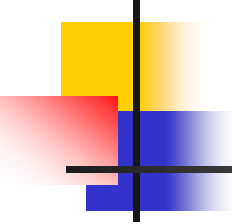


AN INTRODUCTION TO XML...

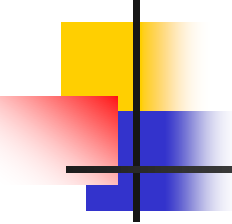
The Web's Universal Data Language

**Terry Garber
South Carolina DOR
Chair, TIGERS**



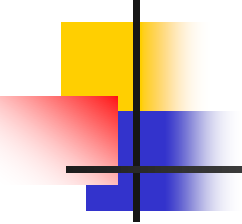
WHAT IS XML?

- **Provisional definition:**
- **Extensible Markup Language (XML) is a way of marking up a “document” or data file to indicate data content.**



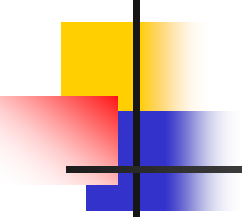
XML FEATURES

- **Selected data is bracketed between a “start tag” <...> and an “end tag” </...>.**
- **Descriptive tags indicate data contents, for example:**
<TaxpayerName>John Smith</TaxpayerName>
- **Computer program can interpret data and reformat it for additional processing**
- **Data can be stored in a database**



NOT A FLAT FILE

- **Simple elements with or without attributes**
- **Complex “types” containing subordinate elements with or without attributes**
- **Elements and complex types can occur multiple times if needed**
- **Can “nest” elements and complex types to create variable hierarchical structures**
- **XPath map through layers of hierarchy**



XML EXAMPLE

<Taxpayer>

<TaxpayerName> John Smith </TaxpayerName>

<TaxpayerSSN> 987654321 </TaxpayerSSN>

<Dependent>

<DependentName> Johnny Smith </DependentName>

<DependentSSN> 123456789 </DependentSSN>

</Dependent>

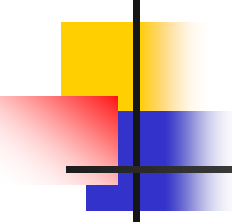
<Dependent>

<DependentName> Susie Smith </DependentName>

<DependentSSN> 246813579 </DependentSSN>

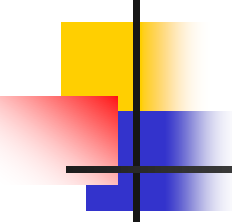
</Dependent>

</Taxpayer>



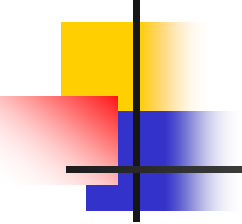
WHERE DID XML COME FROM?

- Like HTML, it is derived from Standard Generalized Markup Language (ISO 8879)
- XML itself is NOT a standard, but as close as you can get in the web world
- XML is a recommendation of the World Wide Web consortium (W3C)
- “Extensible” means you make up the tags!



WELL-FORMED XML

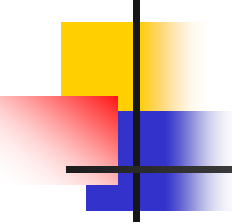
- **Can be read and processed by an XML parser, which can convert the data to another format as needed**
- **Syntax is correct**
- **All the tags match up, and do not intersect or overlap**
- **Doesn't validate document content**



WHAT ABOUT ADDING BUSINESS RULES?

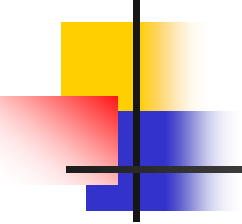
For example:

- **Each taxpayer must have exactly one name and one Social Security Number.**
- **Each taxpayer may have any number of dependents, but doesn't have to have any.**
- **Each dependent must have exactly one name and one Social Security Number.**



BUSINESS RULES IN XML

- **Schema (.xsd)**
 - **Defines an XML document**
 - **Comprehensive data definition and edit capabilities**
 - **Defines nesting structures**
 - **Coded using an XML-formatted data definition language**
 - **Schemas themselves must be well-formed and valid**



SCHEMA EXAMPLE

```
<element name="Taypayer" type="TaxpayerType"/>
```

```
<complexType name="TaxpayerType">
```

```
  <element name="TaxpayerName"/>
```

```
  <element name="TaxpayerSSN" type="SSNType"/>
```

```
  <element name="Dependent" type="DependentType"  
    minOccurs="0" maxOccurs="unbounded"/>
```

```
</complexType>
```

