



Standard Guide for XML DTD Design, Architecture and Implementation¹

This standard is issued under the fixed designation E2183; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 Recently, there has been widespread use of the Internet by the health care industry to represent information. Much of the success of the Internet can be attributed to the World Wide Web (WWW) and its browsing system, which make it possible to navigate the Internet by pointing and clicking. Web pages, electronic documents located on the Internet, are published to a computer screen by a browser. These pages are developed using HyperText Markup Language (HTML).

1.1.1 The next generation of the WWW and web pages has arrived with the specification of eXtensible Markup Language (XML). Just like HTML, XML can define how pages appear in web browser. Additionally XML provides a context for information. While HTML tells the browser what a page should look like, XML defines what the piece of information actually means. Both HTML and XML are markup languages; that is, they insert markup, or additional information, into text. The markup is known as a tag. For example, HTML marks up text on a page by inserting a **** tag when the text should be bolded and by inserting a **<H1>** tag around the first heading of a web page. In HTML, there is no way to distinguish between a medical record number or an telephone number because they are both tagged **** for bold. XML, on the other hand, marks each item with special tags so that computers can tell the difference. With XML, humans and computers can tell the difference between a **<DIAGNOSIS>** tag and **<SYMPTOM>** or a **<Medical.Record.Number>** from a **<Phone.Number>**.

1.1.2 The names of the tags and the rules for using them are contained in the **DTD** or Document Type Definition (DTD). The DTD describes the structure of the document and defines the names of tags it contains. Additionally, the DTD declares the order in which the tags occur and how often the tags can appear; that is, the DTD defines the hierarchy of the tags. A DTD for a prescription might contain structural elements for the medication prescribed, the dosage, the form, the quantity, and so forth.

1.2 This guide provides a compendium of information for the use of E2183 XML DTDs within health care. This guide describes design considerations, the architecture of the DTDs, and implementing systems using the E2183 DTDs.

1.3 This guide refers to and makes use of recommendations from the World Wide Web consortium, the W3C (<http://www.w3.org>).

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory requirements prior to use.*

2. Referenced Documents

2.1 ASTM Standards:²

E1239 Practice for Description of Reservation/Registration-Admission, Discharge, Transfer (R-ADT) Systems for Electronic Health Record (EHR) Systems²

E1384 Practice for Content and Structure of the Electronic Health Record (EHR)²

E1633 Specification for Coded Values Used in the Electronic Health Record²

E2182 Specification for Clinical XML DTDs in Healthcare

2.2 W3C World Wide Web Consortium Recommendations, (<http://www.w3.org>)

XML 1.0 Recommendation, <http://www.w3.org/TR/REC-xml>

XHTML Basic Recommendation, <http://www.w3.org/TR/xhtml-basic>

XML Linking (Xlink) 1.0 Recommendation, <http://www.w3.org/TR/xlink>

XML Namespaces Recommendation, <http://www.w3.org/TR/REC-xml-names>

XPointer, <http://www.w3.org/TR/xptr>

XSLT, <http://www.w3.org/TR/xslt>

2.3 HL7 Standards, (<http://www.HL7.org>)

Informative Document: Using XML as an Alternative Message Syntax for HL7 Version 2.3.x

Clinical Document Architecture: Level 1

2.4 IETF Specifications, (<http://www.ietf.org>)

¹ This guide is under the jurisdiction of ASTM Committee E31 on Healthcare Informatics and is the direct responsibility of Subcommittee E31.25 on Healthcare Data Management, Security, Confidentiality, and Privacy.

Current edition approved July 17, 2006. Published March 2002. DOI: 10.1520/E2183-02.

² Annual Book of ASTM Standards, Vol 14.01.

RFC 2396: Uniform Resource Identifiers (URI), <http://ietf.org/rfc/rfc2396>

3. Terminology

3.1 Definitions:

3.1.1 *clinical document, n*—defined in terms of the native XML and hypertext architecture. The Document Type Definition (DTD) is the base method of defining the structure of a clinical document. XML schema languages such as the W3C XML Schema Language and the OASIS RELAXNG Schema language provide alternative methods of defining clinical document structures.

3.1.2 *clinical document architecture (CDA), n*—the HL7 CDA is a document markup standard for the structure and semantics of exchanged clinical documents. A clinical document is a documentation of observations and other services with the following characteristics: Persistence, Stewardship, Potential for Authentication, Wholeness, and Human Readability. A clinical document is a defined and complete information object that can exist outside of a message, and can include text, sounds, and other multimedia content.

