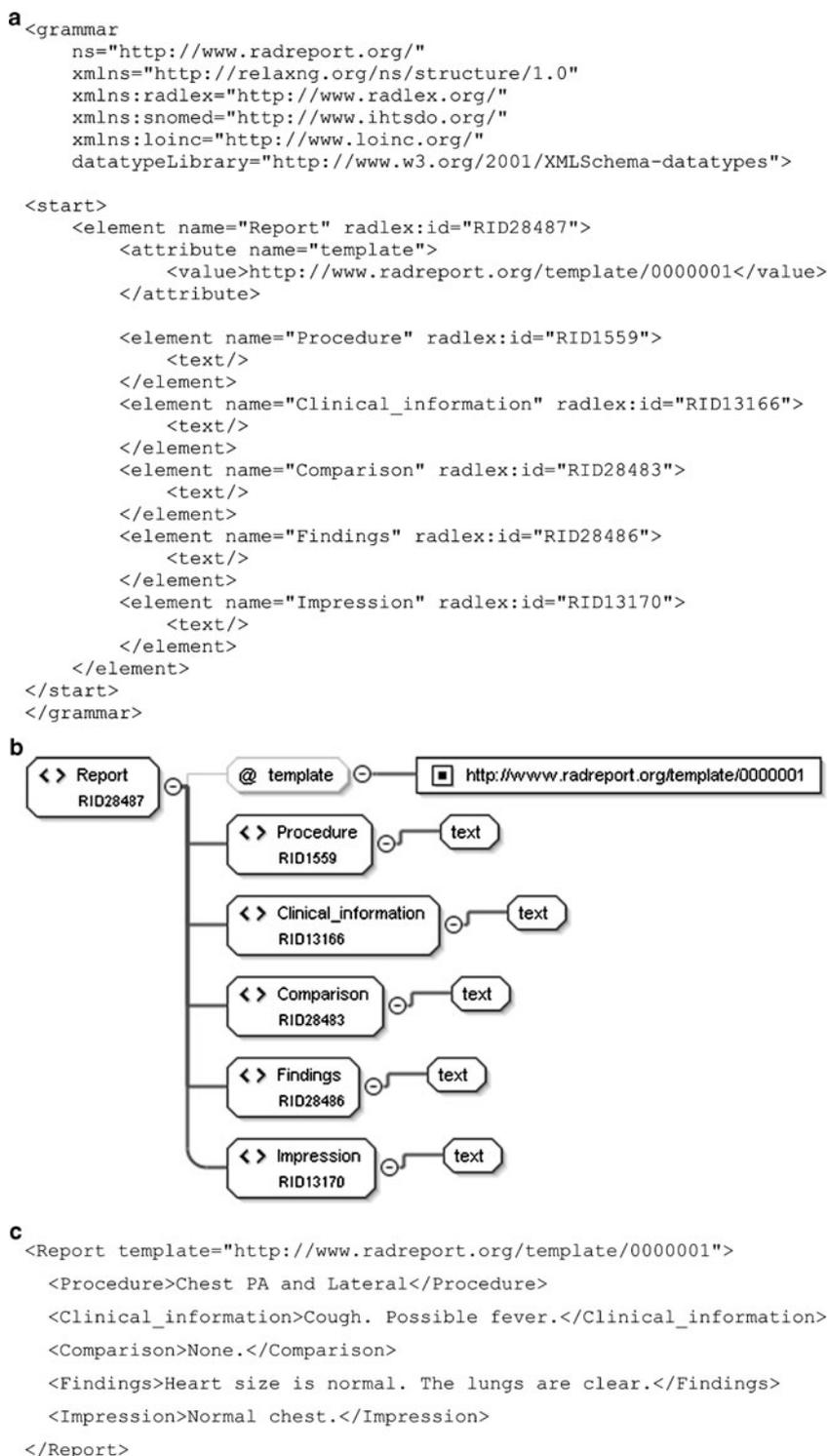
RELAX NG offered distinct advantages over other formalisms for specifying the grammar of XML documents, such as the XML Schema language or the Document Type Definition (DTD) language [22, 23]. From the viewpoint of formal language theory, RELAX NG is more powerful than XML Schema [23]. RELAX NG is relatively simple and easy to understand. RELAX NG is flexible and supports unordered content, data typing, and XML namespaces. RELAX NG also integrates attributes into content models and

supports context-sensitive content models [15]. XML Schema cannot specify such a dependency between the content of an attribute and child elements. RELAX NG is free of most of the limitations of XML Schema concerning non-determinism, and can differentiate elements having the same name and different content models. Moreover, most RELAX NG schemas can be converted algorithmically into XML Schema or DTD documents, but the reverse is not true. RELAX NG is based on tree logic, which allows one to create graphical representations of its schemas [15, 16]. RELAX NG has been recognized as a standard by the International Organization for Standardization (ISO) [24] and by the W3C. Because RELAX NG is an XML-based language, existing tools for manipulating XML documents can be used to validate the syntax of RELAX NG reporting templates, transform them, and extract information from them.

