XML Schemas

A CURA DI BELUSSI ALBERTO

(ESTRATTI DA MATERIALE DISPONIBILE SUL SITO W3C)

HTTP://WWW.W3.ORG/TR/XMLSCHEMA-0/ (PRIMER)
HTTP://WWW.W3.ORG/TR/XMLSCHEMA-1/ (STRUCTURES)
HTTP://WWW.W3.ORG/TR/XMLSCHEMA-2/ (DATATYPES)

30 Second Intro

- On the next 3 slides is a very quick, high-level introduction to XML Schemas. The purpose is to give you the "big picture" before we jump into all the details of creating XML Schemas.

What is XML Schemas?

Answer

An XML vocabulary for expressing your data's business rules

Example


```
<location>
        <latitude>32.904237</latitude>
        <longitude>73.620290</longitude>
        <uncertainty units="meters">2</uncertainty>
</location>
```

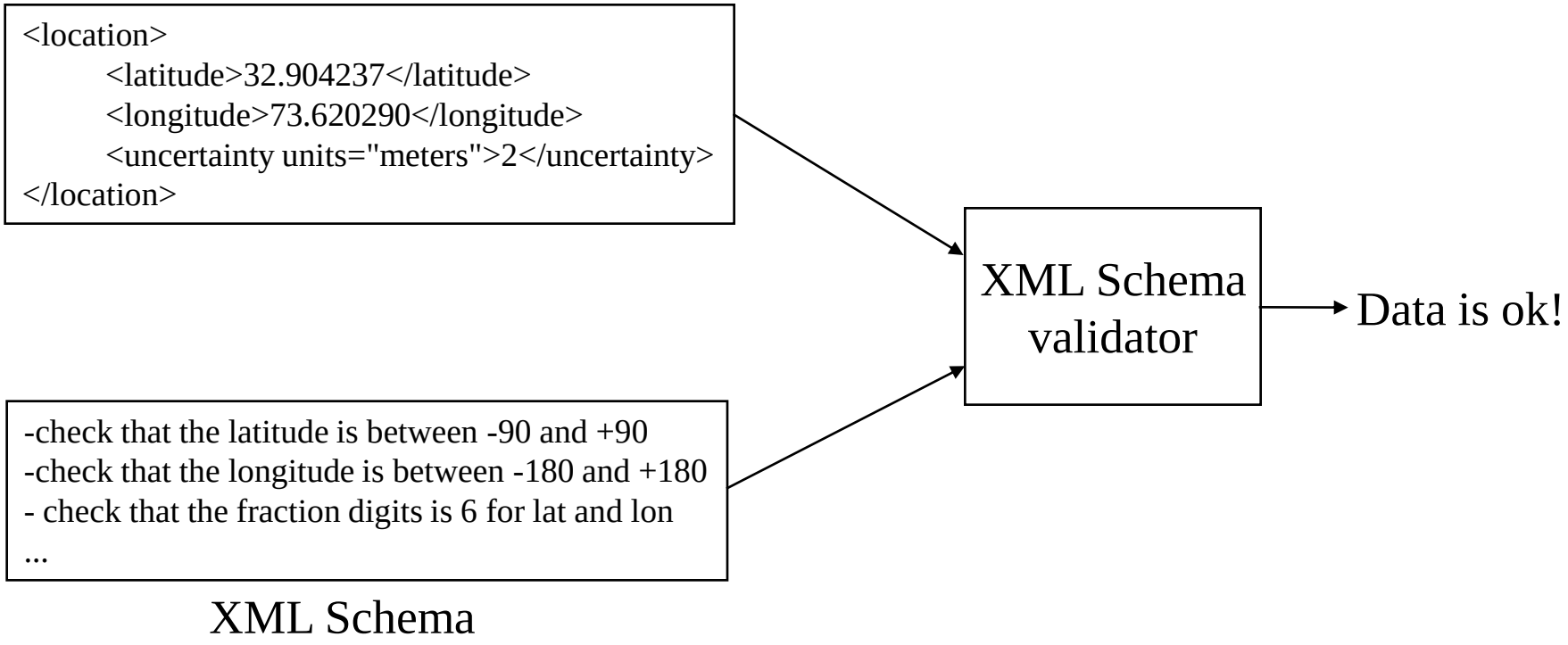
Is this data valid?

To be valid, it must meet these **constraints (data business rules)**:

1. The location must be comprised of a latitude, followed by a longitude, followed by an indication of the uncertainty of the lat/lon measurements.
2. The latitude must be a decimal with a value between -90 to +90
3. The longitude must be a decimal with a value between -180 to +180
4. For both latitude and longitude the number of digits to the right of the decimal point must be exactly six digits.
5. The value of uncertainty must be a non-negative integer
6. The uncertainty units must be either meters or feet.

We can express all these data constraints using XML Schemas

Validating your data



Purpose of XML Schemas (and DTDs)

- Specify:
  - the *structure* of instance documents
    - "this element contains these elements, which contains these other elements, etc"
  - the *datatype* of each element/attribute
    - "this element shall hold an integer with the range 0 to 12,000" (DTDs don't do too well with specifying datatypes like this)

Motivation for XML Schemas

People are NOT satisfied with DTDs

- It's a different syntax
  - You write your XML (instance) document using one syntax and the DTD using another syntax --> bad, inconsistent
- Limited datatype capability
  - DTDs support a very limited capability for specifying datatypes. You can't, for example, express "I want the <elevation> element to hold an integer with a range of 0 to 12,000"
  - Desire a set of datatypes compatible with those found in databases
    - DTD supports 10 datatypes; XML Schemas supports 44+ datatypes

Highlights of XML Schemas

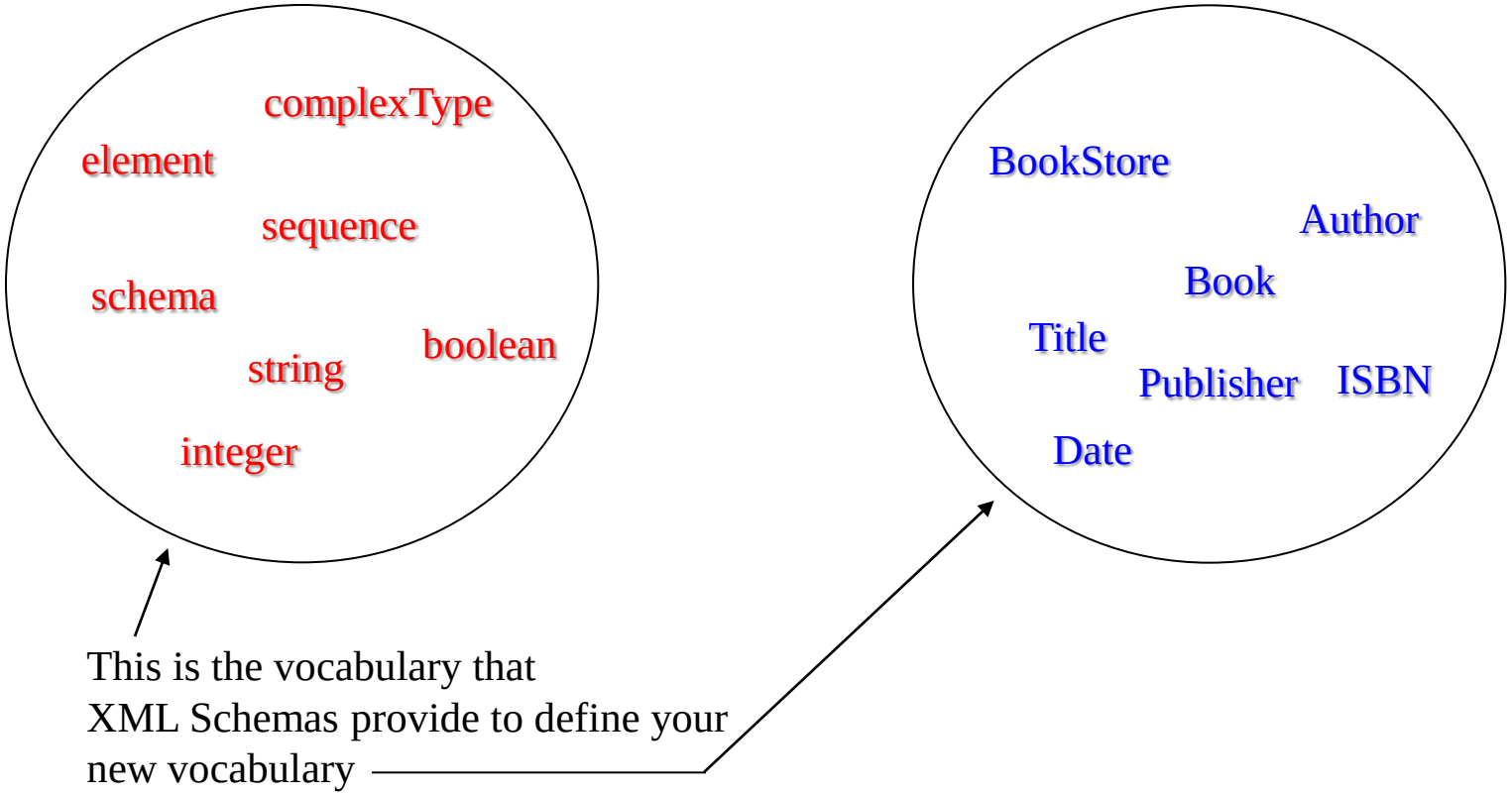
- XML Schemas are a “tremendous” advancement over DTDs:
  - Enhanced datatypes
    - 44+ versus 10
    - Can create your own datatypes
      - Example: "This is a new type based on the string type and elements of this type must follow this pattern: ddd-dddd, where 'd' represents a digit".
  - Written in the same syntax as instance documents
    - less syntax to remember
  - Object-oriented'ish
    - Can extend or restrict a type (derive new type definitions on the basis of old ones)
  - Can express sets, i.e., can define the child elements to occur in any order
  - Can specify element content as being unique (keys on content) and uniqueness within a region
  - Can define elements with nil content
  - Can define substitutable elements - e.g., the "Book" element is substitutable for the "Publication" element.

Let's Get Started!

- Specify the syntax of the BookStore example (next page) using XML Schema
  - for this first example we will make a straight, one-to-one representation, i.e., Title, Author, Date, ISBN, and Publisher will hold strings
  - We will gradually modify the XML Schema to use stronger types

http://www.w3.org/2001/XMLSchema

http://www.books.org (*targetNamespace*)



One difference between XML Schemas and DTDs is that the XML Schema vocabulary is associated with a name (namespace). Likewise, the new vocabulary that you define must be associated with a name (namespace). *With DTDs neither set of vocabulary is associated with a name (namespace) [because DTDs pre-dated namespaces]*.

BookStore.xsd

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

(explanations on succeeding pages)

xsd = Xml-Schema Definition

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

All XML Schemas have "schema" as the root element.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

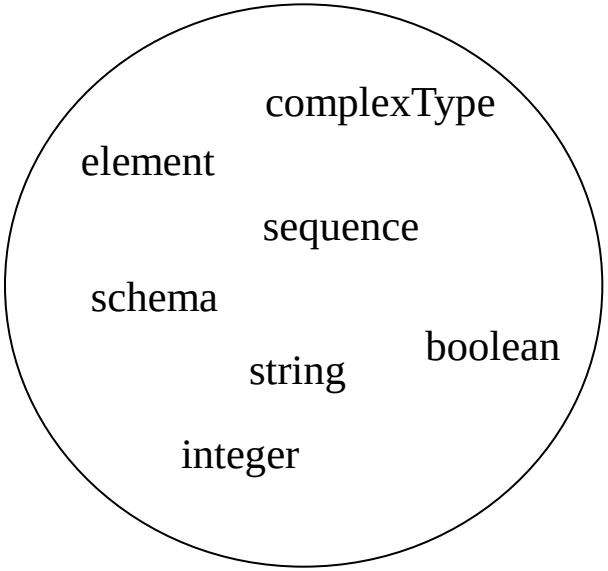
The elements and datatypes that are used to construct schemas

- schema
- element
- complexType
- sequence
- string

come from the http://…/XMLSchema namespace

XMLSchema Namespace

http://www.w3.org/2001/XMLSchema



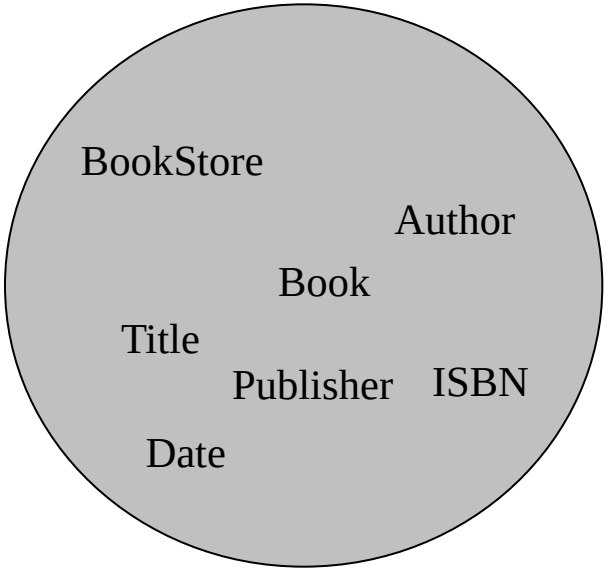
```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

Indicates that the elements defined by this schema
- BookStore
- Book
- Title
- Author
- Date
- ISBN
- Publisher

are to go in the http://…/books.org namespace

Book Namespace (targetNamespace)

http://www.books.org (targetNamespace)



```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

The default namespace is http://www.books.org which is the targetNamespace!

This is referencing a Book element declaration. The Book in what namespace? Since there is no namespace qualifier it is referencing the Book element in the default namespace, which is the targetNamespace! Thus, this is a reference to the Book element declaration in this schema.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Book" minOccurs="1" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Book">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
        <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Title" type="xsd:string"/>
  <xsd:element name="Author" type="xsd:string"/>
  <xsd:element name="Date" type="xsd:string"/>
  <xsd:element name="ISBN" type="xsd:string"/>
  <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

This is a directive to any instance documents which conform to this schema: Any elements used by the instance document which were declared in this schema must be namespace qualified.

