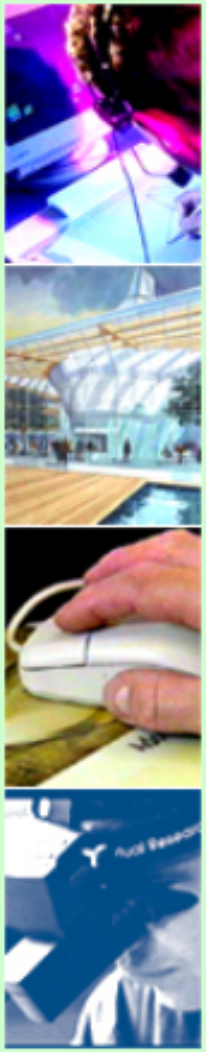
ISO 9001

XML Syntax
Writing XML and Designing DTD's

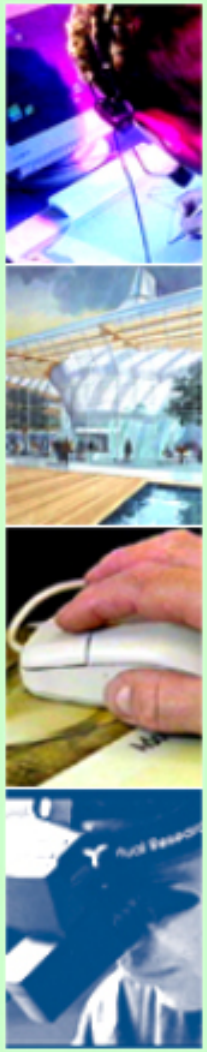


XML Document

```
<booklist title="Some XML Books">
 <book>
    <author>
      <name>St. Laurent</name>
         <initial>S</initial>
    </author>
    <date>1998</date>
    <title edition="Second">XML: A Primer</title>
    <publisher>MIS Press</publisher>
    <website href="http://www.simonstl.com/xmlprim/" />
    <rating stars="4"/>
  </book>
</booklist>
```

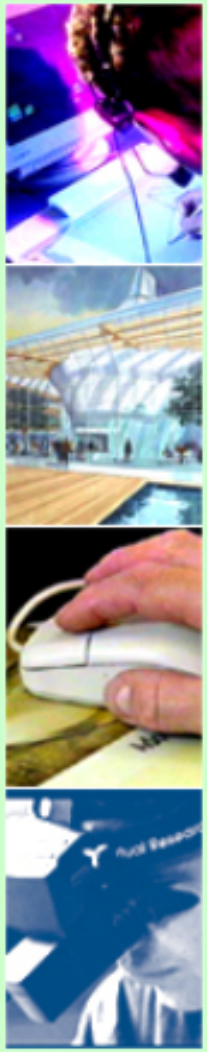
Character Data Declaration

- For occasions when text must contain uninterpreted markup characters,use CDATA sections:
  - Press <<<ENTER>>>
  - <![CDATA[Press <<<ENTER>>>]]>



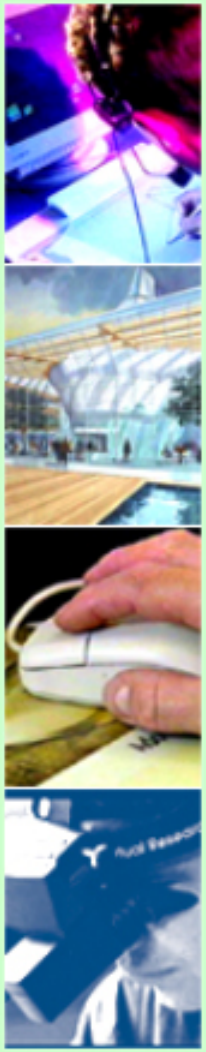
What is a DTD?

- A template for document markup :
  - A file which contains a formal definition of a particular type of document.
  - It's file that constrains or restricts certain elements and attributes to exist in XML document.
- A DTD describes :
  - What names can be used for element types
  - Where & how element types can occur
  - Specifies document hierarchy
  - Specifies names and types of element attributes



Why have a DTD?