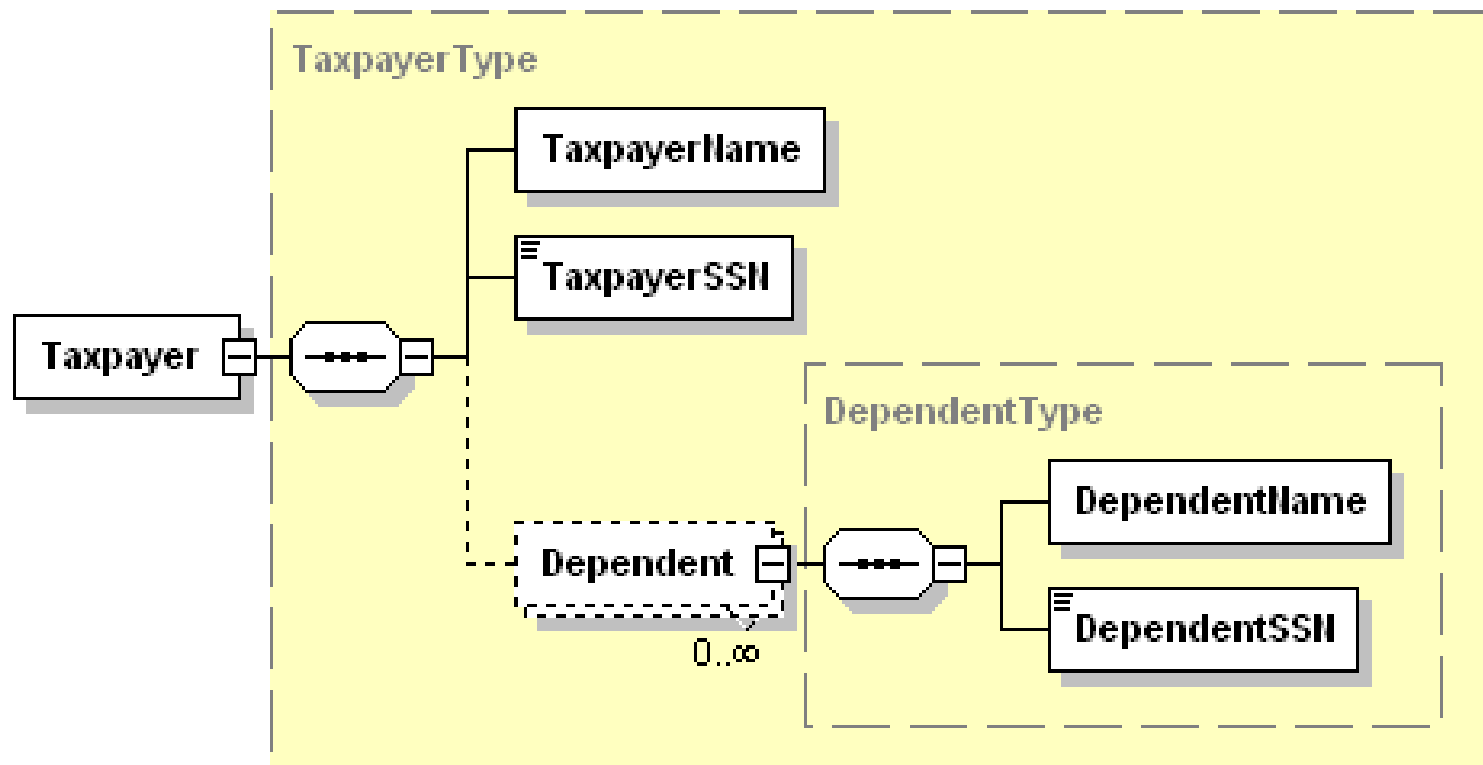
```
<complexType name="DependentType">
```

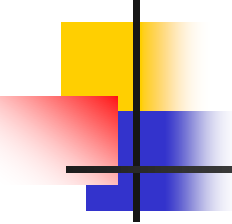
```
  <element name="DependentName"/>
```

```
  <element name="DependentSSN" type="SSNType"/>
```

```
</complexType>
```

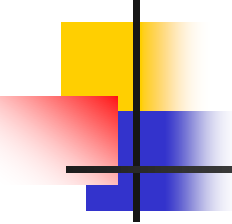
SCHEMA DIAGRAM





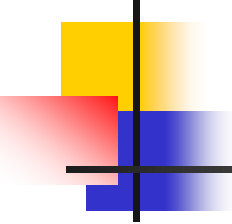
SCHEMA PARAMETERS

- **Data types such as string, integer, non-negative integer**
- **minOccurs and maxOccurs, maxLength, totalDigits**
- **Restrictions on length or value**
- **Patterns, such as [1-9]{9} for SSN**
- **Enumerated values for elements**
- **Cannot make the value of one element dependent on the value of another element**