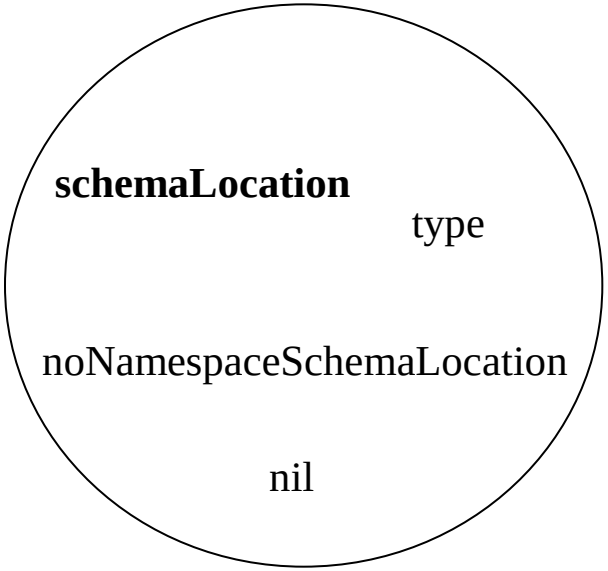
Referencing a schema in an **XML instance** document


```
<?xml version="1.0"?>
<BookStore xmlns ="http://www.books.org"  (1)
           xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  (3)
           xsi:schemaLocation="http://www.books.org
                               BookStore.xsd">  (2)
    <Book>
        <Title>My Life and Times</Title>
        <Author>Paul McCartney</Author>
        <Date>July, 1998</Date>
        <ISBN>94303-12021-43892</ISBN>
        <Publisher>McMillin Publishing</Publisher>
    </Book>
    ...
</BookStore>
```

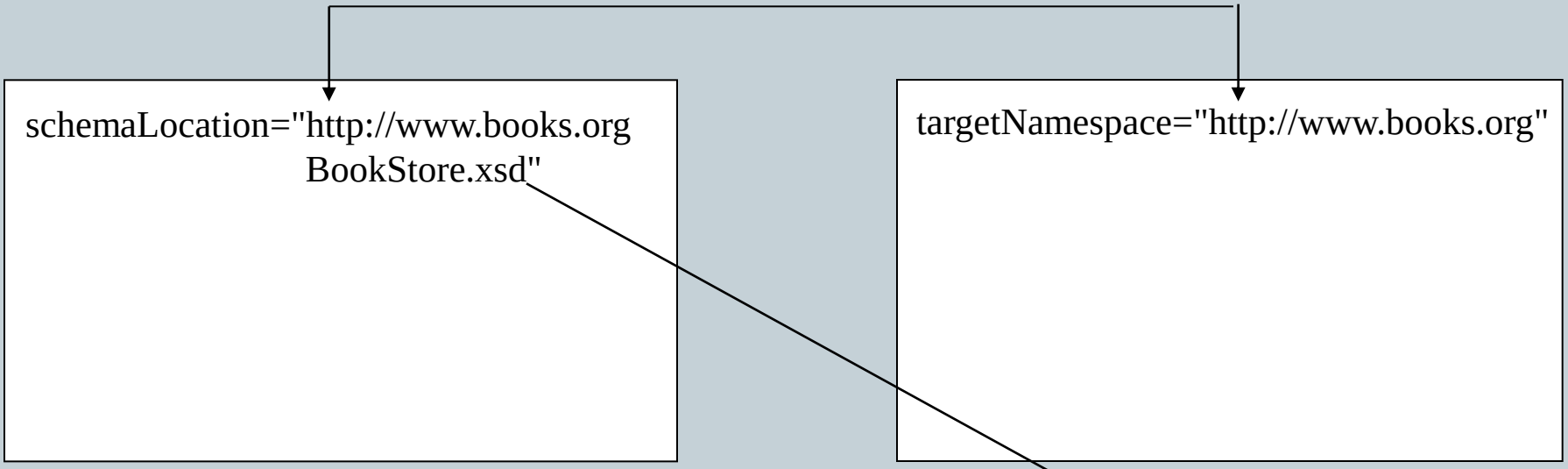
1. First, using a **default namespace declaration**, tell the schema-validator that all of the elements used in this instance document come from the *http://www.books.org* namespace.
2. Second, with schemaLocation tell the schema-validator that the *http://www.books.org* namespace is defined by BookStore.xsd (i.e., **schemaLocation contains a pair of values**).
3. Third, tell the schema-validator that the **schemaLocation attribute** we are using is the one in the XMLSchema-instance namespace.

XMLSchema-instance Namespace

http://www.w3.org/2001/XMLSchema-instance



Referencing a schema in an XML instance document



BookStore.xml

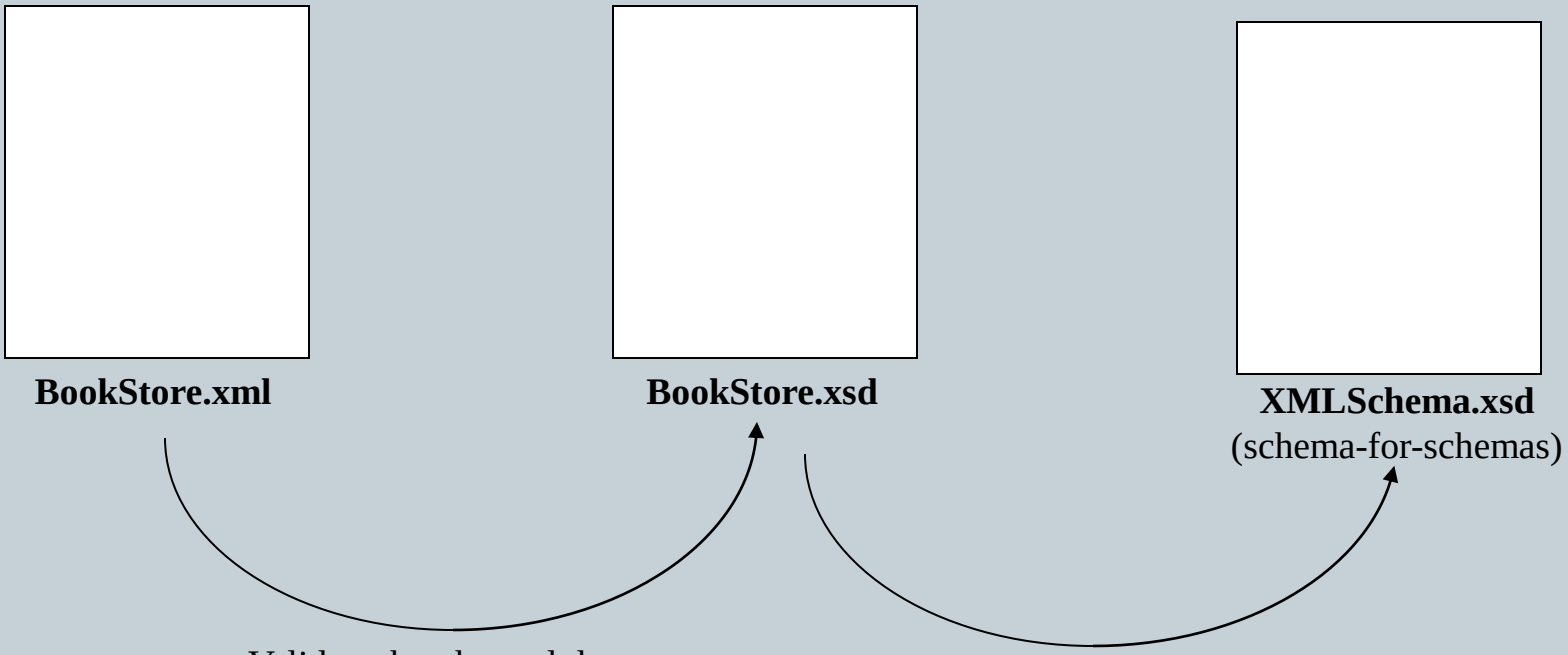
- *uses* elements from namespace http://www.books.org

BookStore.xsd

- *defines* elements in namespace http://www.books.org

A schema *defines* a new vocabulary. Instance documents *use* that new vocabulary.

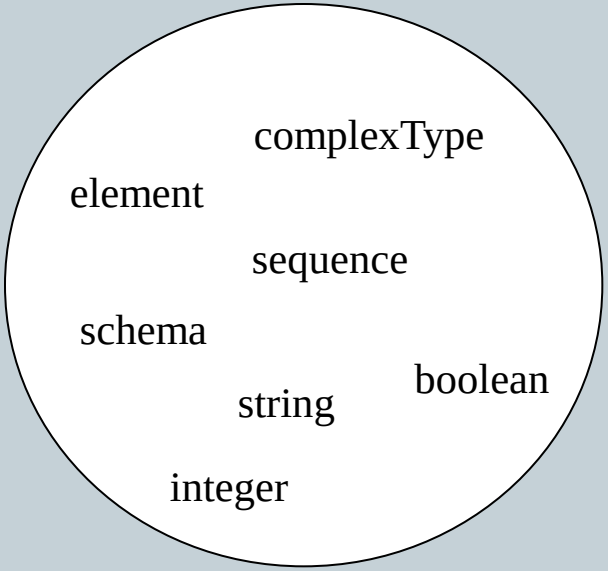
Note multiple levels of checking



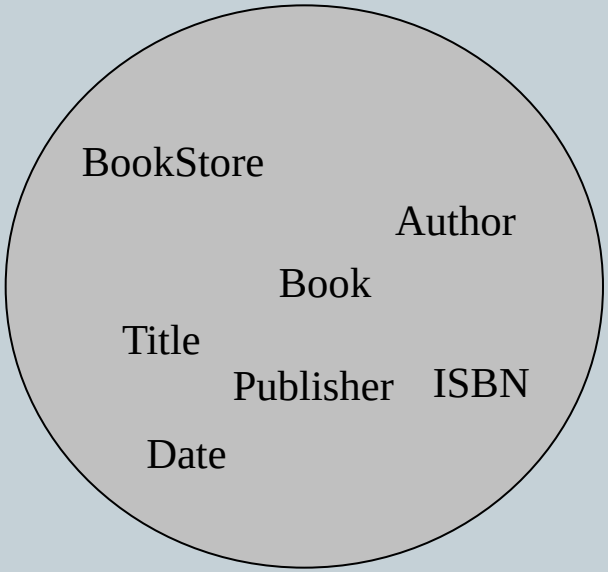
Qualify XMLSchema, Default targetNamespace

- In the first example, we explicitly qualified all elements from the XML Schema namespace. The targetNamespace was the default namespace.

http://www.w3.org/2001/XMLSchema



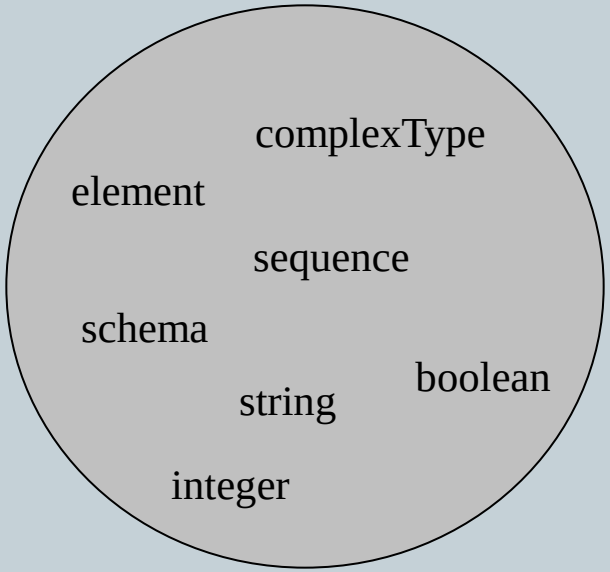
http://www.books.org (targetNamespace)



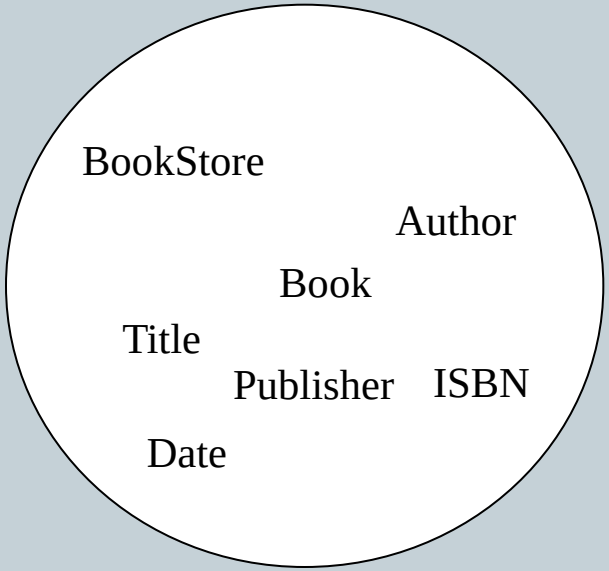
Default XMLSchema, Qualify targetNamespace

- Alternatively (equivalently), we can design our schema so that XMLSchema is the default namespace.

http://www.w3.org/2001/XMLSchema



http://www.books.org (targetNamespace)



```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns:bk="http://www.books.org"
            elementFormDefault="qualified">
   <element name="BookStore">
      <complexType>
         <sequence>
            <element ref="bk:Book" maxOccurs="unbounded"/>
         </sequence>
      </complexType>
   </element>
   <element name="Book">
      <complexType>
         <sequence>
            <element ref="bk:Title"/>
            <element ref="bk:Author"/>
            <element ref="bk:Date"/>
            <element ref="bk:ISBN"/>
            <element ref="bk:Publisher"/>
         </sequence>
      </complexType>
   </element>
   <element name="Title" type="string"/>
   <element name="Author" type="string"/>
   <element name="Date" type="string"/>
   <element name="ISBN" type="string"/>
   <element name="Publisher" type="string"/>
</schema>
```

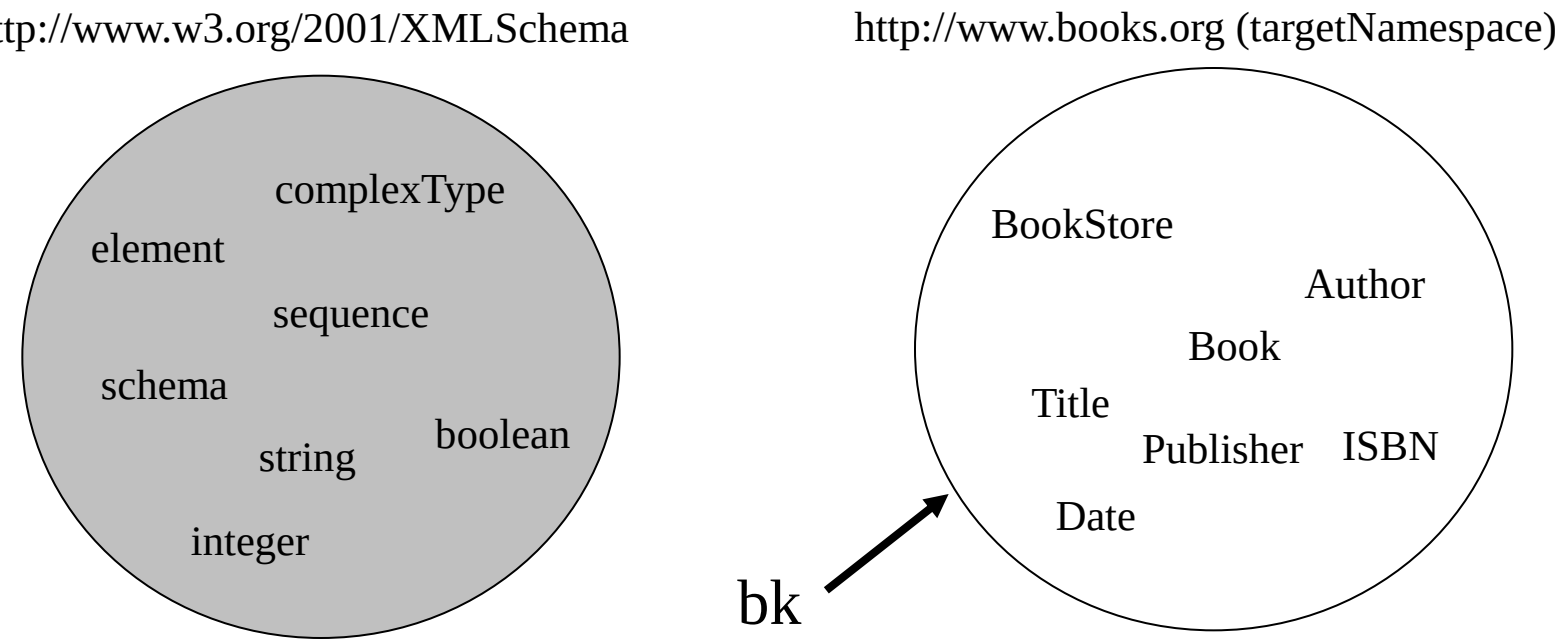
Note that http://…/XMLSchema is the default namespace. Consequently, there are no namespace qualifiers on

- schema
- element
- complexType
- sequence
- string

```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns:bk="http://www.books.org"
            elementFormDefault="qualified">
   <element name="BookStore">
      <complexType>
         <sequence>
            <element ref="bk:Book" minOccurs="1" maxOccurs="unbounded"/>
         </sequence>
      </complexType>
   </element>
   <element name="Book">
      <complexType>
         <sequence>
            <element ref="bk:Title"/>
            <element ref="bk:Author"/>
            <element ref="bk:Date"/>
            <element ref="bk:ISBN"/>
            <element ref="bk:Publisher"/>
         </sequence>
      </complexType>
   </element>
   <element name="Title" type="string"/>
   <element name="Author" type="string"/>
   <element name="Date" type="string"/>
   <element name="ISBN" type="string"/>
   <element name="Publisher" type="string"/>
</schema>
```

Here we are referencing a Book element. Where is that Book element defined? In what namespace? The bk: prefix indicates what namespace this element is in. **bk:** has been set to be the same as the targetNamespace.