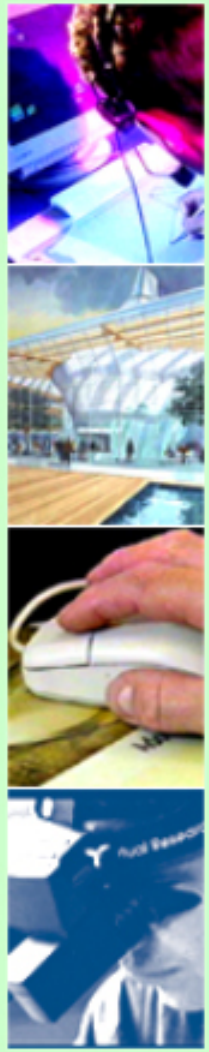
- Validating XML parser can check the structure of the XML file against a DTD and check that it is valid.
- DTD can be a mechanism for standardization and hence document/data manipulation and exchange.
- DTD declares all allowed components of XML document as well as the structural relationships among those components.



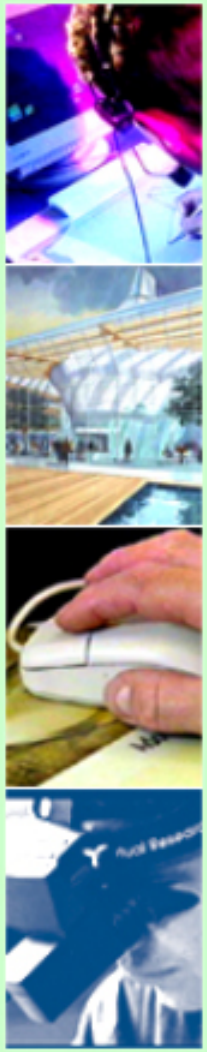
Internal DTD Definition

- Include in the DOCTYPE declaration

```
<?xml version="1.0"?>
<!DOCTYPE root _ elem [
    <!-- DTD appears here -->
    <! ... >
    <! ... >
  ]>
  <!-- Rest of XML file -->
```



External DTD Definition



- Document type declaration :links the DTD file with the XML document.

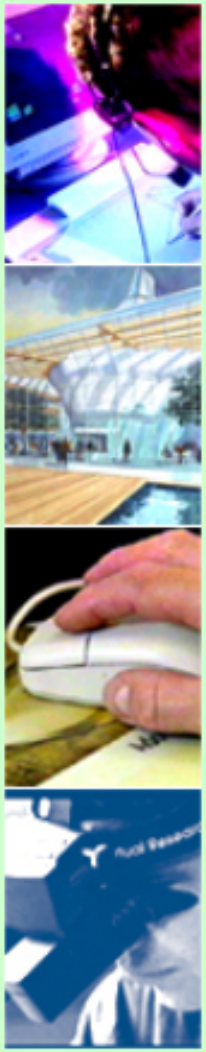
```
<!DOCTYPE rootelem SYSTEM “file.dtd”>
```

- Reference external DTD file as pathname in DOCTYPE declaration

```
<!DOCTYPE MyDoc SYSTEM “fullPath">
```

Designing a DTD

- Identify features of the data that need markup
- For each feature, determine
  - Can it be given a name
  - Does it always appear
  - May there be more than one
  - Does it deconstruct to smaller features
  - Is some of the textual content always the same
  - How is it associated with other features



Element Declarations

- Used to define new elements and their content

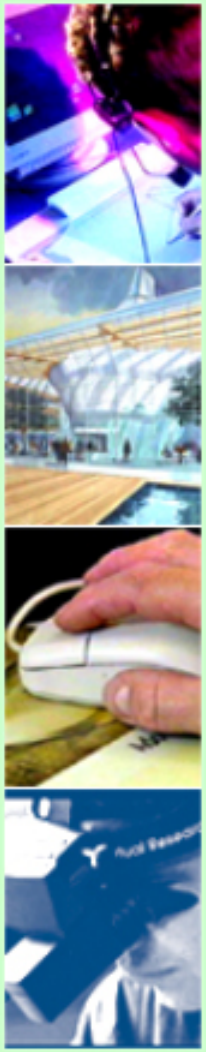
  <!ELEMENT Title (#PCDATA)>

  ⇒ <Title>Company Name </Title>
- Empty element has no content

  <!ELEMENT name EMPTY> ⇒ <name/>
- When children are allowed, use model groups

  <!ELEMENT person (name, e-mail)>
- When declare element the can contain any type of content