3.1.3 *document type definition (DTD), n*—formal definition of the elements, structures, and rules for enabling platform independent data access via XML or for marking up a given type of SGML document.

3.1.4 *extensible markup language (XML), n*—a standard from the World Wide Web Consortium (W3C) that provides for tagging of information content within documents, offering a means for representation of content in a format which is both human and machine readable. Through the use of customizable style sheets and schemas, information can be represented in a uniform way, allowing for interchange of both content (data) and format (metadata).

3.1.5 *Health Level 7 (HL7), n*—a standards organization traditionally focused on message-oriented standards for health-care. HL7 messages are the dominant standard for peer-to-peer exchange of clinical, text-based information. HL7 has recently introduced a document-oriented standard, the *Clinical Document Architecture*.

3.1.6 *hypertext markup language (HTML), n*—the language used in creating a web page. Its origin is an implementation of SGML DTD, but whose most recent update is defined in XML (XHTML). It provides tags regarding the way a document should be displayed in the text of an HTML document, which act as commands that a browser interprets when downloading an HTML file.

3.1.7 *namespaces, n*—provide a simple method for qualifying element and attribute names used in XML documents. This is accomplished by associating a particular tag set by associating a prefix with a URI reference. XML Namespaces provide a mechanism for authoring compound documents (documents consisting of elements and attributes from multiple DTDs or schemas) in such a way that will provide global identification without collisions of names that are the same but are used differently.

3.1.8 *parser, n*—a specialized software program that recognizes markup in a document and differentiates the content from the markup. A parser that reads a DTD and checks and reports on markup errors is a validating XML parser. A parser can be

built into an XML editor to prevent incorrect tagging and to check whether a document contains all the required elements.

3.1.9 *schema, n*—defines the elements that can appear within the document and the attributes that can be associated with an element. It also defines the structure of the document: which elements are child elements of others, the sequence in which the child elements can appear, and the number of child elements. It defines whether an element is empty or can include text. The schema can also define default values for attributes. Schema is a W3C term for the next generation of DTDs.

3.1.10 *stylesheet, n*—the XSL Transformations (XSLT) describes a vocabulary recognized by an XSLT processor to transform information from an organization in the source file into a different organization suitable for continued downstream processing. The Extensible Stylesheet Language (XSL) describes a vocabulary recognized by a rendering agent to reify abstract expressions of format into a particular medium of presentation.

3.1.11 *URI, n*—identifies a Web resource according to IETF RFC 2396. Used in practice to form links between, to name, and to locate web documents. Also used to name XML namespaces.

3.1.12 *valid XML document, n*—well-formed with internal or DOCTYPE reference to element definition of tags within the document.

3.1.13 *W3C (World Wide Web Consortium), n*—develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding.

3.1.14 *well-formed XML document, n*—an XML document that conforms to the syntax as specified by the W3C XML 1.0 recommendation.

3.1.15 *XHTML, n*—HTML documents that are well formed and can be processed by an XML parser.

3.1.16 *XLINK, n*—the W3C XML Linking Language defines a method for specifying links between documents and parts of documents using URIs. Links may be simple, representing a 1:1 mapping or extended, representing an $N:N$ mapping.

3.1.17 *XPointer, n*—the XML Pointer Language, defines how individual parts of a document are addressed. XLinks point to a URI (in practice, a URL) that specifies a particular resource. The URL may include an XPointer part that more specifically identifies the desired part or section of the targeted resource or document. XPointer, the XML Pointer Language, defines an addressing scheme for individual parts of an XML document. XLinks point to a URI (in practice, a URL) that specifies a particular resource. The URI may include an XPointer part that more specifically identifies the desired part or element of the targeted resource or document. XPointers use the same XPath syntax familiar to XSL transformations to identify the parts of the document they point to, along with a few additional pieces.

4. Summary of Guide

4.1 *Philosophy of E31 XML DTDs*—The philosophy of ASTM Subcommittee E31.28 was initially based on the premise that a set of document types for documents that

exist. It is not known if health care documents have common and identifiable structures. It is generally accepted that some regularity exists and that this regularity could be used to create standard high level structures. These standard structures could then be easily manifested in XML DTDs. Regularity in document structure is identified by document analysis, review of health care informatics standards and research and complementary technology standards.