"bk:" References the targetNamespace



Consequently, *bk:Book* refers to the Book element in the targetNamespace.

Inlining Element Declarations

- In the previous examples we declared an element and then we ref'ed to that element declaration. Alternatively, we can *inline* the element declarations.
- On the following slide is an alternate (equivalent) way of representing the schema shown previously, using *inlined element declarations*.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book"  maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="Title" type="xsd:string"/>
              <xsd:element name="Author" type="xsd:string"/>
              <xsd:element name="Date" type="xsd:string"/>
              <xsd:element name="ISBN" type="xsd:string"/>
              <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Note that we have moved all the element declarations inline, and we are no longer ref'ing to the element declarations. This results in a much more compact schema!

This way of designing the schema - by inlining everything - is called the *Russian Doll design*.

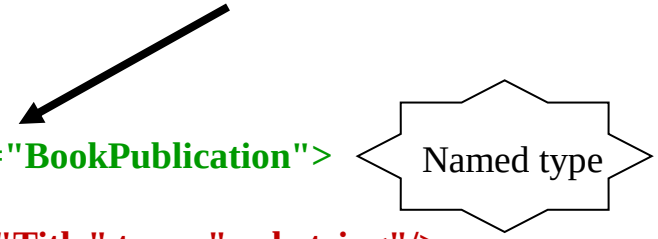
```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book"  maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="Title" type="xsd:string"/>
              <xsd:element name="Author" type="xsd:string"/>
              <xsd:element name="Date" type="xsd:string"/>
              <xsd:element name="ISBN" type="xsd:string"/>
              <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Anonymous types (no name)

Named Types

- The following slide shows an alternate (equivalent) schema which uses a named complexType.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book" type="BookPublication" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:complexType name="BookPublication">
    <xsd:sequence>
      <xsd:element name="Title" type="xsd:string"/>
      <xsd:element name="Author" type="xsd:string"/>
      <xsd:element name="Date" type="xsd:string"/>
      <xsd:element name="ISBN" type="xsd:string"/>
      <xsd:element name="Publisher" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```



The advantage of splitting out Book's element declarations and wrapping them in a named type is that now this type can be *reused* by other elements.

Please note that:

```
<xsd:element name="A" type="T1"/>
<xsd:complexType name="T1">
    <xsd:sequence>
        <xsd:element name="B" …/>
        <xsd:element name="C" …/>
    </xsd:sequence>
</xsd:complexType>
```

Element A *references* the complexType T1.

is equivalent to:

```
<xsd:element name="A">
    <xsd:complexType>
        <xsd:sequence>
            <xsd:element name="B" …/>
            <xsd:element name="C" …/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:element>
```

Element A has the complexType definition *inlined* in the element declaration.

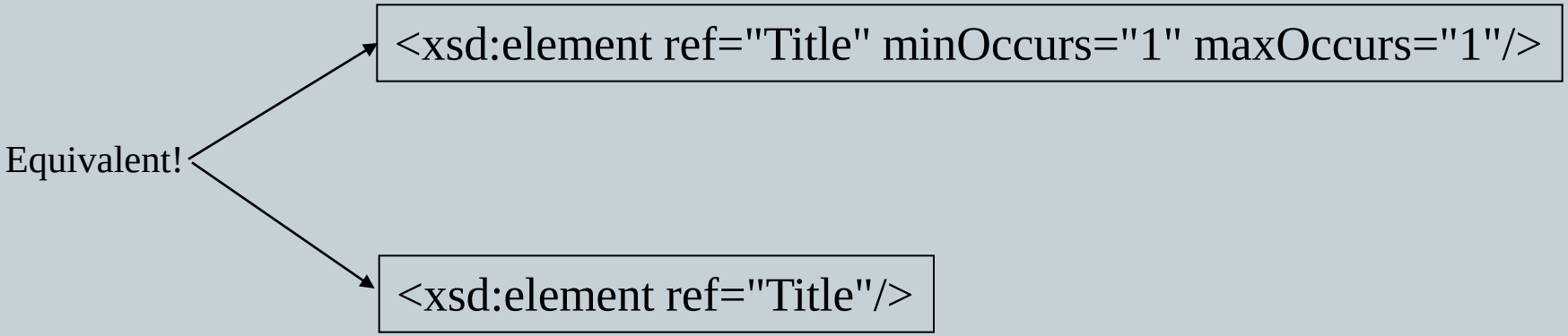
type Attribute or complexType Child Element, but not Both!

- An element declaration can have a type attribute, or a complexType child element, but it cannot have **both** a type attribute and a complexType child element.

```
<xsd:element name="A" type=“T1">
    <xsd:complexType>
        …
    </xsd:complexType>
</xsd:element>
```

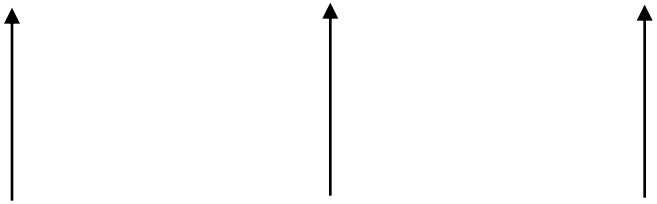
Default Value for minOccurs and maxOccurs

- The default value for minOccurs is "1"
- The default value for maxOccurs is "1"



Summary of Declaring Elements (two ways to do it)

1 <xsd:element name="*name*" type="*type*" minOccurs="*int*" maxOccurs="*int*"/>



A simple type (e.g., xsd:string) or the name of a complexType (e.g., BookPublication)

A nonnegative integer

A nonnegative integer or "unbounded"

Note: minOccurs and maxOccurs can only be used in nested (local) element declarations.

2

```
<xsd:element name="name" minOccurs="int" maxOccurs="int">
    <xsd:complexType>
        …
    </xsd:complexType>
</xsd:element>
```

Problem

- Defining the Date element to be of type string is unsatisfactory (it allows any string value to be input as the content of the Date element, including non-date strings).
  - We would like to constrain the allowable content that Date can have. Modify the BookStore schema to restrict the content of the Date element to just date values (actually, year values. See next two slides).
- Similarly, constrain the content of the ISBN element to content of this form: d-ddddd-ddd-d or d-ddd-ddddd-d or d-dd-dddddd-d, where 'd' stands for 'digit'

The **date** Datatype

- It is a *built-in datatype* (i.e., schema validators know about this datatype)
- The **date** datatype is used to represent a specific day (year-month-day)
- Elements declared to be of type date must follow this form: CCYY-MM-DD
  - range for CC is: 00-99
  - range for YY is: 00-99
  - range for MM is: 01-12
  - range for DD is:
    - 01-28 if month is 2
    - 01-29 if month is 2 and the gYear is a leap gYear
    - 01-30 if month is 4, 6, 9, or 11
    - 01-31 if month is 1, 3, 5, 7, 8, 10, or 12
  - Example: 1999-05-31 represents May 31, 1999

The **gYear** Datatype

- It is a built-in datatype (Gregorian calendar year)
- Elements declared to be of type **gYear** must follow this form: CCYY
  - range for CC is: 00-99
  - range for YY is: 00-99
  - Example: 1999 indicates the gYear 1999

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:simpleType name="ISBNType">
    <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{1}-\d{5}-\d{3}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{3}-\d{5}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{2}-\d{6}-\d{1}"/>
    </xsd:restriction>
  </xsd:simpleType>
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book"  maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="Title" type="xsd:string"/>
              <xsd:element name="Author" type="xsd:string"/>
              <xsd:element name="Date" type="xsd:gYear"/>
              <xsd:element name="ISBN" type="ISBNType"/>
              <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Here we are defining a new (user-defined) data-type, called ISBNType.

Declaring Date to be of type gYear, and ISBN to be of type ISBNType (defined above)

```
<xsd:simpleType name="ISBNType">
   <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{1}-\d{5}-\d{3}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{3}-\d{5}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{2}-\d{6}-\d{1}"/>
   </xsd:restriction>
</xsd:simpleType>
```

"I hereby declare a new type called ISBNType. It is a restricted form of the string type. Elements declared of this type must conform to one of the following patterns:

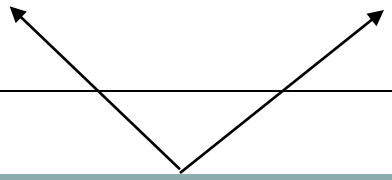
- First Pattern: 1 digit followed by a dash followed by 5 digits followed by another dash followed by 3 digits followed by another dash followed by 1 more digit, or
- Second Pattern: 1 digit followed by a dash followed by 3 digits followed by another dash followed by 5 digits followed by another dash followed by 1 more digit, or
- Third Pattern: 1 digit followed by a dash followed by 2 digits followed by another dash followed by 6 digits followed by another dash followed by 1 more digit."

These patterns are specified using *Regular Expressions*.

Equivalent Expressions


```
<xsd:simpleType name="ISBNType">
   <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{1}-\d{5}-\d{3}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{3}-\d{5}-\d{1}"/>
      <xsd:pattern value="\d{1}-\d{2}-\d{6}-\d{1}"/>
   </xsd:restriction>
</xsd:simpleType>
```

```
<xsd:simpleType name="ISBNType">
   <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{1}-\d{5}-\d{3}-\d{1}|\d{1}-\d{3}-\d{5}-\d{1}|\d{1}-\d{2}-\d{6}-\d{1}"/>
   </xsd:restriction>
</xsd:simpleType>
```



The vertical bar means "or"

<xsd:complexType> or <xsd:simpleType>?

- When do you use the complexType element and when do you use the simpleType element?
  - Use the complexType element when you want to define the structure of an element: i.e., child elements and/or attributes of an element
  - Use the simpleType element when you want to create a new type that is a refinement of a built-in type (string, date, gYear, etc…)

Built-in Datatypes

- Primitive Datatypes
  - string
  - boolean
  - decimal
  - float
  - double
  - duration
  - dateTime
  - time
  - date
  - gYearMonth
  - gYear
  - gMonthDay

- Atomic, built-in
  - **"Hello World"**
  - **{true, false, 1, 0}**
  - **7.08**
  - **12.56E3, 12, 12560, 0, -0, INF, -INF, NAN**
  - **12.56E3, 12, 12560, 0, -0, INF, -INF, NAN**
  - **P1Y2M3DT10H30M12.3S**
  - format: *CCYY-MM-DDThh:mm:ss*
  - format: *hh:mm:ss.sss*
  - format: *CCYY-MM-DD*
  - format: *CCYY-MM*
  - format: *CCYY*
  - format: *--MM-DD*

Note: 'T' is the date/time separator
INF = infinity
NAN = not-a-number

Built-in Datatypes (cont.)

- Primitive Datatypes
  - gDay
  - gMonth
  - hexBinary
  - base64Binary
  - anyURI
  - QName
  - NOTATION

- Atomic, built-in
  - format: ---*DD* (note the 3 dashes)
  - format: --*MM*--
  - a hex string
  - a base64 string
  - **http://www.xfront.com**
  - a namespace qualified name
  - a NOTATION from the XML spec

Built-in Datatypes (cont.)

Derived types	Subtype of primitive datatype
normalizedString	A string without tabs, line feeds, or carriage returns
token	String w/o tabs, l/f, leading/trailing spaces, consecutive spaces
language	any valid xml:lang value, e.g., **EN**, **FR**, ...
IDREFS	must be used only with attributes
ENTITIES	must be used only with attributes
NMTOKEN	must be used only with attributes
NMTOKENS	must be used only with attributes
Name	
NCName	**part** (no namespace qualifier)
ID	must be used only with attributes
IDREF	must be used only with attributes
ENTITY	must be used only with attributes
integer	**456**
nonPositiveInteger	negative infinity to 0

Built-in Datatypes (cont.)

Derived types	Subtype of primitive datatype
negativeInteger	negative infinity to -1
long	**-9223372036854775808** to **9223372036854775807**
int	**-2147483648** to **2147483647**
short	**-32768** to **32767**
byte	**-127** to **128**
nonNegativeInteger	0 to infinity
unsignedLong	**0** to **18446744073709551615**
unsignedInt	**0** to **4294967295**
unsignedShort	**0** to **65535**
unsignedByte	**0** to **255**
positiveInteger	**1** to infinity

Note: the following types can only be used with attributes (which we will discuss later):
ID, IDREF, IDREFS, NMTOKEN, NMTOKENS, ENTITY, and ENTITIES.

Creating your own Datatypes

- A new datatype can be defined from an existing datatype (called the "base" type) by specifying values for one or more of the optional *facets* for the base type.
- Example. The string primitive datatype has six optional facets:
  - length
  - minLength
  - maxLength
  - pattern
  - enumeration
  - whitespace (legal values: preserve, replace, collapse)

Example of Creating a New Datatype by Specifying Facet Values


```
<xsd:simpleType name="TelephoneNumber"> (1)
   <xsd:restriction base="xsd:string">(2)
      <xsd:length value="8"/> (3)
      <xsd:pattern value="\d{3}-\d{4}"/>(4)
   </xsd:restriction>
</xsd:simpleType>
```

1. This creates a new datatype called 'TelephoneNumber'.
2. Elements of this type can hold string values,
3. But the string length must be exactly 8 characters long *and*
4. The string must follow the pattern: ddd-dddd, where 'd' represents a 'digit'.

(Obviously, in this example the regular expression makes the length facet redundant.)

Another Example


```
<xsd:simpleType name="shape">
   <xsd:restriction base="xsd:string">
      <xsd:enumeration value="circle"/>
      <xsd:enumeration value="triangle"/>
      <xsd:enumeration value="square"/>
   </xsd:restriction>
</xsd:simpleType>
```

This creates a new type called shape.
An element declared to be of this type
must have either the value circle, or triangle, or square.

Facets of the integer Datatype

- The integer datatype has 8 optional facets:
  - totalDigits
  - pattern
  - whitespace
  - enumeration
  - maxInclusive
  - maxExclusive
  - minInclusive
  - minExclusive

Example


```
<xsd:simpleType name= "EarthSurfaceElevation">
   <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="-1290"/>
      <xsd:maxInclusive value="29035"/>
   </xsd:restriction>
</xsd:simpleType>
```

This creates a new datatype called 'EarthSurfaceElevation'. Elements declared to be of this type can hold an integer. However, the integer is restricted to have a value between -1290 and 29035, inclusive.

General Form of Creating a New Datatype by Specifying Facet Values


```
<xsd:simpleType name= "name">
   <xsd:restriction base= "xsd:source">
      <xsd:facet value= "value"/>
      <xsd:facet value= "value"/>
      …
   </xsd:restriction>
</xsd:simpleType>
```

Facets:
- length
- minlength
- maxlength
- pattern
- enumeration
- minInclusive
- maxInclusive
- minExclusive
- maxExclusive

...

Sources:
- string
- boolean
- number
- float
- double
- duration
- dateTime
- time

...

Multiple Facets - "*and*" them together, or "*or*" them together?