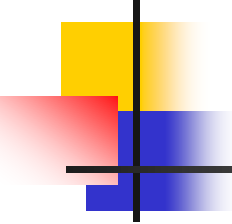
VALIDATING XML

- **XML document specifies the schema to which it should conform**
- **Parser checks XML document both for syntax and for conformance to schema**
- **XML document is “valid” if it conforms to the business rules specified by the schema**



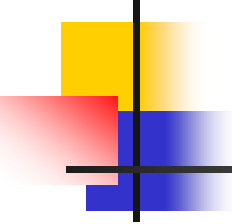
REFINED DEFINITION

- **Extensible Markup Language (XML) is a method of formatting data content according to defined business rules and structures.**



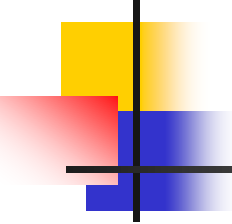
ADVANTAGES OF SCHEMA VALIDATION

- **Parser edits data at point of entry**
- **Only clean data makes it to the processing system**
- **Software developers can test their own data using the schema, before testing with the tax and revenue agency**
- **Standard schemas can be published to provide consistency across multi-state and fed/state programs**



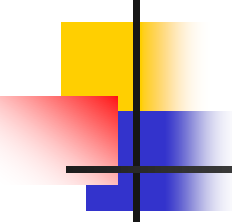
HOW IS THE SCHEMA SHARED BETWEEN PARTIES?

- **The schema may be transmitted along with the XML document**
- **More generally, the XML document specifies a “URI” or location for the schema, which is generally a Website**
- **The receiving party retrieves the schema using the URI and uses it for validation**



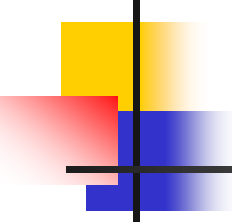
ADVANTAGES OF XML OVER PROPRIETARY FORMATS

- **Human readable using current browser**
- **Tools for developing schemas, and parsers for validation, are comparatively inexpensive**
- **Business rules can be shared and validated via a common website**
- **Only need to agree on tags for specific applications**



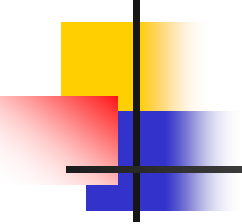
DISPLAYING XML

- **XSL - Extensible Stylesheet Language**
- **All the power of HTML – for example, can duplicate a tax form**
- **Can “attach” a style sheet to an XML document**
- **Browser can interpret XSL to display the XML document**



WHERE IS XML BEING USED TODAY?

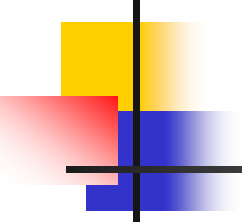
- **Web applications that transfer data between displays and databases**
- **Online catalogs, and Web purchasing applications**
- **Foundation of Services Oriented Architecture using web services to communicate application to application**



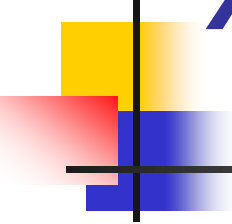
EXAMPLES OF XML USE IN TAX FILING

- **TaXML - Microsoft sponsored Personal Income Tax electronic filing in the UK**
- **IRS 940/941 e-file**
- **IRS Modernized e-file, including Fed/State 1120 and Fed/State 1065 – Fed/State 1040 will be migrated to XML in 2009**
- **Streamlined Sales Tax**

WHY XML FOR THESE PROGRAMS?

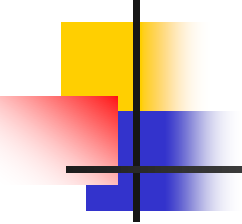


- **Provides cost-effective tools for building Web-enabled applications**
- **Provide simple application-to-application interfaces between front-end Web applications and legacy systems**
- **Provide a common format for data interchange between two parties**
- **Platform independent**
- **Single XML-based eFile architecture across multiple tax types**



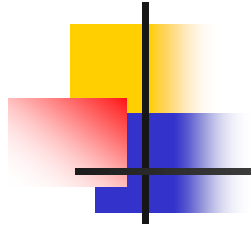
XML IS NOT PERFECT...

- **XML isn't free – States must provide infrastructure**
 - **Authoring tools**
 - **Parsers**
 - **XML processors**
- **Not transmission efficient**
 - **Compression helps**
- **States must build interfaces from the XML transmission to their legacy systems**



XML STANDARDS DEVELOPMENT

- **Need to agree on common tag names – for example <AdjustedGrossIncome> rather than <AGI> to encourage uniformity**
- **Need to agree on common schema structures, such as Header, Financial Transaction, and Binary Attachments**
- **Need to allow flexibility for tax forms, which vary from state to state**
- **This is the work that TIGERS does, in creating XML standards for e-file**



QUESTIONS?