4.2 Two workshops were held to design and develop the DTDs. The results of the workshops identified the following:

4.2.1 The sections or categories that exist in clinical documents,

4.2.2 The order of the sections, if the sequence is important, and

4.2.3 The number of times the different sections appear in clinical document types.

4.3 The supporting materials and knowledge of clinical documentation was used to determine the sections, the order of the sections and the number of times the sections appeared in document. This included a review of transcribed reports for common areas, sections and sub-sections, government forms for categories for data entry, and demographic information contained within the header of the Clinical Document Architecture. Using the results of the workshop, this guide, Standard Practice for XML DTDs for Health Care: GEM: A Document Model for Clinical Practice Guidelines, Specification E2182, and other DTDs would be developed.

4.4 The philosophy in DTD design includes the following:

4.4.1 Simplicity of XML document structure,

4.4.2 Agreement on high level structures in clinical documents,

4.4.3 Ability for International Representations in the XML,

4.4.4 Make use of existing standards where possible, and

4.4.5 Maintain readability.

4.5 *Architecture and Design of DTDs*—The ASTM DTDs have been designed to have a common architecture. This architecture defines a document to consist of a document header and a document body. The header may be specific to a particular document type. For instance, the Standard Practice for XML DTDs for Health Care: GEM: A Document Model includes a header that is specific to guidelines in Health Care.

4.6 *Format and Common Features*—Formatting information and common features frequently found in these documents include:

4.6.1 Basic text (including headings, paragraphs, and lists),

4.6.2 Hyperlinks and links to related documents,

4.6.3 Forms,

4.6.4 Tables,

4.6.5 Images, and

4.6.6 Meta information.

4.7 The design provides for tags that represent common items such as paragraphs, images and tables. This is provided by inclusion of the XHTML Basic DTD and the tags, which it contains. XHTML Basic is designed as a common base that