```
<xsd:simpleType name="TelephoneNumber">
   <xsd:restriction base="xsd:string">
      <xsd:length value="8"/>
      <xsd:pattern value="\d{3}-\d{4}"/>
   </xsd:restriction>
</xsd:simpleType>
```

An element declared to be of type TelephoneNumber must be a string of length=8 *and* the string must follow the pattern: 3 digits, dash, 4 digits.

```
<xsd:simpleType name="shape">
   <xsd:restriction base="xsd:string">
      <xsd:enumeration value="circle"/>
      <xsd:enumeration value="triangle"/>
      <xsd:enumeration value="square"/>
   </xsd:restriction>
</xsd:simpleType>
```

An element declared to be of type shape must be a string with a value of *either* circle, *or* triangle, *or* square.

Patterns, enumerations => "or" them together

All other facets => "and" them together

Regular Expressions

- Recall that the string datatype has a ***pattern*** facet. The value of a pattern facet is a regular expression. Below are some examples of regular expressions:

Regular Expression	Example
- Chapter \d	- Chapter 1
- Chapter \d	- Chapter 1
- a*b	- b, ab, aab, aaab, …
- [xyz]b	- xb, yb, zb
- a?b	- b, ab
- a+b	- ab, aab, aaab, …
- [a-c]x	- ax, bx, cx

Regular Expressions (cont.)

- Regular Expression
  - [a-c]x
  - [-ac]x
  - [ac-]x
  - [^0-9]x
  - \Dx
  - Chapter\s\d
  - (ho){2} there
  - (ho\s){2} there
  - .abc
  - (a|b)+x
- Example
  - ax, bx, cx
  - -x, ax, cx
  - ax, cx, -x
  - any non-digit char followed by x
  - any non-digit char followed by x
  - Chapter followed by a blank followed by a digit
  - hoho there
  - ho ho there
  - any (*one*) char followed by abc
  - ax, bx, aax, bbx, abx, bax,…

Regular Expressions (cont.)

- a{1,3}x
- a{2,}x
- \w\s\w
- [a-zA-Z]*
- \.

- ax, aax, aaax
- aax, aaax, aaaax, …
- word **character** (alphanumeric plus dash) followed by a space followed by a word character
- A string comprised of any lower and upper case letters
- The period "." (Without the backward slash the period means "any character")

Creating a simpleType from another simpleType

- Thus far we have created a simpleType using one of the built-in datatypes as our base type.
- However, we can create a simpleType that uses another simpleType as the base. See next slide.

```
<xsd:simpleType name= "EarthSurfaceElevation">
   <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="-1290"/>
      <xsd:maxInclusive value="29035"/>
   </xsd:restriction>
</xsd:simpleType>
```

```
<xsd:simpleType name= "BostonAreaSurfaceElevation">
   <xsd:restriction base="EarthSurfaceElevation">
      <xsd:minInclusive value="0"/>
      <xsd:maxInclusive value="120"/>
   </xsd:restriction>
</xsd:simpleType>
```

This simpleType uses EarthSurfaceElevation as its base type.

Fixing a Facet Value

- Sometimes when we define a simpleType we want to require that one (or more) facet have an unchanging value. That is, we want to make the facet a constant in all “subtypes”.

```
<xsd:simpleType name= "ClassSize">
   <xsd:restriction base="xsd:nonNegativeInteger">
      <xsd:minInclusive value="10" fixed="true"/>
      <xsd:maxInclusive value="60"/>
   </xsd:restriction>
</xsd:simpleType>
```

simpleTypes which derive from this simpleType may not change this facet.

```
<xsd:simpleType name= "ClassSize">
   <xsd:restriction base="xsd:nonNegativeInteger">
      <xsd:minInclusive value="10" fixed="true"/>
      <xsd:maxInclusive value="60"/>
   </xsd:restriction>
</xsd:simpleType>
```

```
<xsd:simpleType name= "BostonIEEEClassSize">
   <xsd:restriction base="ClassSize">
      <xsd:minInclusive value="15"/>
      <xsd:maxInclusive value="60"/>
   </xsd:restriction>
</xsd:simpleType>
```

Error! Cannot change the value of a fixed facet!

Element Containing a User-Defined Simple Type

Example. Create a schema element declaration for an elevation element.
Declare the elevation element to be an integer with a range -1290 to 29035

```
<elevation>5240</elevation>
```

Here's one way of declaring the elevation element:

```
<xsd:simpleType name="EarthSurfaceElevation">
   <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="-1290"/>
      <xsd:maxInclusive value="29035"/>
   </xsd:restriction>
</xsd:simpleType>
<xsd:element name="elevation" type="EarthSurfaceElevation"/>
```

Element Containing a User-Defined Simple Type (cont.)

Here's an alternative method for declaring elevation:

```
<xsd:element name="elevation">
   <xsd:simpleType>
      <xsd:restriction base="xsd:integer">
         <xsd:minInclusive value="-1290"/>
         <xsd:maxInclusive value="29035"/>
      </xsd:restriction>
   </xsd:simpleType>
</xsd:element>
```

The simpleType definition is defined inline, it is an *anonymous* simpleType definition.

The disadvantage of this approach is that this simpleType may not be reused by other elements.

Summary of Declaring Elements (three ways to do it)

1

```
<xsd:element name="name" type="type" minOccurs="int" maxOccurs="int"/>
```

2

```
<xsd:element name="name" minOccurs="int" maxOccurs="int">
    <xsd:complexType>
        …
    </xsd:complexType>
</xsd:element>
```

3

```
<xsd:element name="name" minOccurs="int" maxOccurs="int">
    <xsd:simpleType>
        <xsd:restriction base="type">
        …
        </xsd:restriction>
    </xsd:simpleType>
</xsd:element>
```

Annotating Schemas

- The <annotation> element is used for documenting the schema, both for humans and for programs.
  - Use <documentation> for providing a comment to humans
  - Use <appinfo> for providing a comment to programs
    - The content is any well-formed XML
- Note that annotations have no effect on schema validation

```
<xsd:annotation>
  <xsd:documentation>
      The following constraint is not expressible with XML Schema: The value of element A should be greater
      than the value of element B.  So, we need to use a separate tool (e.g., Schematron) to check this constraint.
     We will express this constraint in the appinfo section (below).
  </xsd:documentation>
  <xsd:appinfo>
      <assert test="A &gt; B">A should be greater than B</assert>
  </xsd:appinfo>
<xsd:/annotation>
```

Where Can You Put Annotations?

- You cannot put annotations at just any random location in the schema.
- Here are the rules for where an annotation element can go:
  - annotations may occur before and after any global component
  - annotations may occur only at the beginning of non-global components

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                  targetNamespace="http://www.books.org"
                 xmlns="http://www.books.org"
                 elementFormDefault="qualified">
   <xsd:element name="BookStore">
      <xsd:complexType>
         <xsd:sequence>
            <xsd:element name="Book"  maxOccurs="unbounded">
               <xsd:complexType>
                  <xsd:sequence>
                     <xsd:element name="Title" type="xsd:string"/>
                     <xsd:element name="Author" type="xsd:string"/>
                     <xsd:element name="Date" type="xsd:string"/>
                     <xsd:element name="ISBN" type="xsd:string"/>
                     <xsd:element name="Publisher" type="xsd:string"/>
                  </xsd:sequence>
               </xsd:complexType>
            </xsd:element>
         </xsd:sequence>
      </xsd:complexType>
   </xsd:element>
</xsd:schema>
```

Can put annotations only at these locations

Suppose that you want to annotate, say, the Date element declaration. What do we do? See next page ...

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                 targetNamespace="http://www.books.org"
                xmlns="http://www.books.org"
                elementFormDefault="qualified">
   <xsd:element name="BookStore">
      <xsd:complexType>
         <xsd:sequence>
            <xsd:element name="Book"  maxOccurs="unbounded">
               <xsd:complexType>
                  <xsd:sequence>
                     <xsd:element name="Title" type="xsd:string"/>
                     <xsd:element name="Author" type="xsd:string"/>
                     <xsd:element name="Date" type="xsd:string">
                        <xsd:annotation>
                           <xsd:documentation>This is how to annotate the Date element!</xsd:documentation>
                        </xsd:annotation>
                     </xsd:element>
                     <xsd:element name="ISBN" type="xsd:string"/>
                     <xsd:element name="Publisher" type="xsd:string"/>
                  </xsd:sequence>
               </xsd:complexType>
            </xsd:element>
         </xsd:sequence>
      </xsd:complexType>
   </xsd:element>
</xsd:schema>
```

Inline the annotation within the Date element declaration.

Two Optional Attributes for the documentation Element

- In the previous example we showed <xsd:documentation> with no attributes. Actually, it can have two attributes:
  - source: this attribute contains a URL to a file which contains supplemental information
  - xml:lang: this attribute specifies the language that the documentation was written in

```
<xsd:documentation source="http://www.xfront.com/BookReview.txt"  xml:lang="FR"/>
```

One Optional Attribute for the appinfo Element

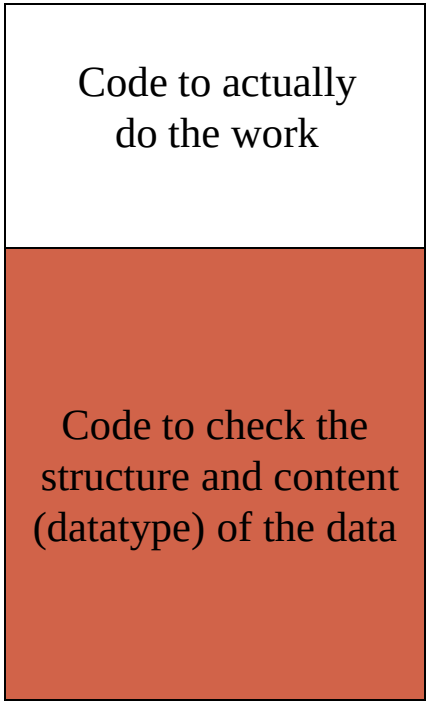
- In the previous example we showed <xsd:appinfo> with no attributes. Actually, it can have one attribute:
  - source: this attribute contains a URL to a file which contains supplemental information