  <!ELEMENT DOCUMENT ANY>

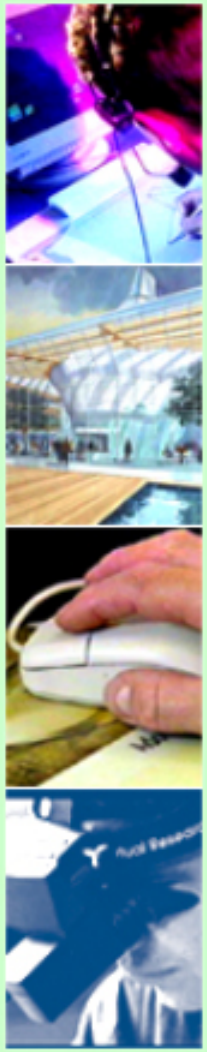


Model Group Quantity Indicators

- Describe constraints on elements in DTD:

A?	May occur [0..1]
A+	Must occur [1..*]
A*	May occur [0..*]
A | B	Either A or B
A, B	A followed by B

- <!Element father (daughter|son)>

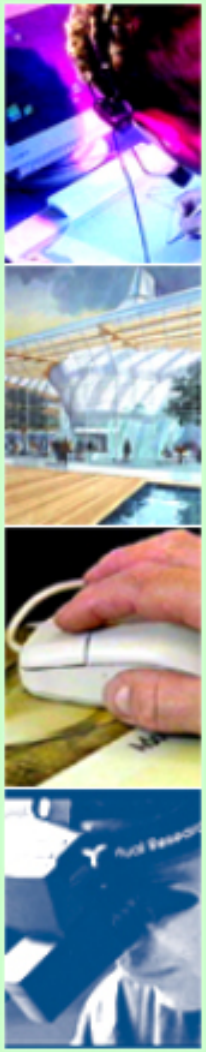


Mixed content Declarations

- Used to define Mixed content elements:

```
<!ELEMENT father (#PCDATA|job|age)*>
```

⇒
```
<father>Ibrahim
    <job>engineer</job>
    <age>55</age>
</father>
```



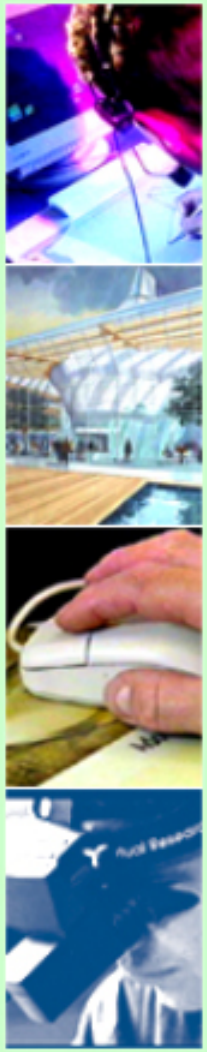
Mixed content Declarations (cont.)

When using mixed content :

- You can't specify the number of occurrences or the order of appearance of the child elements .
- So Mixed Content is to kept as minimum as possible

Attribute Declarations

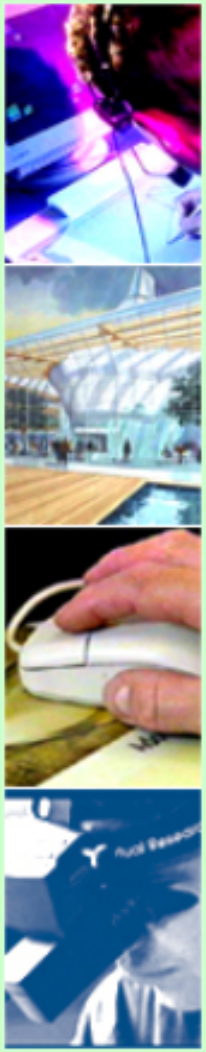
- Attributes can be attached to elements.
- Declared separately in an “ATTLIST” declaration
- Attribute name
  - <!ATTLIST tag name type default>
  - <!ATTLIST tag first_attr …
    secon_attr …
    third_attr … >
- Attribute types include: CDATA, enumeration,ID