The approach described here has several limitations. Creating and editing report templates in RELAX NG is time-consuming and requires expertise in both XML and radiology. Templates can become large and complex. Although commercial systems are available for editing RELAX NG documents, there is no graphical editing environment at present for building radiology reporting templates. Although the current template format allows reporting elements to be mapped to RadLex and/or SNOMED terms, it requires modification of the present format for reporting templates to comply with the HL7 CDA standard.

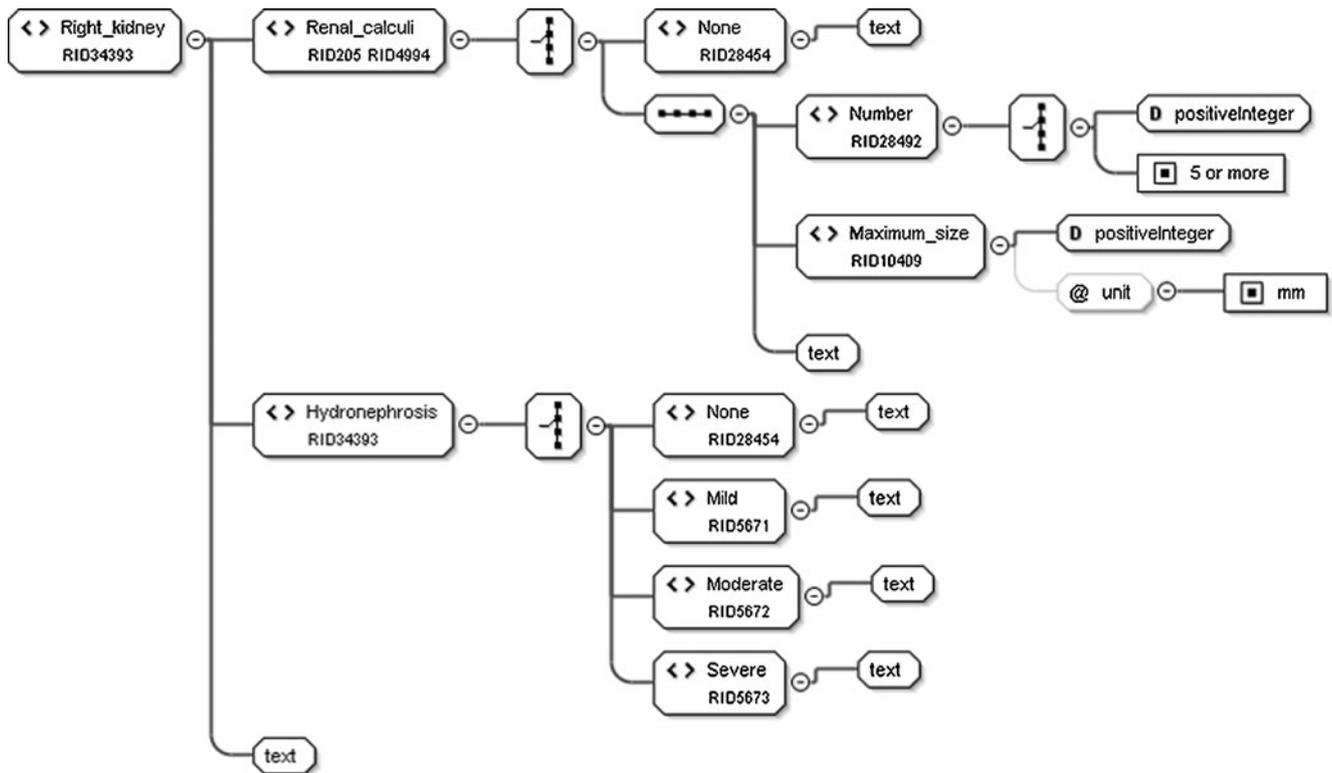
Each reporting template defines a “class” or “family” of valid XML documents that represents clinical reports. A