```
<xsd:appinfo source="http://www.xfront.com/Assertions.xml"/>
```

Up for a Breath

- Wow! We have really been into the depths of XML Schemas.
- Let's back up for a moment and look at XML Schemas from a "big picture" point of view.

Save $$$ using XML Schemas

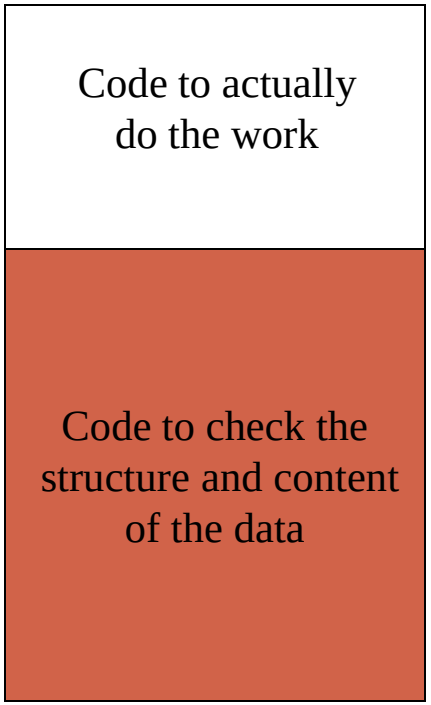


"In a typical program, up to 60% of the code is spent checking the data!"

- source unknown

Continued

Save $$$ using XML Schemas (cont.)

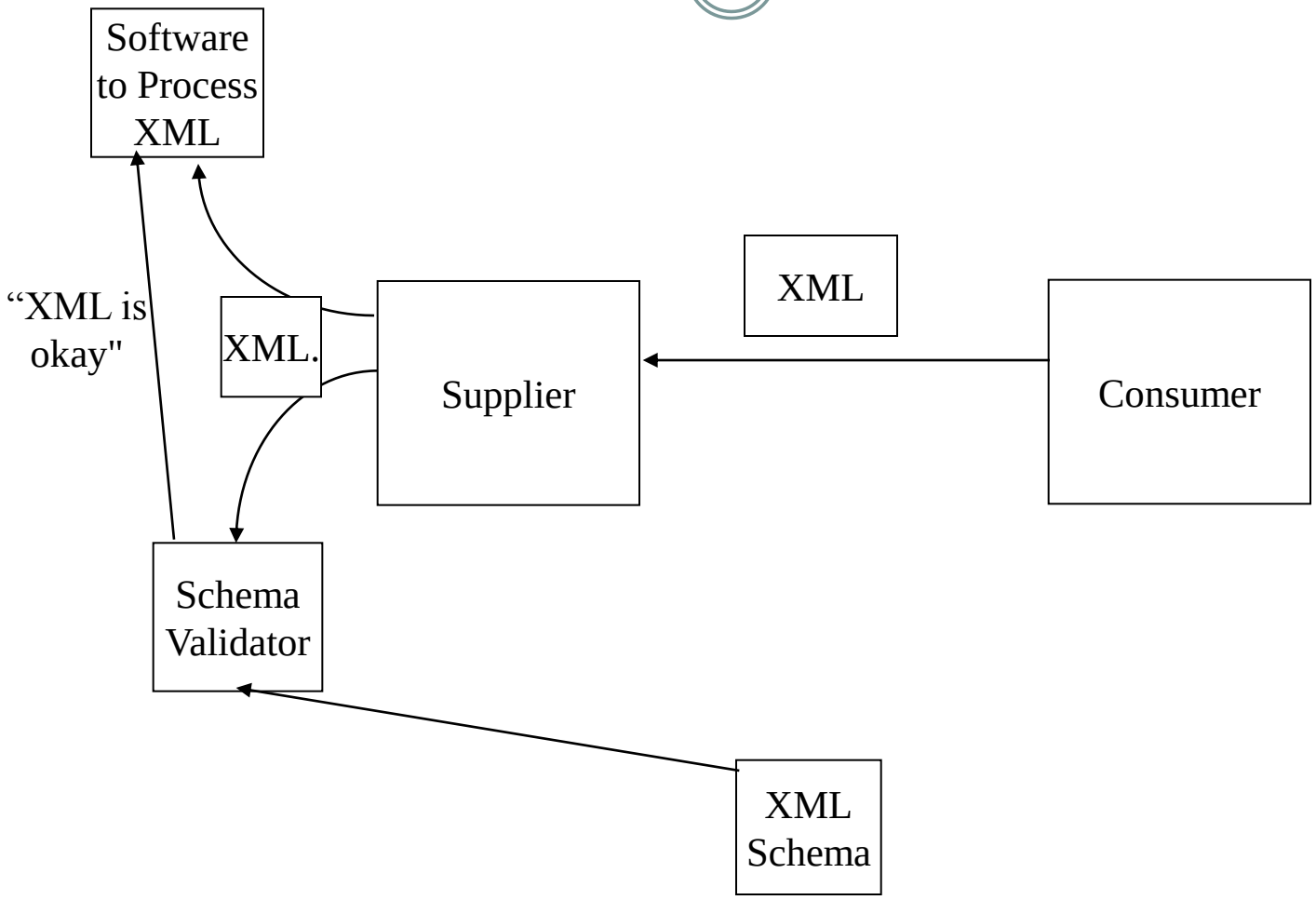


If your data is structured as XML, and there is a schema, then you can hand the data-checking task off to a schema validator.

Thus, your code is reduced by up to 60%!!!

Big $$ savings!

Classic use of XML Schemas



(Schema at third-party, neutral web site)

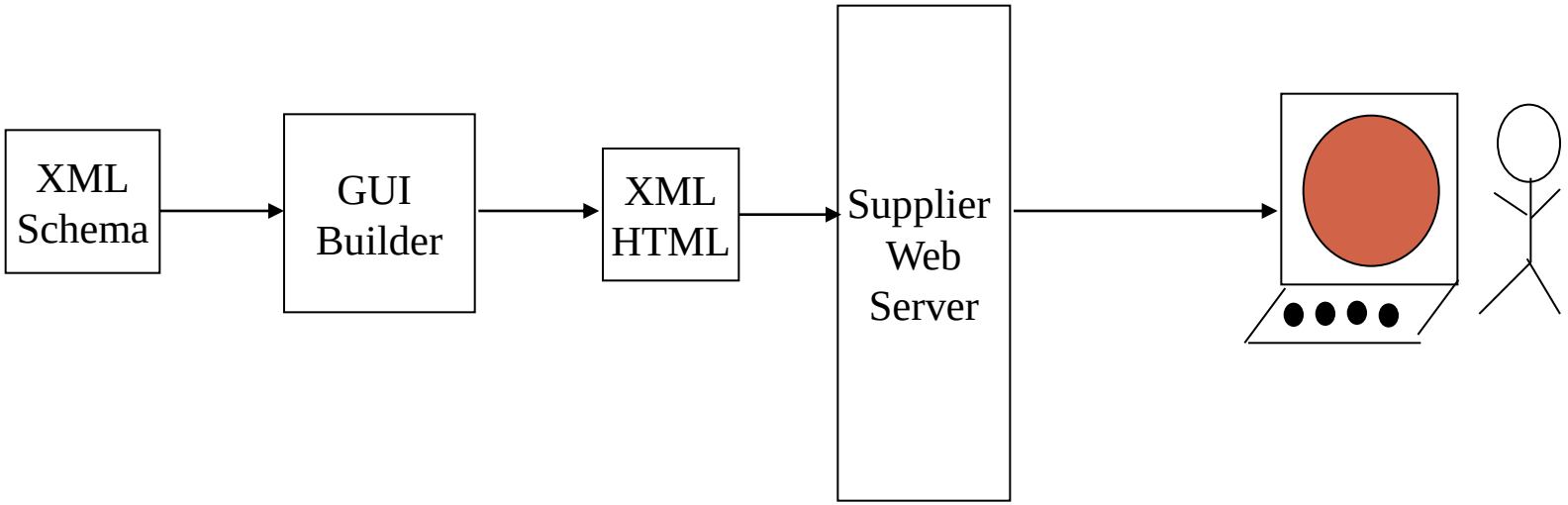
What are XML Schemas?

- **Data Model**
  - With XML Schemas you specify how your XML data will be organized, and the datatypes of your data. That is, with XML Schemas you model how your data is to be represented in an instance document.
- **A Contract**
  - Organizations agree to structure their XML documents in conformance with an XML Schema. Thus, the XML Schema acts as a contract between the organizations.
- **A rich source of metadata**
  - An XML Schema document contains lots of data about the data in the XML instance documents, such as the datatype of the data, the data's range of values, how the data is related to another piece of data (parent/child, sibling relationship), i.e., XML Schemas contain metadata

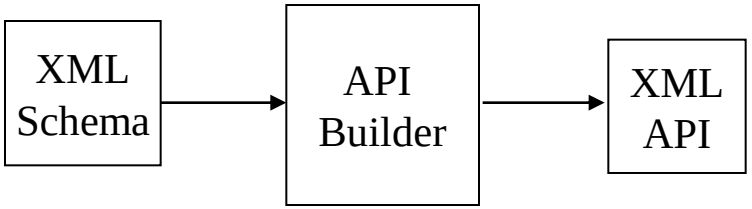
No Limits

- Two slides back we showed the classic use of XML Schemas - to validate your data (so that you don't have to write code to do it)
- However, there are many other uses for XML Schemas. Schemas are a wonderful source of metadata.
- Truly, *your imagination is the only limit on its usefulness*.
- On the next slide I show how to use the metadata provided by XML Schemas to create a GUI. The slide after that shows how to automatically generate an API using the metadata in XML Schemas. Following that is a slide showing how to create a "smart editor" using XML Schemas.

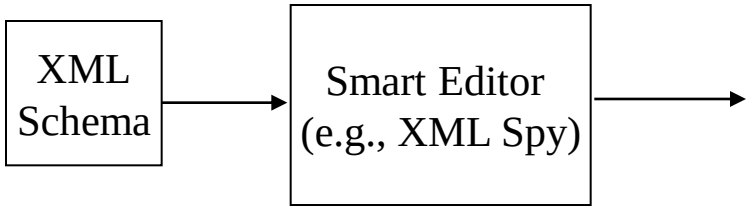
XML Schema --> GUI



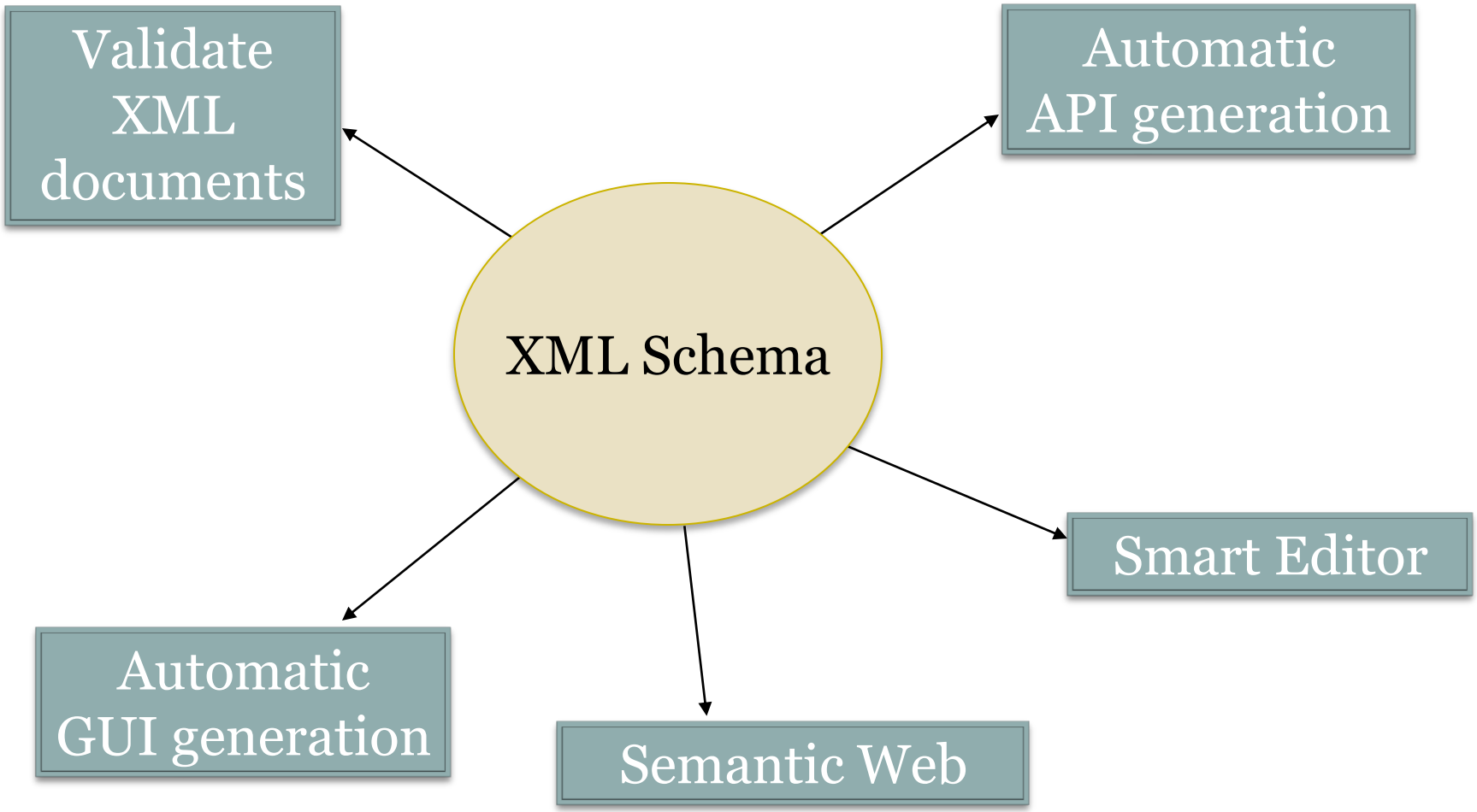
XML Schema --> API



XML Schema --> Smart Editor



Helps you build your instance documents. For example, it pops up a menu showing you what is valid next. It knows this by looking at the XML Schema!


Validate XML documents
Automatic API generation
XML Schema
Smart Editor
Automatic GUI generation
Semantic Web

Derived Types

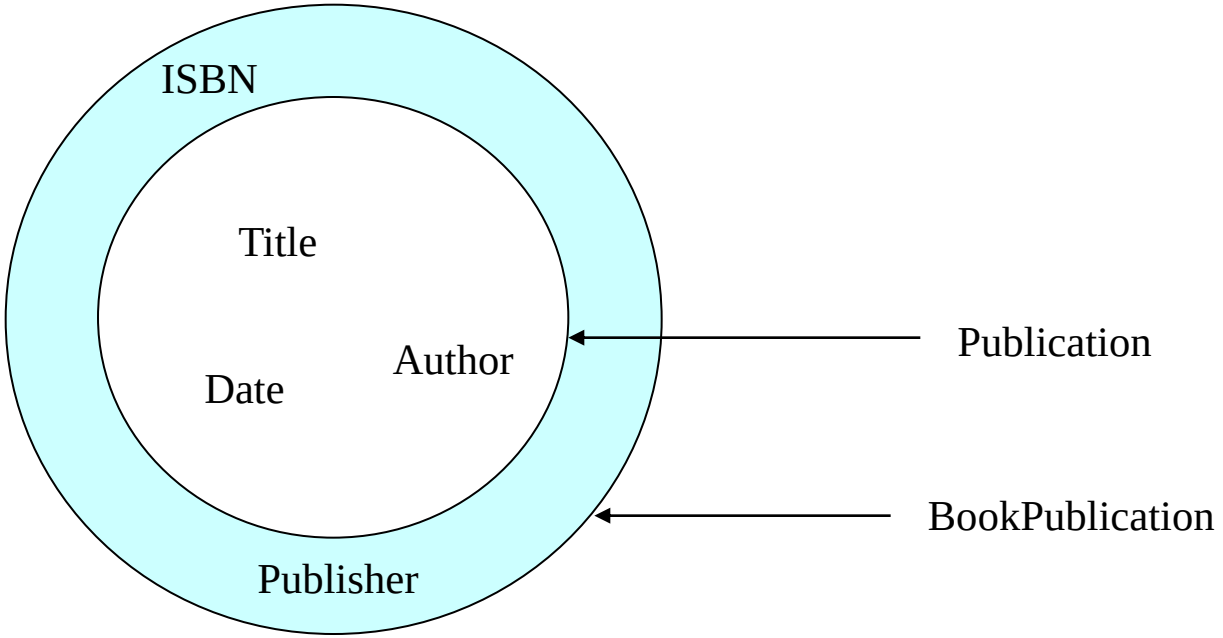
- We can do a form of **subclassing** complexType definitions. We call this "derived types"
  - derive by extension: extend the parent complexType with more elements
  - derive by restriction: create a type which is a subset of the base type. There are two ways to subset the elements:
    - redefine a type element to have a **restricted range of values**, or
    - redefine a type element to have a more **restricted number of occurrences**.

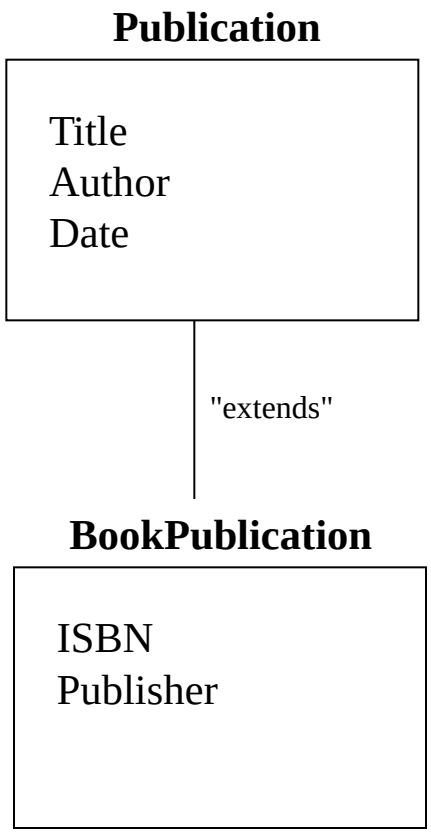
```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                targetNamespace="http://www.books.org"
                xmlns="http://www.books.org"
               elementFormDefault="qualified">
  <xsd:complexType name="Publication">
    <xsd:sequence>
      <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Date" type="xsd:gYear"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="BookPublication">
    <xsd:complexContent>
      <xsd:extension base="Publication" >
        <xsd:sequence>
          <xsd:element name="ISBN" type="xsd:string"/>
          <xsd:element name="Publisher" type="xsd:string"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book" type="BookPublication" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Note that BookPublication extends the Publication type, i.e., we are doing Derive by Extension

```
<xsd:complexType name="Publication">
    <xsd:sequence>
      <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Date" type="xsd:gYear"/>
    </xsd:sequence>
</xsd:complexType >
<xsd:complexType name="BookPublication">
   <xsd:complexContent>
      <xsd:extension base="Publication">
         <xsd:sequence>
            <xsd:element name="ISBN" type="xsd:string"/>
            <xsd:element name="Publisher" type="xsd:string"/>
         </xsd:sequence>
      </xsd:extension>
   </xsd:complexContent>
</xsd:complexType >
```

Elements declared to be of type BookPublication will have 5 child elements - Title, Author, Date, ISBN, and Publisher. Note that the elements in the derived type are **appended** to the elements in the base type.


ISBN
Title
Author
Date
Publisher
Publication
BookPublication


Publication
Title
Author
Date
"extends"
BookPublication
ISBN
Publisher

Derive by Restriction


```
<xsd:complexType name="Publication">
      <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Date" type="xsd:gYear"/>
      </xsd:sequence>
</xsd:complexType>
<xsd:complexType name= "SingleAuthorPublication">
   <xsd:complexContent>
      <xsd:restriction base="Publication">
         <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Author" type="xsd:string"/>
            <xsd:element name="Date" type="xsd:gYear"/>
         </xsd:sequence>
      </xsd:restriction>
   </xsd:complexContent>
</xsd:complexType>
```

Elements of type SingleAuthorPublication will have 3 child elements - Title, Author, and Date. However, there must be exactly one Author element.

Note that in the restriction type you must repeat all the declarations from the base type (except when the base type has an element with minOccurs="0" and the subtype wishes to delete it. See next slide).

