



SGML / XML Tutorial

**Ryan Augsburger
Continental DataGraphics
S1000D EPWG Chairman**

**2008 ATA e-Business Forum and
S1000D User Forum
2008-10-21/23
Budapest, Hungary**





Outline



- 10,000 foot (3048 meter) view
- Structure and content
- How SGML/XML works
- DTDs and Schemas
- XML Related Specifications
- Summary



10,000 foot (3048 meter) view



- **SGML = Standard Generalized Markup Language**
 - International Organization for Standardization (ISO) Standard
- **XML = Extensible Markup Language**
 - World Wide Web Consortium (W3C) Recommendation
 - Profile of SGML (initially)
- **Provides exchange of information**
 - Any level of complexity
 - Understood by people, software, storage (database), presentation systems
 - Without regard to manufacturer
 - Authority of international standards



Structure and content



- Ways to look at a document structure
 - Files
 - May represent a title page, a table, a graphic, a chunk of text, etc
 - A fragment of a complete document
 - Logical structure
 - Combination of objects that produce a complete document
 - A hierarchical structure of objects

- Two kinds of content

- Content (data)

- Data about the subject

John Smith
+1-123-456-7890

- Information about the content (markup)

- Information about what the content represents
 - Hierarchical structure
 - Modifiers to the content

name
phone no.

This phone number
belongs to “John Smith”

The phone number
is a Fax number



How SGML/XML works



- All SGML/XML documents are structured data
- Provides **markup** to mark the data.
 - Markup tells what each data item means
 - Markup tells the structure of the data
 - The system determines how to format the data
 - Good practice is to separate content from format
 - An SGML/XML document may indicate a setting of **bold** for a word, but should not indicate details such as Bold Helvetica 16 pt font



How SGML/XML works



- **Elements**

- The building block of SGML/XML
- Each element has a:
 - start-tag
 - content
 - end-tag

<name>John Smith</name>

<phone>+1-123-456-7890</phone>



How SGML/XML works



- **Nested elements create structure**
 - Elements can contain data or other elements

```
<customer>  
  <name>  
    <first>John</first>  
    <last>Smith</last>  
  </name>  
  <phone>+1-123-456-7890</phone>  
</customer>
```



How SGML/XML works



- **Attributes**

- Attributes are name/value pairs that express more information about an element
- May modify the content in some way
- May be used as pointers to other places

- **Attributes**

```
<customer id="123">  
  <name>  
    <first>John</first>  
    <last>Smith</last>  
  </name>  
  <phone type="fax" countryCode="001">  
    123-456-7890</phone>  
  <referredBy customerId="456"/>  
</customer>
```



What is Missing?

What is Next?



DTDs and Schemas



- In order to exchange data, there must be an agreed upon meaning for the elements and attributes
 - The phone “type” to you may mean something different than the phone “type” to me
 - Computing systems cannot discern meaning from names
- DTDs and Schemas define the allowable structure and semantics of an SGML or XML document



DTDs and Schemas



- Document Type Definition (DTD) declares
 - Element names and allowable content
 - Attribute names and allowable values
 - Start-tag and end-tag omission rules (SGML)
 - In short, all possible *document structures* are declared in the DTD
- DTDs can be used for SGML or XML
 - Some minor differences exist between them
 - Written in a DTD language



DTDs and Schemas



- XML Schema
 - Performs the same basic function as DTD
- Schema can be used for XML only
 - Written in XML itself
 - Provides greater controls over DTDs for
 - Definition of allowable content
 - Definition of allowable values
 - Lacks some of the more advanced DTD features



DTDs and Schemas



- Provides the ability to validate an SGML or XML document against the “rules” specified in a DTD or Schema
 - A “well formed” document must adhere to SGML or XML rules
 - A “valid” document must adhere to both SGML or XML rules plus the DTD or Schema rules
 - An SGML or XML document will declare the DTD or Schema it is written to at the top

- The W3C supports a family of specifications related to XML, some of the more popular are:
 - **XSL/XSLT**: Extensible Stylesheet Language
Used to transform XML to another form, usually display
 - **XSL-FO**: XSL Formatting Objects
Vocabulary for specifying formatting semantics
 - **XLink**: XML Linking Language
Create and describe links between resources
 - **XPath**: XML Path Language
Expression language to refer to parts of an XML document
 - **XPointer**: XML Pointer Language
Identifies a resource fragment



Summary



- SGML and XML provide a standard mechanism for exchanging data only
- Specifications such as ATA iSpec 2200 or S1000D provide the specific allowable structures and semantics
 - The specifications define the semantics
 - DTDs and Schemas define and enforce the allowable structures



Questions?



References



- **Electronic Thesis and Dissertation Markup Language (ETD-ML) User's Guide**
Neill A. Kipp
Virginia Polytechnic Institute and State University
March 9, 1997 - Version 1.3
Blacksburg, Virginia
<http://etd.vt.edu/etd-ml/userguid.htm>
- **SoftQuad's Quick Reference Guide to the Essentials of the Standard: The SGML Needed for Reading a DTD and Marked-Up Documents and Discussing Them Reasonably.**
<http://xml.coverpages.org/sqprimerIntro.html>
- **World Wide Web Consortium**
<http://www.w3c.org>