Attribute types

- **CDATA :** simple character data .
- **Enumeration :** list of values

```
<!ATTLIST son sport( football | tennis)
                        "tennis">
<!ATTLIST son job CDATA #IMPLIED>
<!ATTLIST son age CDATA #REQUIRED>
```



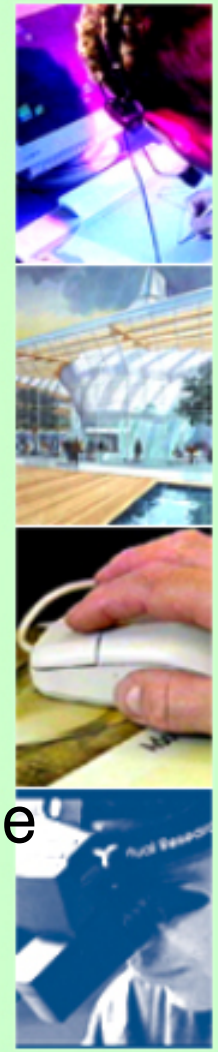
Attribute types (cont.)

ID :

- Applications can use the ID value of elements to uniquely identify those elements ,so they declare attribute type ID (Note attribute doesn't have to be named id).
- Attribute of type ID should have to be unique contain only characters permitted for NMTOKEN and must start with a letter. (ex: C101)

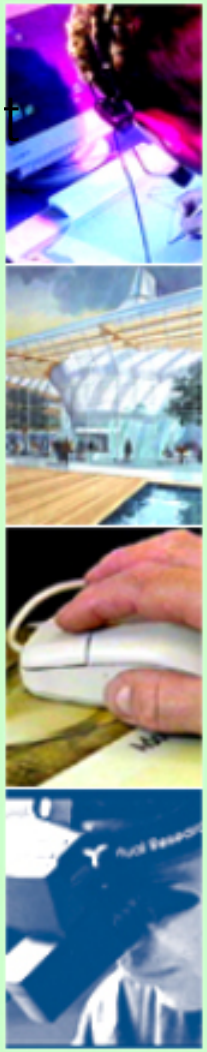
- No element type may have more than one ID, and the value of attribute defined as ID should be unique

<!ATTLIST invoice num ID #REQUIRED>



Default Attribute Values

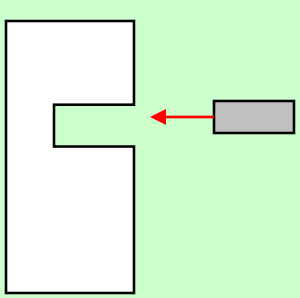
- Can specify a default attribute value for when its missing from XML document, or state that value must be entered
  - #REQUIRED — Must be specified
  - #IMPLIED — May be specified
  - "default" — Default value if unspecified
  - #FIXED — Only one value allowed
- <!ATTLIST son sports (football|basketball) “football”>
- <!ATTLIST son age CDATA #REQUIRED>

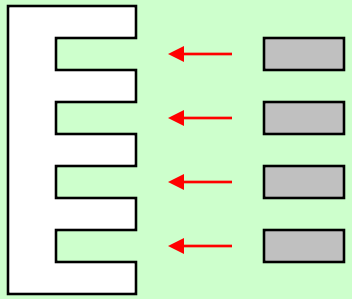


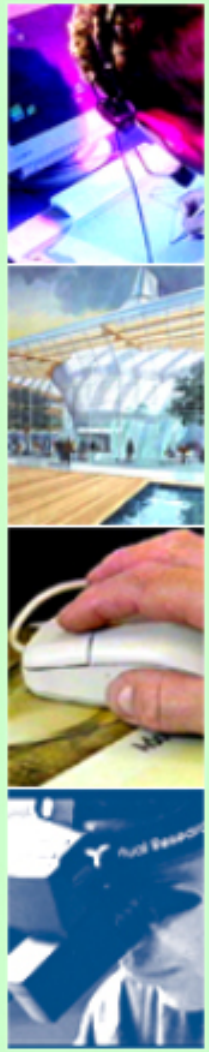
ISO 9001

Entities

- XML document may be distributed among units of information:
  - Each unit of information is called an **entity**
  - Each entity has a name to identify it
  - Defined using an entity declaration
  - Used by calling an entity reference