Deleting an element in the base type


```
<xsd:complexType name="Publication">
      <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Author" type="xsd:string" minOccurs="0"/>
            <xsd:element name="Date" type="xsd:gYear"/>
      </xsd:sequence>
</xsd:complexType>
<xsd:complexType name= "ZeroAuthorPublication">
   <xsd:complexContent>
      <xsd:restriction base="Publication">
         <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Date" type="xsd:gYear"/>
         </xsd:sequence>
      </xsd:restriction>
   </xsd:complexContent>
</xsd:complexType>
```

Note that in this subtype we have eliminated the Author element, i.e., the subtype is just comprised of an unbounded number of Title elements followed by a single Date element.

If the base type has an element with minOccurs="0", and the subtype wishes to not have that element, then it can simply leave it out.

Derive by Restriction (cont.)

- You might (legitimately) ask:
  - why do I have to repeat all the declarations from the base type? Why can't I simply show the delta (i.e., show those declarations that are changed)?
  - What's the advantage of doing derived by restriction if I have to repeat everything? I'm certainly not saving on typing.
- Answer:
  - Even though you have to retype everything in the base type there are advantages to explicitly associating a type with a base type. In a few slides we will see *element substitution* - the ability to substitute one element for another. A restriction of element substitution is that the substituting element have a type that derives from the type of the element it is substituting. Thus, it is beneficial to link the type.
  - Also, an element's content model may be substituted by the content model of derived types. Thus, the content of an element that has been declared to be of type Publication can be substituted with a SingleAuthorPublication content since SingleAuthorPublication derives from Publication. We will discuss this *type substitutability* in detail later.

Prohibiting Derivations

- Sometimes we may want to create a type and disallow all derivations of it, or just disallow extension derivations, or disallow restriction derivations.
  - Rationale: "For example, I may create a complexType and make it publicly available for others to use. However, I don't want them to extend it with their proprietary extensions or subset it to remove, say, copyright information." (Jon Cleaver)

```
<xsd:complexType name="Publication" final="#all" …>   Publication cannot be extended nor restricted

<xsd:complexType name="Publication" final="restriction" …>   Publication cannot be restricted

<xsd:complexType name="Publication" final="extension" …>   Publication cannot be extended
```

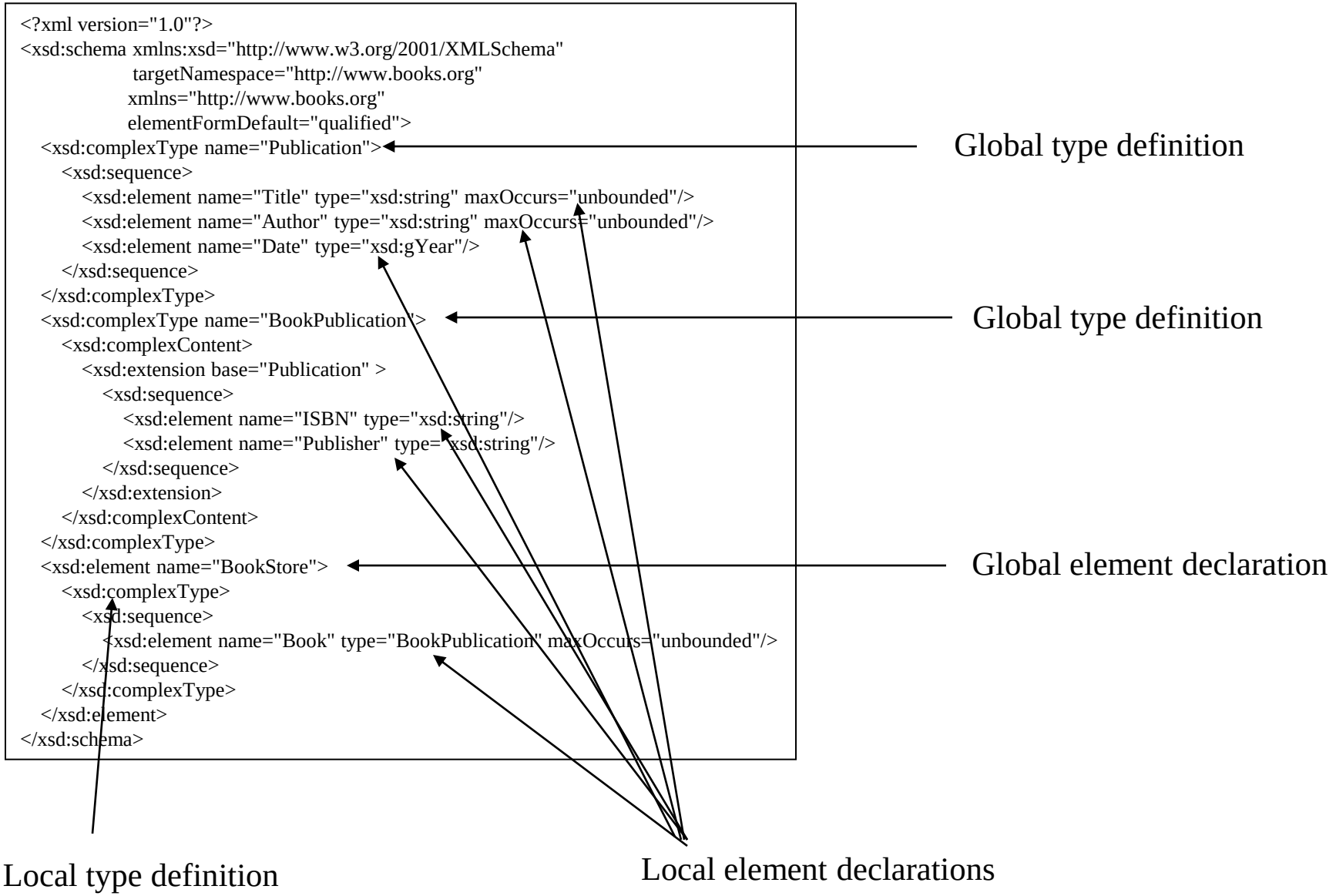
Terminology: Declaration vs Definition

- In a schema:
  - You *declare* elements and attributes. Schema components that are *declared* are those that have a representation in an XML instance document.
  - You *define* components that are used just within the schema document(s). Schema components that are *defined* are those that have no representation in an XML instance document.

Declarations:	**Definitions**:
- element declarations - attribute declarations	- type (simple, complex) definitions - attribute group definitions - model group definitions

Terminology: Global versus Local

- Global element declarations, global type definitions:
  - These are element declarations/type definitions that are immediate children of <schema>
- Local element declarations, local type definitions:
  - These are element declarations/type definitions that are nested within other elements/types.


<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.books.org"
xmlns="http://www.books.org"
elementFormDefault="qualified">
<xsd:complexType name="Publication">
<xsd:sequence>
<xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
<xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
<xsd:element name="Date" type="xsd:gYear"/>
</xsd:sequence>
</xsd:complexType>
<xsd:complexType name="BookPublication">
<xsd:complexContent>
<xsd:extension base="Publication" >
<xsd:sequence>
<xsd:element name="ISBN" type="xsd:string"/>
<xsd:element name="Publisher" type="xsd:string"/>
</xsd:sequence>
</xsd:extension>
</xsd:complexContent>
</xsd:complexType>
<xsd:element name="BookStore">
<xsd:complexType>
<xsd:sequence>
<xsd:element name="Book" type="BookPublication" maxOccurs="unbounded"/>
</xsd:sequence>
</xsd:complexType>
</xsd:element>
</xsd:schema>
Global type definition
Global type definition
Global element declaration
Local type definition
Local element declarations

Global vs Local … What's the Big Deal?

- So what if an element or type is global or local. What practical impact does it have?
  - Answer: **only global elements/types can be referenced (i.e., reused)**. Thus, if an element/type is local then it is effectively invisible to the rest of the schema (and to other schemas).

Element Substitution

- Oftentimes in daily conversation there are several ways to express something.
  - In Boston we use the words "T" and "subway" interchangeably. For example, "we took the T into town", or "we took the subway into town".
    - Thus, "T" and "subway" are substitutable. Which one is used may depend upon what part of the state you live in, what mood you're in, or any number of factors.
- We would like to be able to express this substitutability in XML Schemas.
  - That is, we would like to be able to declare in a schema an element called "subway", an element called "T", and state that "T"may be substituted for "subway". Instance documents can then use either <subway> or <T>, depending on their preference.

substitutionGroup

- We can define a group of substitutable elements (called a substitutionGroup) by declaring an element (called the head) and then declaring other elements which state that they are substitutable for the head element.

```
<xsd:element name="subway" type="xsd:string"/>
<xsd:element name="T" substitutionGroup="subway" type="xsd:string"/>
```

subway is the *head* element
T is substitutable for subway

So what's the big deal?
- Anywhere a head element can be used in an instance document, any member of the substitutionGroup can be substituted!

Schema:

```
<xsd:element name="subway" type="xsd:string"/>
<xsd:element name="T" substitutionGroup="subway" type="xsd:string"/>
<xsd:element name="transportation">
   <xsd:complexType>
      <xsd:sequence>
         <xsd:element ref="subway"/>
      </xsd:sequence>
   </xsd:complexType>
</xsd:element>
```

Instance doc:

```
<transportation>
   <subway>Red Line</subway>
</transportation>
```

Alternative instance doc (substitute T for subway):

```
<transportation>
   <T>Red Line</T>
</transportation>
```

This example shows the <subway> element being substituted with the <T> element.

International Clients

- We can use substitutionGroups to create elements customized for our international clients. On the next slide is shown a Spanish version of the element.

Schema:

```
<xsd:element name="subway" type="xsd:string"/>
<xsd:element name="metro" substitutionGroup="subway" type="xsd:string"/>
<xsd:complexType name="transport">
    <xsd:sequence>
        <xsd:element ref="subway"/>
    </xsd:sequence>
</xsd:complexType>
<xsd:element name="transportation" type="transport"/>
<xsd:element name="transporte" substitutionGroup="transportation"/>
```

Instance doc:

```
<transportation>
    <subway>Red Line</subway>
</transportation>
```

Alternative instance doc (customized for our Spanish clients):

```
<transporte>
    <metro>Linea Roja</metro>
</transporte>
```

Notes about using substitutionGroup

- The elements that are declared to be in the substitution group (e.g., subway and T) **must be declared as global elements**
- If the type of a substitutionGroup element is the same as the head element then you can omit it (the type)
  - In our Subway example we could have omitted the type attribute in the declaration of the T element since it is the same as Subway's type (xsd:string).

```
<xsd:element name="T" substitutionGroup="subway"/>
```

Notes about using substitutionGroup (cont.)

- The type of every element in the substitutionGroup must be the same as, or derived from, the type of the head element.

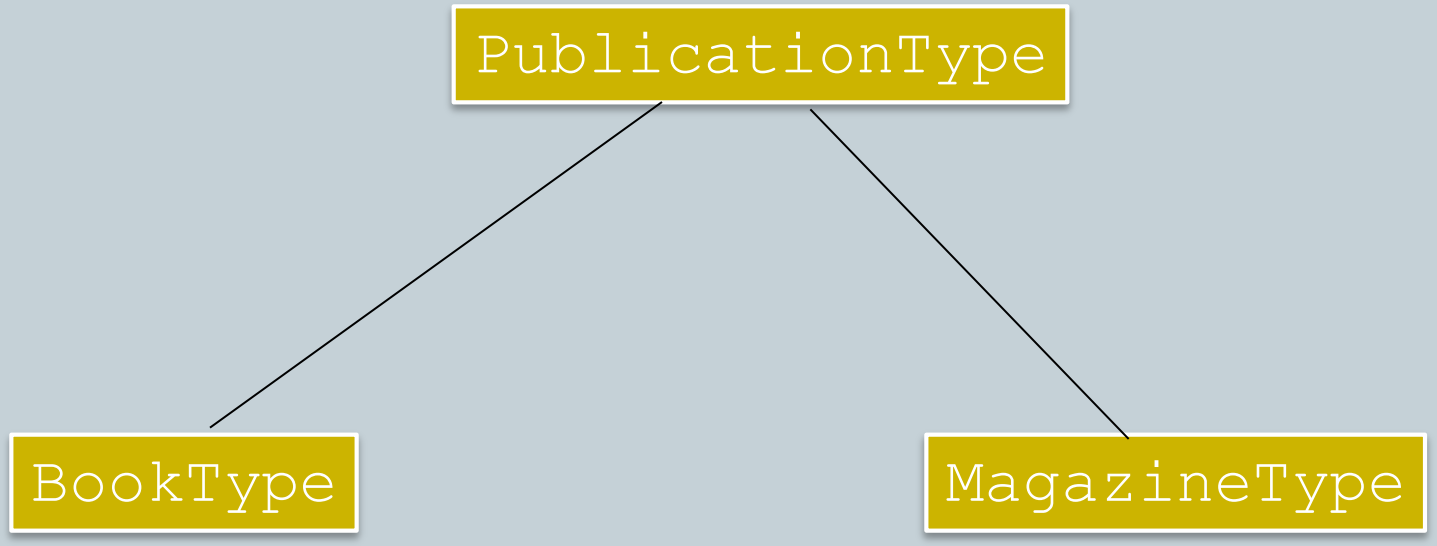
```
<xsd:element name="A" type="xxx"/>

<xsd:element name="B"   substitutionGroup="A" type="yyy"/>
```


This type must be the same as "xxx" or, it must be derived from "xxx".

Element Substitution with Derived Types

```
<xsd:element name="Publication" type="PublicationType"/>
<xsd:element name="Book" substitutionGroup="Publication" type="BookType"/>
<xsd:element name="Magazine" substitutionGroup="Publication" type="MagazineType"/>
<xsd:element name="BookStore">
   <xsd:complexType>
         <xsd:sequence>
            <xsd:element ref="Publication" maxOccurs="unbounded"/>
         </xsd:sequence>
   </xsd:complexType>
</xsd:element>
```

BookType and MagazineType derive from PublicationType



In order for Book and Magazine to be in a substitutionGroup with Publication, their type (BookType and MagazineType, respectively) must be the same as, or derived from Publication's type (PublicationType)

```
<xsd:complexType name="PublicationType">
      <xsd:sequence>
         <xsd:element name="Title" type="xsd:string"/>
         <xsd:element name="Author" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>
         <xsd:element name="Date" type="xsd:gYear"/>
      </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="BookType">
      <xsd:complexContent>
         <xsd:extension base="PublicationType" >
            <xsd:sequence>
               <xsd:element name="ISBN" type="xsd:string"/>
               <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:sequence>
         </xsd:extension>
      </xsd:complexContent>
</xsd:complexType>
<xsd:complexType name="MagazineType">
      <xsd:complexContent>
         <xsd:restriction base="PublicationType">
            <xsd:sequence>
               <xsd:element name="Title" type="xsd:string"/>
               <xsd:element name="Date" type="xsd:gYear"/>
            </xsd:sequence>
         </xsd:restriction>
      </xsd:complexContent>
</xsd:complexType>
```

```
<?xml version="1.0"?>
<BookStore …>
      <Book>
            <Title>Illusions: The Adventures of a Reluctant Messiah</Title>
            <Author>Richard Bach</Author>
            <Date>1977</Date>
            <ISBN>0-440-34319-4</ISBN>
            <Publisher>Dell Publishing Co.</Publisher>
      </Book>
      <Magazine>
            <Title>Natural Health</Title>
            <Date>1999</Date>
      </Magazine>
      <Book>
            <Title>The First and Last Freedom</Title>
            <Author>J. Krishnamurti</Author>
            <Date>1954</Date>
            <ISBN>0-06-064831-7</ISBN>
            <Publisher>Harper &amp;amp; Row</Publisher>
      </Book>
</BookStore>
```

<BookStore> can contain any element in the substitutionGroup with Publication!