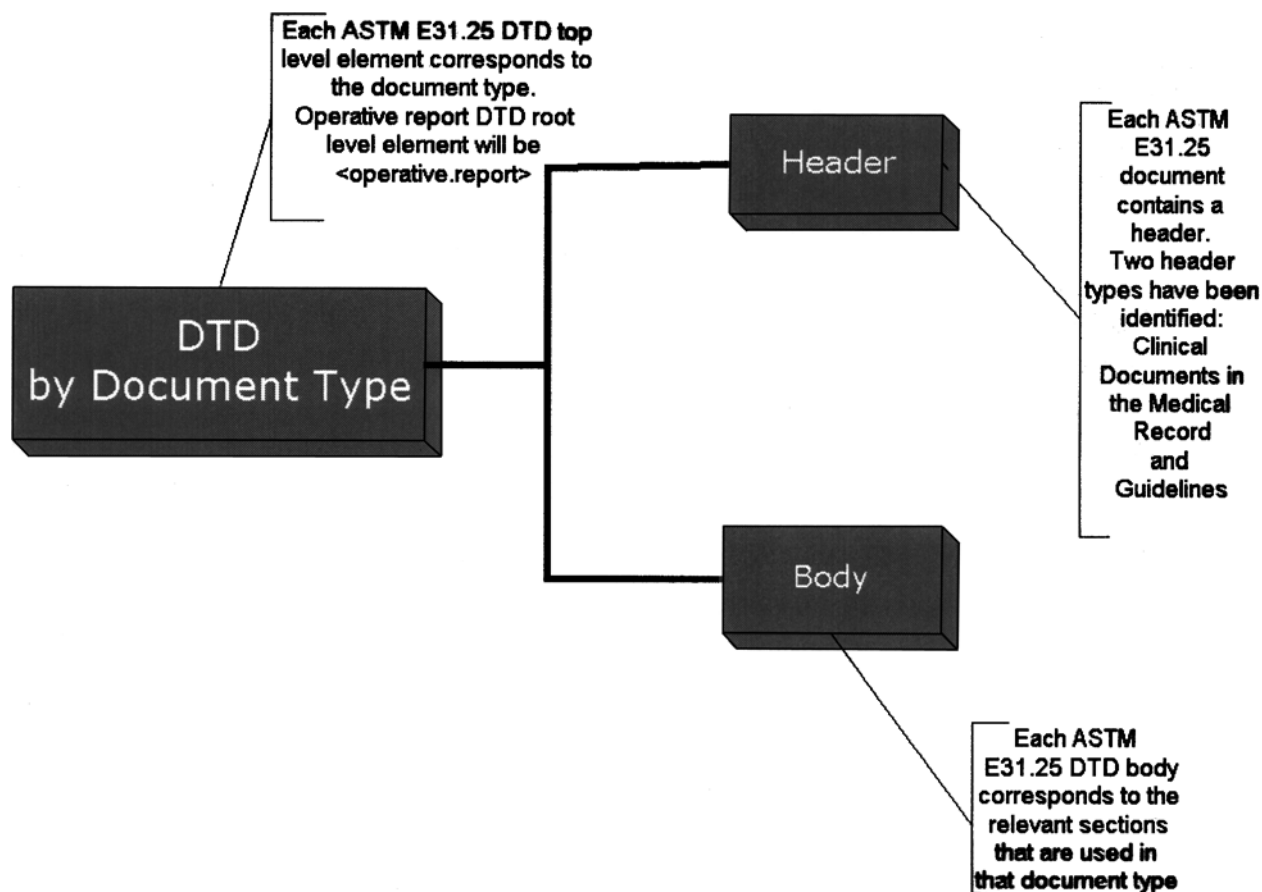


FIG. 1 DTD Architecture

may be extended. The motivation for using XHTML Basic is to incorporate an XHTML document type that can be shared across communities, is rich enough to be used for simple content authoring, and has the ability to be extended as future versions of this guide are developed.

4.8 By allowing the inclusion of XHTML tags, format information may be included within the documents in a standard way. The author of the DTD determines which parts of the XHTML standard are to be used. A subset of XHTML has been developed for use in E2183 DTDs and is included in this guide. Some modules of XHTML were not included in the ASTM XHTML. These modules include forms and web specific tags (such as the <HTML> tag itself). Additionally, some attributes of XHTML, which specify the language of the document, are used to indicate sections that are in different languages (such as French or German). Below is an example operative report, which makes use of the E31.25 architecture and XHTML within the sections.

4.9 *XML Namespaces*—XML Namespaces is a W3C recommendation that allows for a single XML document to contain elements and attributes that are defined for and used by multiple software modules. For instance, the following XML tags listed below are both valid <Citation> elements. The

Citation elements are defined differently in the DTDs. One representation makes use of elements and the other makes use of attributes.

```
<Citation
  Author="Tunstall-Pedoe, H."
  Title="What's in a name?"
  Source="W3C, 1999 Nov; 317:1-4"/>
```

```
<Citation>
  <Author>
    Tunstall-Pedoe, H.
  </Author>
  <Title>
    What's in a name?
  </Title>
  <Source>
    W3C, 1999 Nov;
    317:1-4
  </Source>
</Citation>
```

4.10 Both versions are defined in a DTD and are defined differently. The January 1999 W3C recommendation on namespaces solves the problem of different content models. Namespaces allow for a prefix to be added to the element. This prefix allows for tag names to be unique. Conflicts are avoided for elements and attributes with the same name.