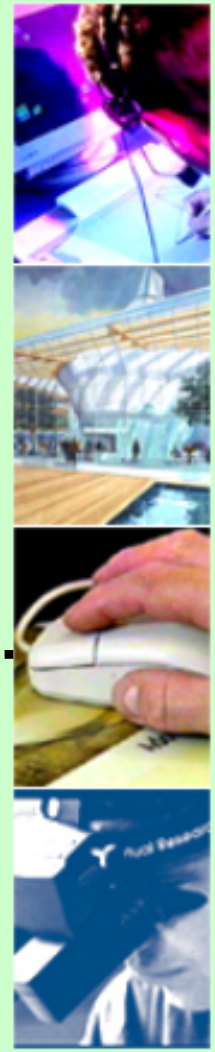
General Entities

- Declared in 'Document Type Definition'
  - <!ENTITY name "*replacement text*">

<!ENTITY xml "eXtensible Markup Language">

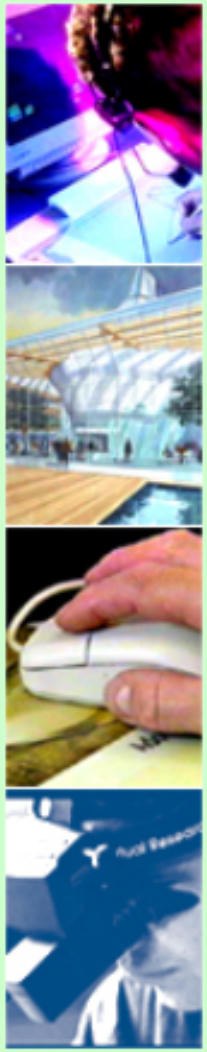
<!ENTITY copyright "©">

The &xml; and &copyright; includes entities.



Restrictions on Entities

- General text entities
  - Can appear in element content
    - <para> … &ent; … </para>
  - Can appear in attribute value
    - <para name="&ent;"> … </para>
  - Can appear in internal entity content
    - <!ENTITY cod "&ent;">



XML Document

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ITI SYSTEM "ITIhierarchy.dtd">
<ITI>
 <ITI-manager>Dr.Heba Saleh</ITI-manager>
 <ITI.vice-manager>Eng.Amr El Shafey
     <ITI.ScientificDept mgr="Eng.Sherif Sharaf">Scientific Departments
          <platform mgr="Eng. Hany Safwat">System Development</platform>
          <platform mgr="Eng. Asmaa">Java</platform>
          <platform mgr="Eng. Shahinaz">ERP</platform>
          platform mgr="Eng. Rania">UNIX</platform>
    </ITI.ScientificDept>
    <ITI.SpecializedTracks name="Specialized Tracks">
          <ProgramManager>Eng.Bahaa</ProgramManager>
          <Track mgr="Eng.Dalia Rasmy">Short Training </Track>
          <Track>Community Development Program</Track>
    </ITI.SpecializedTracks>
 </ITI.vice-manager>
</ITI>
```

DTD Document

```
<!ELEMENT ITI (ITI-manager, ITI.vice-manager)>
<!ELEMENT ITI-manager (#PCDATA)>
<!ELEMENT ITI.vice-manager (#PCDATA | ITI.ScientificDept | ITI.SpecializedTracks)*>
<!ELEMENT ITI.ScientificDept (#PCDATA | platform)*>
<!ELEMENT platform (#PCDATA)>
<!ELEMENT ITI.SpecializedTracks (ProgramManager?, Track+)>
<!ELEMENT ProgramManager (#PCDATA)>
<!ELEMENT Track (#PCDATA)>
<!ATTLIST platform mgr CDATA #IMPLIED>
<!ATTLIST ITI.ScientificDept mgr CDATA #REQUIRED>

<!ATTLIST ITI.SpecializedTracks  name CDATA #FIXED "Specialized Tracks"
>
<!ATTLIST Track mgr CDATA #IMPLIED
>
```

ISO 9001

Disadv. Of DTDs

- DTD has its own syntax.
- Precise number of element repetitions can't be achieved.
- Limited number of data types.
- Only 1 DTD can be referenced from within the XML document.