Blocking Element Substitution

- An element may wish to block other elements from substituting with it. This is achieved by adding a block attribute.

```
<xsd:element name="…" type="…" block="substitution"/>
```

Schema:

```
<xsd:element name="subway" type="xsd:string" block="substitution"/>
<xsd:element name="T" substitutionGroup="subway"/>
<xsd:element name="transportation">
   <xsd:complexType>
      <xsd:sequence>
         <xsd:element ref="subway"/>
      </xsd:sequence>
   </xsd:complexType>
</xsd:element>
```

Instance doc:

```
<transportation>
   <subway>Red Line</subway>
</transportation>
```

Not allowed!

```
<transportation>
   <T>Red Line</T>
</transportation>
```

Note: there is no error in declaring T to be substitutable with subway.
The error occurs only when you try to do substitution in the instance document.

One more Note about substitutionGroup

1. **Transitive**: if element A can substitute for element B, and element B can substitute for element C, then element A can substitute for element C.

   A --> B --> C then A --> C

2. **Non-symmetric**: if element A can substitute for element B, it is not the case that element B can substitute for element A.

Attributes

- On the next slide I show a version of the BookStore DTD that uses attributes. Then, on the following slide I show how this is implemented using XML Schemas.

```
<!ELEMENT BookStore (Book+)>
<!ELEMENT Book (Title, Author, Date, ISBN, Publisher)>
<!ATTLIST Book
        Category (autobiography | non-fiction | fiction) #REQUIRED
        InStock (true | false) "false"
        Reviewer CDATA " ">
<!ELEMENT Title (#PCDATA)>
<!ELEMENT Author (#PCDATA)>
<!ELEMENT Date (#PCDATA)>
<!ELEMENT ISBN (#PCDATA)>
<!ELEMENT Publisher (#PCDATA)>
```

BookStore.dtd

```
<xsd:element name="BookStore">
   <xsd:complexType>
      <xsd:sequence>
         <xsd:element name="Book"  maxOccurs="unbounded">
            <xsd:complexType>
               <xsd:sequence>
                  <xsd:element name="Title" type="xsd:string"/>
                  <xsd:element name="Author" type="xsd:string"/>
                  <xsd:element name="Date" type="xsd:string"/>
                  <xsd:element name="ISBN" type="xsd:string"/>
                  <xsd:element name="Publisher" type="xsd:string"/>
               </xsd:sequence>
               <xsd:attributeGroup ref="BookAttributes"/>
            </xsd:complexType>
         </xsd:element>
      </xsd:sequence>
   </xsd:complexType>
</xsd:element>
<xsd:attributeGroup name="BookAttributes">
   <xsd:attribute name="Category" use="required">
      <xsd:simpleType>
         <xsd:restriction base="xsd:string">
            <xsd:enumeration value="autobiography"/>
            <xsd:enumeration value="non-fiction"/>
            <xsd:enumeration value="fiction"/>
         </xsd:restriction>
      </xsd:simpleType>
   </xsd:attribute>
   <xsd:attribute name="InStock" type="xsd:boolean" default="false"/>
   <xsd:attribute name="Reviewer" type="xsd:string" default=" "/>
</xsd:attributeGroup>
```

Category (autobiography | non-fiction | fiction) #REQUIRED

InStock (true | false) "false"

Reviewer CDATA " "

```
<xsd:attribute name="Category"  use="required">
        <xsd:simpleType>
            <xsd:restriction base="xsd:string">
                <xsd:enumeration value="autobiography"/>
                <xsd:enumeration value="non-fiction"/>
                <xsd:enumeration value="fiction"/>
            </xsd:restriction>
        </xsd:simpleType>
</xsd:attribute>
```

"Instance documents are required to have the Category attribute (as indicated by use="required"). The value of Category must be either autobiography, non-fiction, or fiction (as specified by the enumeration facets)."

Note: attributes can only have simpleTypes (i.e., attributes cannot have child elements).

Summary of Declaring Attributes (two ways to do it)

1 <xsd:attribute name="*name*" type="*simple-type*" use="*how-its-used*" default/fixed="*value*"/>

xsd:string
xsd:integer
xsd:boolean
...

required
optional
prohibited

The "use" attribute must be optional if you use default or fixed.

2

```
<xsd:attribute name="name" use="how-its-used" default/fixed="value">
    <xsd:simpleType>
        <xsd:restriction base="simple-type">
            <xsd:facet value="value"/>
            …
        </xsd:restriction>
    </xsd:simpleType>
</xsd:attribute>
```

use --> use it only with Local Attribute Declarations

- The "use" attribute only makes sense in the context of an element declaration. Example: "for each Book element, the Category attribute is required".
- When declaring a global attribute do not specify a "use"

```
<xsd:element name="Book">
   <xsd:complexType>
      <xsd:sequence>
         …
      </xsd:sequence>
      <xsd:attribute ref="Category" use="required"/>
      …
   </xsd:complexType>
</xsd:element>
<xsd:attribute name="Category">
   <xsd:simpleType>
      <xsd:restriction base="xsd:string">
         <xsd:enumeration value="autobiography"/>
         <xsd:enumeration value="fiction"/>
         <xsd:enumeration value="non-fiction"/>
      </xsd:restriction>
   </xsd:simpleType>
</xsd:attribute>
```

Local attribute declaration. Use the "use" attribute here.

Global attribute declaration. Should NOT have a "use" ("use" only makes sense in the context of an element)

Inlining Attributes

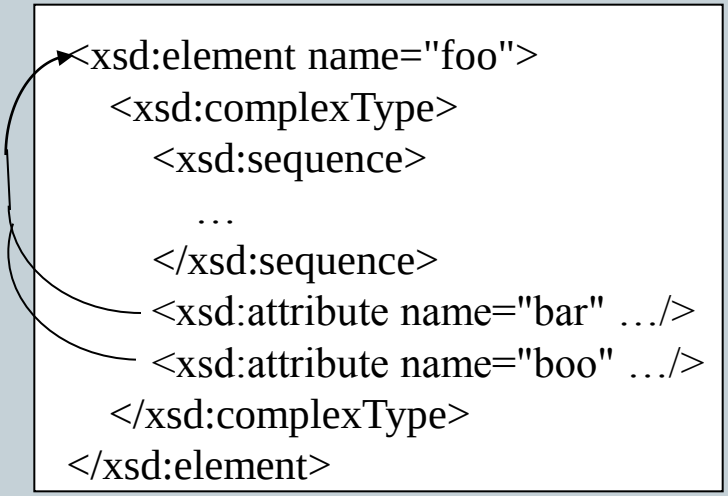
- On the next slide is another way of expressing the last example - the attributes are inlined within the Book declaration rather than being separately defined in an attributeGroup.

```
<xsd:element name="Book"  maxOccurs="unbounded">
    <xsd:complexType>
        <xsd:sequence>
            <xsd:element name="Title" type="xsd:string"/>
            <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Date" type="xsd:string"/>
            <xsd:element name="ISBN" type="xsd:string"/>
            <xsd:element name="Publisher" type="xsd:string"/>
        </xsd:sequence>
        <xsd:attribute name="Category" use="required">
            <xsd:simpleType>
                <xsd:restriction base="xsd:string">
                    <xsd:enumeration value="autobiography"/>
                    <xsd:enumeration value="non-fiction"/>
                    <xsd:enumeration value="fiction"/>
                </xsd:restriction>
            </xsd:simpleType>
        </xsd:attribute>
        <xsd:attribute name="InStock" type="xsd:boolean" default="false"/>
        <xsd:attribute name="Reviewer" type="xsd:string" default=" "/>
    </xsd:complexType>
</xsd:element>
```

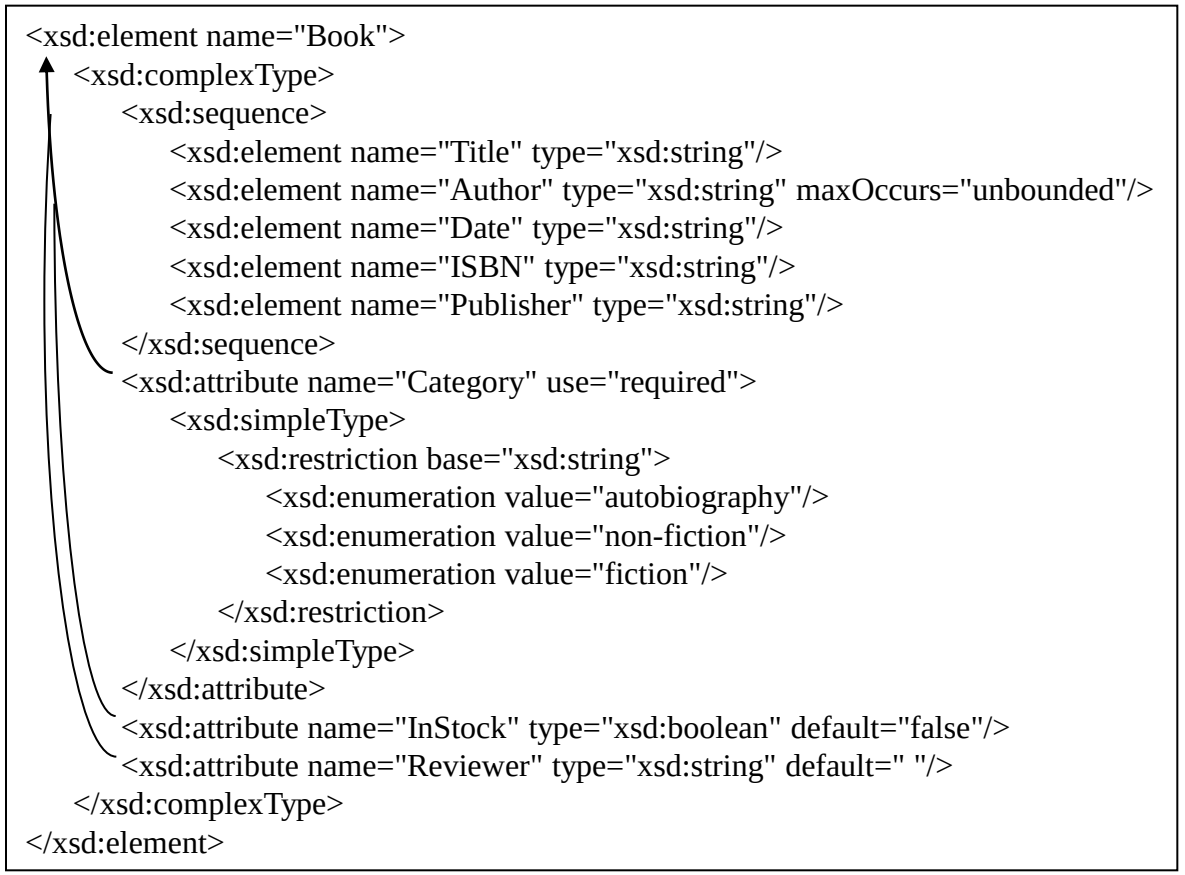
Notes about Attributes

- The attribute declarations always come last, after the element declarations.
- *The attributes are always with respect to the element that they are defined (nested) within.*

"bar and boo are attributes of foo"



These attributes apply to the element they are nested within (Book) That is, Book has three attributes - Category, InStock, and Reviewer.



Element with Simple Content and Attributes

Example. Consider this:

```
<elevation units="feet">5440</elevation>
```

The elevation element has these two constraints:

- it has a simple (integer) content
- it has an attribute called units

How do we declare elevation? (see next slide)

```
<xsd:element name="elevation">
   <xsd:complexType> (1)
      <xsd:simpleContent>(2)
         <xsd:extension base="xsd:integer">(3)
            <xsd:attribute name="units" type="xsd:string" use="required"/> (4)
         </xsd:extension>
      </xsd:simpleContent>
   </xsd:complexType>
</xsd:element>
```

1. elevation contains an attribute.
   - therefore, we must use <xsd:complexType>
2. However, elevation does not contain child elements (which is what we generally use <complexType> to indicate). Instead, elevation contains simpleContent.
3. We wish to extend the simpleContent (an integer) ...
4. with an attribute.

elevation - use Stronger Datatype

- In the declaration for elevation we allowed it to hold any integer. Further, we allowed the units attribute to hold any string.
- Let's restrict elevation to hold an integer with a range 0 - 12,000 and let's restrict units to hold either the string "feet" or the string "meters"

```
<xsd:simpleType name="elevationType">
   <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="0"/>
      <xsd:maxInclusive value="12000"/>
   </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="unitsType">
   <xsd:restriction base="xsd:string">
      <xsd:enumeration value="feet"/>
      <xsd:enumeration value="meters"/>
   </xsd:restriction>
</xsd:simpleType>
<xsd:element name="elevation">
   <xsd:complexType>
      <xsd:simpleContent>
         <xsd:extension base="elevationType">
            <xsd:attribute name="units" type="unitsType" use="required"/>
         </xsd:extension>
      </xsd:simpleContent>
   </xsd:complexType>
</xsd:element>
```

Summary of Declaring Elements

1. **Element with Simple Content.**

Declaring an element using a **built-in type**:

```
<xsd:element name="numStudents" type="xsd:positiveInteger"/>
```

Declaring an element using a **user-defined simpleType**:

```
<xsd:simpleType name="shapes">
        <xsd:restriction base="xsd:string">
           <xsd:enumeration value="triangle"/>
           <xsd:enumeration value="rectangle"/>
           <xsd:enumeration value="square"/>
        </xsd:restriction>
</xsd:simpleType>
<xsd:element name="geometry" type="shapes"/>
```

An alternative formulation of the above shapes example is to **inline the simpleType** definition:

```
<xsd:element name="geometry">
        <xsd:simpleType>
           <xsd:restriction base="xsd:string">
              <xsd:enumeration value="triangle"/>
              <xsd:enumeration value="rectangle"/>
              <xsd:enumeration value="square"/>
           </xsd:restriction>
        </xsd:simpleType>
 </xsd:element>
```

Summary of Declaring Elements (cont.)

2. **Element Contains Child Elements**

Defining the child elements **inline**:

```
<xsd:element name="Person">
   <xsd:complexType>
      <xsd:sequence>
         <xsd:element name="Title" type="xsd:string"/>
         <xsd:element name="FirstName" type="xsd:string"/>
         <xsd:element name="Surname" type="xsd:string"/>
      </xsd:sequence>
   </xsd:complexType>
</xsd:element>
```

An alternate formulation of the above Person example is to create a **named complexType** and then use that type:

```
<xsd:complexType name="PersonType">
      <xsd:sequence>
         <xsd:element name="Title" type="xsd:string"/>
         <xsd:element name="FirstName" type="xsd:string"/>
         <xsd:element name="Surname" type="xsd:string"/>
      </xsd:sequence>
</xsd:complexType>
<xsd:element name="Person" type="PersonType"/>
```

Summary of Declaring Elements (cont.)

3. **Element Contains a complexType that is an Extension of another complexType**

```
<xsd:complexType name="Publication">
   <xsd:sequence>
      <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
      <xsd:element name="Date" type="xsd:gYear"/>
   </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="BookPublication">
   <xsd:complexContent>
      <xsd:extension base=" Publication " >
         <xsd:sequence>
            <xsd:element name="ISBN" type="xsd:string"/>
            <xsd:element name="Publisher" type="xsd:string"/>
         </xsd:sequence>
      </xsd:extension>
   </xsd:complexContent>
</xsd:complexType>
<xsd:element name="Book" type="BookPublication"/>
```

Summary of Declaring Elements (cont.)