**Fig. 1** **a** A generic radiology reporting template, expressed in RELAX NG. The `<grammar>` tag contains attributes that specify various XML namespaces: the `ns` attribute defines the namespace of the grammar's elements, which form the elements of the radiology report. **b** A graphical view of the RELAX NG reporting template shown in Fig. 1a. Illustration created using the Oxygen XML Editor (<http://www.oxygenxml.com/>). **c** A valid XML report instance that complies with the reporting template. The “template” attribute of the Report element provides a reference to the reporting template, which can be used to map the reporting elements (e.g., “Clinical information”) to coded RadLex identifiers in the reporting template



radiology report—here, also called a “report instance”—described the results of a specific imaging procedure for a specific patient; it must comply with the grammar of its corresponding template in order to represent a valid report. The XML documents specified by the reporting templates lack requisite administrative information, such as

information about the patient, examination facility, date, imaging procedure, and authenticating provider. One must transform the results information into an HL7 CDA document for the information to be used in a clinical setting. CDA is a subset of XML, which somewhat simplifies the transformation process.



**Fig. 2** Graphical view of part of a reporting template for renal stone CT. For the reporting element “Right kidney,” the template specifies information about renal calculi, hydronephrosis, and additional (optional) text. For renal calculi, one can choose “None,” or enter the number and

maximum size of observed calculi. The size value is required to be a positive integer, with units of millimeters. The “Hydronephrosis” element has a set of predefined values ranging from “None” to “Severe,” all of which are linked to RadLex terms

## Discussion

We used RELAX NG as an XML schema language because it is relatively simple and easy to understand. RELAX NG offers great flexibility for expressing constraints on the content model of report documents, and can be used both as a formal description of the structural constraints of an XML document type and to validate instance documents. The RELAX NG template also specifies the model for extraction of information from report instances. A reporting template can be converted from RELAX NG into the XML Schema language if needed. Because the RELAX NG template is itself an XML document, one can apply the XSLT to transform the templates and extract information from them. For example, a RELAX NG template can be transformed into a Hypertext Markup Language (HTML) data entry form. One of the most important aspects of the reporting templates is the ability to link the reporting elements to controlled terminologies.

RELAX NG is extensible. One can define and apply additional attributes to the templates. For example, one can link the reporting elements with terms from controlled vocabularies. RELAX NG is a widely recognized XML schema language, and has been adopted by both the ISO and the W3C. RELAX NG documents are human-readable and fairly intuitive.

In radiology, a further challenge to structured reporting is the number of templates to be developed. Structured reporting has been embraced by the fields of cardiology [25] and pathology [26], both of which have fewer unique procedures to describe. RELAX NG has been used by the College of American Pathologists for their synoptic reporting initiative (Madden JF, personal communication, Oct. 2010). Not only does radiology have a large number of imaging procedures, but many procedures have numerous indications that necessitate different information to be reported. For example, chest CT examinations performed for indications of “lung nodule follow-up,” “chest pain,” “suspected coronary artery disease,” and “quantitative assessment of emphysema” have differing exam protocols and widely different reporting templates. Reporting templates generally required 30 to 90 min each to encode; additional time was necessary to review them with clinical experts and to map their elements to RadLex terms. The process of encoding templates into RELAX NG is dependent on the skill and speed of the coder, the complexity of the template, and the tools available.

Further elaboration by standards organizations is in progress. The Digital Imaging and Communication in Medicine (DICOM) organization is incorporating the approach described here to define a formal syntax and semantics of radiology

reporting templates. DICOM Supplement 155, now under development, will specify how to transform template-based reports into HL7 CDA format. This work will complement the DICOM Structured Reporting framework. The Integrating the Healthcare Enterprise Radiology group is creating integration profiles for the exchange and application of reporting templates.

## Conclusions

Structured reporting in radiology should promote effective communication and retrieval of imaging procedure results. Accordingly, radiology reporting templates benefit from a flexible, but standardized approach. XML provides the needed structure and extensibility to capture radiology report information, and XML-based methods and applications have the potential to promote development of radiology reporting systems.

RELAX NG provided an XML-based schema language for XML documents and was well suited for encoding radiology report templates. The schema language captured the structure of radiology reports and the values and data types that appeared in the reports. RELAX NG templates allowed report elements to be mapped to standardized lexicons to assure semantic interoperability.

Efforts are underway to integrate this XML-based approach with the HL7 CDA, which also is based on XML. The approach presented here serves as a basis for development of DICOM Supplement 155 (Radiology Structured Reporting Templates). RELAX NG also serves as the basis for “synoptic reporting” efforts by the College of American Pathologists; the common platform holds promise to help integrate radiology–pathology report information.

**Acknowledgments** This work was supported in part by the National Institute for Biomedical Imaging and Bioengineering (NIBIB).