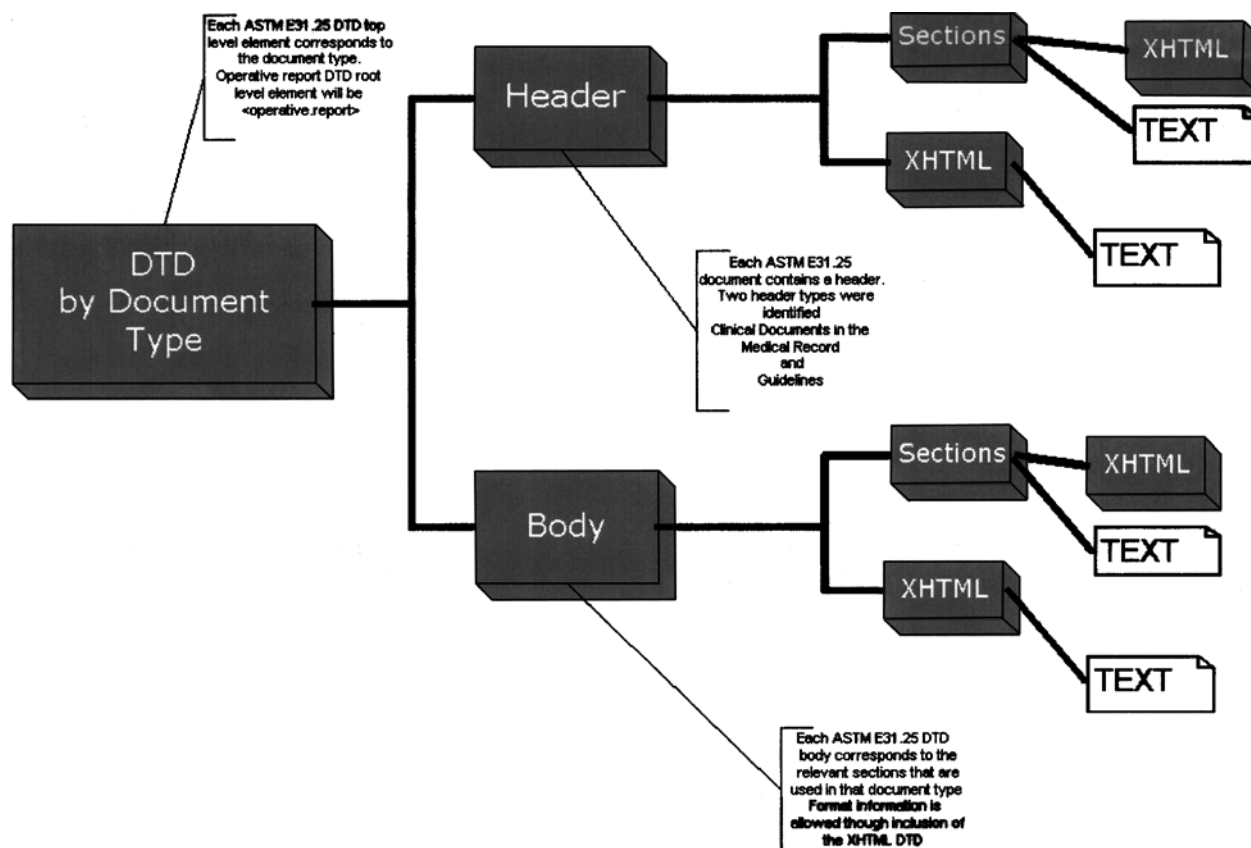


FIG. 2 Architecture and Inclusion of XHTML Basic

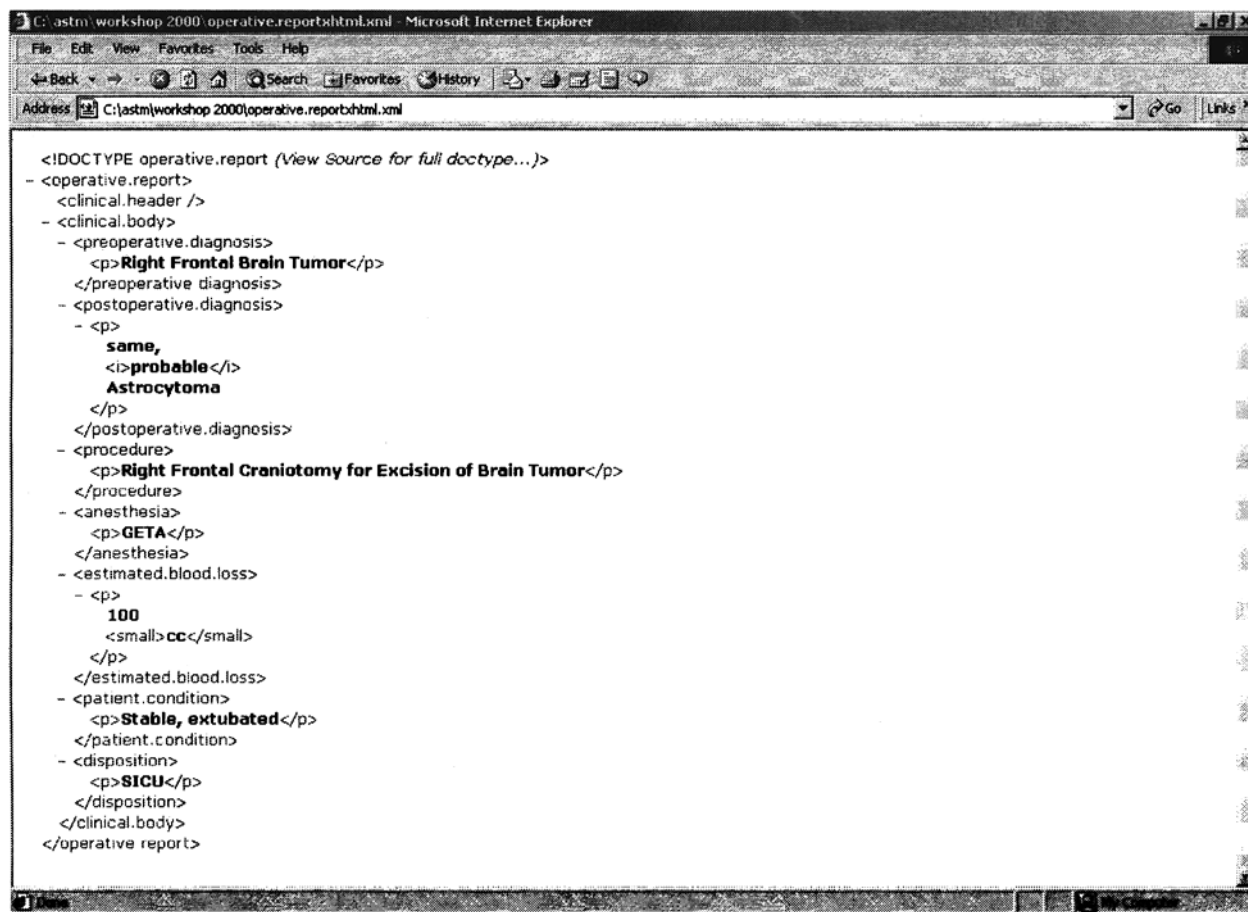


FIG. 3 Use of XHTML in XML Clinical Documents

```
<Citation.List xmlns:Pub1="http://WimeTarnar.com/cit.dtd"
xmlns:Pub2="http://Smiley.com/C.dtd">
<Pub1:Citation Pub1:Author="Tunstall-Pedoe, H."
  Pub1:Title="What's in a name?"
  Pub1:Source="W3C, 1999 Nov; 317:1-4"/>
<Pub2:Citation>
  <Pub2:Author>Tunstall-Pedoe, H.</Pub2:Author>
  <Pub2:Title>What's in a name?</Pub2:Title>
  <Pub2:Source>W3C,1999 Nov;317:1-4</Pub2:Source>
</Pub2:Citation>
</Citation.List>
```

4.11 The E2183 DTDs may make use of namespaces to resolve conflicts with other XML standards and XHTML tags.

4.12 *XLink*—XLink defines the XML Linking Language (XLink), which allows elements to be inserted into XML documents in order to create and describe links between resources. It uses XML syntax to create structures that can describe the simple unidirectional hyperlinks of today's HTML, as well as more sophisticated links. The XLink specification defines the XML Pointer Language (XPointer), the language to be used as a fragment identifier for any URI-reference that locates an XML resource. XPointer supports addressing into the internal structures of XML documents. It allows for traversals of a document tree and choice of its internal parts based on various properties, such as element types, attribute values, character content, and relative position.

4.13 The base E2183 DTD includes the XLink Namespace.

5. Significance and Use

5.1 *Naming Conventions*—E2183 DTDs will use the following Naming Conventions:

5.1.1 Tag names are in English,

5.1.2 Tag names are not shortened; readability is of utmost importance,

5.1.3 Tag names are lowercase except for appropriate abbreviations and acronyms (US, ASTM), and

5.1.4 Tag names with multiple words are separated with a period “.”; for example, patient.name.

5.2 *Architecture*—The following architecture is to be used by ASTM DTDs:

5.2.1 *Root Level*—The root level tag will contain the name of the document type. For instance, the root level element for an operative report will be <operative.report> and for a guideline document <guideline>.

5.2.2 *Header and Body*—The root level element will contain two children elements: a document header and a document body. The document header will contain sections that are relevant for the document. In a guideline document the header may contain information about the guideline developer and meta-information about the guideline such as its title, release date, and whether or not there are companion documents.

5.2.3 *Format*—Format information may be expressed by inclusion of any XHTML tag.



Documents

Document Sharing Hub