4. **Element Contains a complexType that is a Restriction of another complexType**

```
<xsd:complexType name="Publication">
      <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Author" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Date" type="xsd:gYear"/>
      </xsd:sequence>
</xsd:complexType>
<xsd:complexType name= "SingleAuthorPublication">
   <xsd:complexContent>
      <xsd:restriction base=" Publication ">
         <xsd:sequence>
            <xsd:element name="Title" type="xsd:string" maxOccurs="unbounded"/>
            <xsd:element name="Author" type="xsd:string"/>
            <xsd:element name="Date" type="xsd:gYear"/>
         </xsd:sequence>
      </xsd:restriction>
   </xsd:complexContent>
</xsd:complexType>
<xsd:element name="Catalogue" type="SingleAuthorPublication"/>
```

Summary of Declaring Elements (concluded)

5. **Element Contains Simple Content and Attributes**

```
<xsd:element name="apple">
  <xsd:complexType>
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attribute name="variety" type="xsd:string" use="required"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>
</xsd:element>
```

Example. <apple variety="Cortland">Large, green, sour</apple>

complexContent versus simpleContent

- With complexContent you *extend or restrict* a complexType
- With simpleContent you *extend or restrict* a simpleType

```
<xsd:complexType name="…">
   <xsd:complexContent>
      <extension base="X">
         …
      </extension>
   </xsd:complexContent>
</xsd:complexType>
```

X must be a complexType

versus

```
<xsd:complexType name="…">
   <xsd:simpleContent>
      <extension base="Y">
         …
      </extension>
   </xsd:simpleContent>
</xsd:complexType>
```

Y must be a simpleType

Expressing Alternates

DTD:

```
<!ELEMENT transportation (train | plane | automobile)>
```

XML Schema:

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.travel.org"
            xmlns="http://www.travel.org"
            elementFormDefault="qualified">
  <xsd:element name="transportation">
    <xsd:complexType>
      <xsd:choice>
        <xsd:element name="train" type="xsd:string"/>
        <xsd:element name="plane" type="xsd:string"/>
        <xsd:element name="automobile" type="xsd:string"/>
      </xsd:choice>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Note: the choice is an exclusive-or, that is, transportation can contain only **one** element - either train, or plane, or automobile.

Expressing Repeatable Choice

DTD:

```
<!ELEMENT binary-string (zero | one)*>
```

XML Schema:

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.binary.org"
            xmlns="http://www.binary.org"
            elementFormDefault="qualified">
  <xsd:element name="binary-string">
    <xsd:complexType>
      <xsd:choice minOccurs="0" maxOccurs="unbounded">
        <xsd:element name="zero" type="xsd:unsignedByte" fixed="0"/>
        <xsd:element name="one" type="xsd:unsignedByte" fixed="1"/>
      </xsd:choice>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Notes:

1. An element can fix its value, using the fixed attribute.
2. When you don't specify a value for minOccurs, it defaults to "1". Same for maxOccurs. See the last example (transportation) where we used a <choice> element with no minOccurs or maxOccurs.

fixed/default Element Values

- When you declare an element you can give it a fixed or default value.
  - Then, in the instance document, you can leave the element empty.

```
<element name="zero" fixed="0"/>
…
<zero>0</zero>

    or equivalently:

<zero/>
```

```
<element name="color" default="red"/>
…
<color>red</color>

    or equivalently:

<color/>
```

Using <sequence> and <choice>

DTD:

```
<!ELEMENT life ((work, eat)*, (work | play), sleep)* >
```

XML Schema:

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                targetNamespace="http://www.life.org"
                xmlns="http://www.life.org"
                elementFormDefault="qualified">
  <xsd:element name="life">
    <xsd:complexType>
      <xsd:sequence minOccurs="0" maxOccurs="unbounded">
        <xsd:sequence minOccurs="0" maxOccurs="unbounded">
          <xsd:element name="work" type="xsd:string"/>
          <xsd:element name="eat" type="xsd:string"/>
        </xsd: sequence>
        <xsd:choice>
          <xsd:element name="work" type="xsd:string"/>
          <xsd:element name="play" type="xsd:string"/>
        </xsd:choice>
        <xsd:element name="sleep" type="xsd:string"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Expressing Any Order

Problem: create an element, Book, which contains Author, Title, Date, ISBN, and Publisher, *in any order* (Note: this is very difficult and ugly with DTDs).

XML Schema:

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                targetNamespace="http://www.books.org"
                xmlns="http://www.books.org"
                elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book"  maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:all>
              <xsd:element name="Title" type="xsd:string"/>
              <xsd:element name="Author" type="xsd:string"/>
              <xsd:element name="Date" type="xsd:string"/>
              <xsd:element name="ISBN" type="xsd:string"/>
              <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:all>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
```

<all> means that Book must contain all five child elements, but they may occur in any order.

Constraints on using <all>

- Elements declared within <all> must have a maxOccurs value of "1" (minOccurs can be either "0" or "1")
- If a complexType uses <all> and it extends another type, then that parent type must have empty content.
- The <all> element cannot be nested within either <sequence>, <choice>, or another <all>
- The contents of <all> must be just elements. It cannot contain <sequence> or <choice>

Empty Element

DTD:

```
<!ELEMENT image EMPTY>
<!ATTLIST image href CDATA #REQUIRED>
```

Schema:

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.photography.org"
            xmlns="http://www.photography.org"
            elementFormDefault="qualified">
  <xsd:element name="gallery">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="image" maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:attribute name="href" type="xsd:anyURI" use="required"/>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Instance doc (snippet):

```
<image href="http://www.xfront.com/InSubway.gif"/>
```

No targetNamespace (noNamespaceSchemaLocation)

- Sometimes you may wish to create a schema but without associating the elements with a namespace.
- The targetNamespace attribute is actually an optional attribute of <schema>. Thus, if you don't want to specify a namespace for your schema then simply don't use the targetNamespace attribute.
- Consequences of having no namespace
  - 1. In the instance document don't namespace qualify the elements.
  - 2. In the instance document, instead of using schemaLocation use noNamespaceSchemaLocation.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                elementFormDefault="qualified">
   <xsd:element name="BookStore">
      <xsd:complexType>
         <xsd:sequence>
            <xsd:element ref="Book" minOccurs="0" maxOccurs="unbounded"/>
         </xsd:sequence>
      </xsd:complexType>
   </xsd:element>
   <xsd:element name="Book">
      <xsd:complexType>
         <xsd:sequence>
            <xsd:element ref="Title"/>
            <xsd:element ref="Author"/>
            <xsd:element ref="Date"/>
            <xsd:element ref="ISBN"/>
            <xsd:element ref="Publisher"/>
         </xsd:sequence>
      </xsd:complexType>
   </xsd:element>
   <xsd:element name="Title" type="xsd:string"/>
   <xsd:element name="Author" type="xsd:string"/>
   <xsd:element name="Date" type="xsd:string"/>
   <xsd:element name="ISBN" type="xsd:string"/>
   <xsd:element name="Publisher" type="xsd:string"/>
</xsd:schema>
```

Note that there is no targetNamespace attribute, and note that there is no longer a default namespace.

```
<?xml version="1.0"?>
<BookStore xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                xsi:noNamespaceSchemaLocation= "BookStore.xsd">
      <Book>
            <Title>My Life and Times</Title>
            <Author>Paul McCartney</Author>
            <Date>1998</Date>
            <ISBN>1-56592-235-2</ISBN>
            <Publisher>McMillin Publishing</Publisher>
       </Book>
       …
</BookStore>
```

1. Note that there is no default namespace declaration. So, none of the elements are associated with a namespace.
2. Note that we do not use xsi:schemaLocation (since it requires a pair of values - a namespace and a URL to the schema for that namespace). Instead, we use xsi:noNamespaceSchemaLocation.

Assembling an Instance Document from Multiple Schema Documents

- An instance document may be composed of elements from multiple schemas.
- Validation can apply to the entire XML instance document, or to a single element.

Library.xml

```
<?xml version="1.0"?>
<Library xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation=
               "http://www.book.org
               Book.xsd
               http://www.employee.org
               Employee.xsd">
  <Books>
    <Book xmlns="http://www.book.org">
        <Title>My Life and Times</Title>
        <Author>Paul McCartney</Author>
        <Date>1998</Date>
        <ISBN>1-56592-235-2</ISBN>
        <Publisher>Macmillan Publishing</Publisher>
    </Book>
    <Book xmlns="http://www.book.org">
        <Title>Illusions: The Adventures of a Reluctant Messiah</Title>
        <Author>Richard Bach</Author>
        <Date>1977</Date>
        <ISBN>0-440-34319-4</ISBN>
        <Publisher>Dell Publishing Co.</Publisher>
    </Book>
    <Book xmlns="http://www.book.org">
        <Title>The First and Last Freedom</Title>
        <Author>J. Krishnamurti</Author>
        <Date>1954</Date>
        <ISBN>0-06-064831-7</ISBN>
        <Publisher>Harper &amp;amp; Row</Publisher>
    </Book>
  </Books>
  <Employees>
    <Employee xmlns="http://www.employee.org">
      <Name>John Doe</Name>
      <SSN>123-45-6789</SSN>
    </Employee>
    <Employee xmlns="http://www.employee.org">
      <Name>Sally Smith</Name>
      <SSN>000-11-2345</SSN>
    </Employee>
  </Employees>
</Library>
```

Validating against two schemas

The <Book> elements are defined in Book.xsd, and the <Employee> elements are defined in Employee.xsd. The <Library>, <Books>, and <Employees> elements are not defined in any schema!

1. A schema validator will validate each Book element against Book.xsd.
2. It will validate each Employee element against Employee.xsd.
3. It will not validate the other elements.

Summary of Declaring simpleTypes

1. **simpleType that uses a built-in base type:**

```
<xsd:simpleType name= "EarthSurfaceElevation">
   <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="-1290"/>
      <xsd:maxInclusive value="29035"/>
   </xsd:restriction>
</xsd:simpleType>
```

2. **simpleType that uses another simpleType as the base type:**

```
<xsd:simpleType name= "BostonSurfaceElevation">
   <xsd:restriction base="EarthSurfaceElevation">
      <xsd:minInclusive value="0"/>
      <xsd:maxInclusive value="120"/>
   </xsd:restriction>
</xsd:simpleType>
```

Summary of Declaring simpleTypes

3. **simpleType that declares a list type:**

```
<xsd:simpleType name= "LotteryNumbers">
   <xsd:list itemType="OneToNinetyNine"/>
</xsd:simpleType>
```

where the datatype OneToNinetyNine is declared as:

```
<xsd:simpleType name= "OneToNinetyNine">
   <xsd:restriction base="xsd:nonNegativeInteger">
      <xsd:maxInclusive value="99"/>
   </xsd:restriction>
</xsd:simpleType>
```

4. **An alternate form of the above, where the list's datatype is specified using an inlined simpleType:**

```
<xsd:simpleType name= "LotteryNumbers">
   <xsd:list>
      <xsd:simpleType>
         <xsd:restriction base="xsd:nonNegativeInteger">
            <xsd:maxInclusive value="99"/>
         </xsd:restriction>
      </xsd:simpleType>
   </xsd:list>
</xsd:simpleType>
```

Summary of Declaring simpleTypes

5. **simpleType that declares a union type:**

```
<xsd:simpleType name= "maxOccurs">
   <xsd:union memberTypes="xsd:positiveInteger  UnboundedType"/>
</xsd:simpleType>
```

where the datatype UnboundedType is declared as:

```
<xsd:simpleType name= "UnboundedType">
   <xsd:restriction base="xsd:string">
       <xsd:enumeration value="unbounded"/>
   </xsd:restriction>
</xsd:simpleType>
```

6. **An alternate form of the above, where the datatype UnboundedType is specified using an inline simpleType:**

```
<xsd:simpleType name= "maxOccurs">
   <xsd:union memberTypes="xsd:positiveInteger">
      <xsd:simpleType>
          <xsd:restriction base="xsd:string">
              <xsd:enumeration value="unbounded"/>
          </xsd:restriction>
      </xsd:simpleType>
   </xsd:union>
</xsd:simpleType>
```

any Element

- The <any> element enables the instance document author to extend his/her document with elements not specified by the schema.

```
<xsd:element name="Book">
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element name="Title" type="xsd:string"/>
                        <xsd:element name="Author" type="xsd:string"/>
                        <xsd:element name="Date" type="xsd:string"/>
                        <xsd:element name="ISBN" type="xsd:string"/>
                        <xsd:element name="Publisher" type="xsd:string"/>
                        <xsd:any  minOccurs="0"/>
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>
```

Now an instance document author can optionally extend (after <Publisher>) the content of <Book> elements with any element.

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
                targetNamespace="http://www.repository.org"
                xmlns="http://www.repository.org"
                elementFormDefault="qualified">
   <xsd:element name="Reviewer">
      <xsd:complexType>
         <xsd:sequence>
            <xsd:element name="Name">
               <xsd:complexType>
                  <xsd:sequence>
                     <xsd:element name="First" type="xsd:string"/>
                     <xsd:element name="Last" type="xsd:string"/>
                  </xsd:sequence>
               </xsd:complexType>
            </xsd:element>
         </xsd:sequence>
      </xsd:complexType>
   </xsd:element>
</xsd:schema>
```

SchemaRepository.xsd (see example23)

Suppose that the instance document author discovers this schema repository, and wants to extend his/her <Book> elements with a <Reviewer> element. He/she can do so! Thus, the instance document will be extended with an element never anticipated by the schema author. Wow!

```
<?xml version="1.0"?>
<BookStore xmlns="http://www.BookRetailers.org"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation=
                        "http://www. BookRetailers.org
                         BookSeller.xsd
                         http://www. repository.org
                         SchemaRepository.xsd">
    <Book>
        <Title>My Life and Times</Title>
        <Author>Paul McCartney</Author>
        <Date>1998</Date>
        <ISBN>94303-12021-43892</ISBN>
        <Publisher>McMillin Publishing</Publisher>
        <Reviewer xmlns="http://www.repository.org">
          <Name>
            <First>Roger</First>
            <Last>Costello</Last>
          </Name>
        </Reviewer>
    </Book>
    <Book>
        <Title>Illusions: The Adventures of a Reluctant Messiah</Title>
        <Author>Richard Bach</Author>
        <Date>1977</Date>
        <ISBN>0-440-34319-4</ISBN>
        <Publisher>Dell Publishing Co.</Publisher>
    </Book>
</BookStore>
```

This instance document uses components from two different schemas.

Extensible Instance Documents

- The <any> element enables instance document authors to create instance documents containing elements above and beyond what was specified by the schema. The instance documents are said to be *extensible*. Contrast this schema with previous schemas where the content of all our elements were always fixed and *static*.
- We are empowering the instance document author with the ability to define what data makes sense to him/her!

Schema Validators

- Command Line Only
  - XSV by Henry Thompson
    - ftp://ftp.cogsci.ed.ac.uk/pub/XSV/XSV12.EXE
- Has a Programmatic API
  - xerces by Apache
    - http://xerces.apache.org/xerces-j/
  - IBM Schema Quality Checker (Note: this tool is only used to check your schema. It cannot be used to validate an instance document against a schema.)
    - http://www.alphaworks.ibm.com/tech/xmlsqc
  - XML Services (MSXML) 6.0
    - http://www.microsoft.com
- GUI Oriented
  - XML Spy
    - http://www.altova.com/xml-editor/