## References

- Kushner DC, Lucey LL: Diagnostic radiology reporting and communication: the ACR guideline. *J Am Coll Radiol* 2:15–21, 2005
- American College of Radiology. ACR Practice Guideline for Communication of Diagnostic Imaging Findings. (revised 2010) <[http://www.acr.org/secondarymainmenucategories/quality\\_safety/guidelines/dx/comm\\_diag\\_rad.aspx](http://www.acr.org/secondarymainmenucategories/quality_safety/guidelines/dx/comm_diag_rad.aspx)> Accessed 15 May 2011
- Dunnick NR, Langlotz CP: The radiology report of the future: a summary of the 2007 Intersociety Conference. *J Am Coll Radiol* 5:626–629, 2008
- Kahn CE Jr, Langlotz CP, Burnside ES, et al: Toward best practices in radiology reporting. *Radiology* 252:852–856, 2009
- Hong Y, Bozkurt S, Kahn CE Jr: Developing a digital library of radiology reporting templates. *AMIA Annu Symp Proc* 1083, 2010
- Naik SS, Hanbidge A, Wilson SR: Radiology reports: examining radiologist and clinician preferences regarding style and content. *AJR Am J Roentgenol* 176:591–598, 2001
- Bosmans JM, Weyler JJ, De Schepper AM, Parizel PM: The radiology report as seen by radiologists and referring clinicians: results of the COVER and ROVER surveys. *Radiology* 259:184–195, 2011
- Schwartz LH, Panicek DM, Berk AR, Li Y, Hricak H: Improving communication of diagnostic radiology findings through structured reporting. *Radiology* 260:174–181, 2011
- Langlotz CP: RadLex: a new method for indexing online educational materials. *Radiographics* 26:1595–1597, 2006
- Rubin DL: Creating and curating a terminology for radiology: ontology modeling and analysis. *J Digit Imaging* 21:355–362, 2008
- Stearns MQ, Price C, Spackman KA, Wang AY: SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp* 662–666, 2001
- Kahn CE Jr: A generalized mark-up language for platform-independent structured reporting. *Methods Inf Med* 36:163–171, 1997
- Bray T, Paoli J, Sperberg-McQueen CM, Maler E, Yergeau F Eds: Extensible Markup Language (XML) 1.0 (5th Edition). W3C Recommendation 26 November 2008. World Wide Web Consortium (W3C). <<http://www.w3.org/TR/REC-xml/>> Accessed 15 May 2011
- Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A: HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc* 13:30–39, 2006
- RELAX NG Home Page. <<http://www.relaxng.org/>>. Accessed 15 May 2011
- Van der Vlist E: RELAX NG. Sebastopol, CA: O'Reilly, 2004. xviii, 486 pp
- Clark J, Murata M: RELAX NG Specification. The Organization for the Advancement of Structured Information Standards [OASIS], 2001. <<http://relaxng.org/spec-20011203.html>> Accessed 15 May 2011
- Cornet R, de Keizer N: Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 8(Suppl 1):S2, 2008
- Huff SM, Rocha RA, McDonald CJ, et al: Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. *J Am Med Inform Assoc* 5:276–292, 1998
- Forrey AW, McDonald CJ, DeMoor G, et al: Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem* 42:81–90, 1996
- Biron PV, Malhotra A Eds: XML Schema Part 2: Datatypes Second Edition. W3C Recommendation. World Wide Web Consortium (W3C). <<http://www.w3.org/TR/xmlschema-2/>> Accessed 15 May 2011
- Murata M, Lee D, Mani M: Taxonomy of XML schema languages using formal language theory. *Extreme Markup Languages* 153–166, 2001
- Lee D, Chu WW: Comparative analysis of six XML schema languages. *ACM SIGMOD Record* 29(3):76–87, 2000
- International Organization for Standardization. ISO/IEC 19757-2:2008. Information technology—Document Schema Definition Language (DSDL)—Part 2: Regular-grammar-based validation—RELAX NG. 2008, 43 pp
- Douglas PS, Hendel RC, Cummings JE, Dent JM, Hodgson JM, Hoffmann U, et al: ACCF/ACR/AHA/ASE/ASNC/HRS/NASCI/RSNA/SAIP/SCAI/SCCT/SCMR 2008 Health Policy Statement on Structured Reporting in Cardiovascular Imaging. *J Am Coll Cardiol* 53:76–90, 2009
- Leslie KO, Rosai J: Standardization of the surgical pathology report: formats, templates, and synoptic reports. *Semin Diagn Pathol* 11:253–257, 1994