#### XML Schema Nguyễn Hồng Phương Email: phuong.nguyenhong@hust.edu.vn Site: http://is.hut.edu.vn/~phuongnh Bộ môn Hệ thống thông tin Viện Công nghệ thông tin và Truyền thông Đại học Bách Khoa Hà Nội

Nội dung
☐ Giới thiệu
☐ Các kiểu đơn giản
☐ Các kiểu phức tạp
☐ Kiểu dữ liệu

1. Giới thiệu
Là một giải pháp thay thế cho DTD
Một XML Schema mô tả cấu trúc một tài liệu XML
Ngôn ngữ lược đồ XML được biết đến là XML Schema Definition (XSD)

Một lược đồ XML
 định nghĩa các phần tử có thể xuất hiện trong tài liệu.
 định nghĩa các thuộc tính có thể xuất hiện trong tài liệu.
 định nghĩa phần tử con
 định nghĩa thứ tự các phần tử con
 định nghĩa số lượng các phần tử con
 định nghĩa một phần tử là rỗng hay chứa text
 định nghĩa kiểu dữ liệu cho phần tử/thuộc tính
 định nghĩa giá trị mặc định cho phần tử/thuộc tính

 ■ XML Schema là một sự thay thế cho DTD?

 Dễ mở rộng
 Giàu tính biểu đạt hơn DTD
 Được viết trong XML
 Hỗ trợ các kiểu dữ liệu
 Hỗ trợ các không gian tên

 Là khuyến cáo của W3C

XML Schema hỗ trợ các kiểu dữ liệu

☐ Một trong những điểm mạnh của XML Schema là hỗ trợ các kiểu dữ liệu

☐ mô tả nội dung tài liệu được phép

☐ hợp thức hóa tính đúng đắn của dữ liệu

☐ làm việc với dữ liệu của một CSDL

☐ xác định các khía cạnh dữ liệu (các giới hạn)

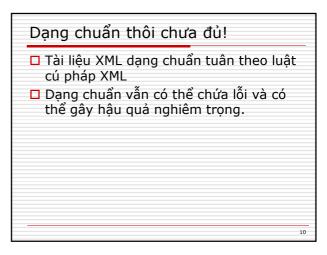
☐ xác định các mẫu dữ liệu (format dữ liệu)

☐ chuyển đổi dữ liệu giữa các kiểu dữ liệu khác nhau

#### XML Schema sử dụng cú pháp XML □ Lược đồ XML được viết trong XML □ Lợi ích: ■ Không phải học ngôn ngữ mới ■ Có thể soạn thảo file schema bằng bộ soạn thảo XML ■ Có thể sử dụng XML parser để parse file schema ■ Có thể thao tác lược đồ với XML DOM ■ Có thể chuyển đổi lược đồ với XSLT

### XML Schema truyền dữ liệu thống nhất Kiểu ngày "03-02-2013" 03 Feb 2013 02 March 2013 Thống nhất bên gửi và bên nhận <date type="date">2013-02-03</date> Kiểu dữ liệu date yêu cầu định dạng yyyymm-dd

# Lược đồ XML có thể mở rộng Vì được viết trong XML nên có thể mở rộng Với một định nghĩa lược đồ có thể mở rộng, chúng ta có thể: Tái sử dụng lược đồ trong các lược đồ khác Tạo ra kiểu dữ liệu riêng dẫn xuất từ kiểu chuẩn Tham chiếu nhiều lược đồ trong cùng tài liệu



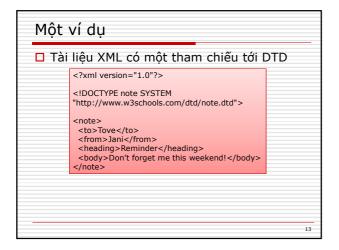
```
Một ví dụ

Quan sát lại file note.xml

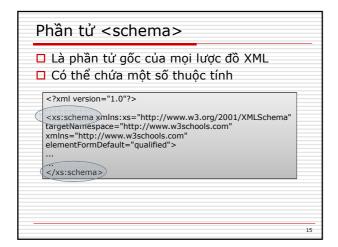
<pr
```

```
Một ví dụ

| File note.xsd
| <?xml version="1.0"?>
| <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" targetNamespace="http://www.w3schools.com" xmlns="http://www.w3schools.com" elementFormDefault="qualified">
| <xs:element name="note">
| <xs:complexType>
| <xs:sequence>
| <xs:element name="to" type="xs:string"/>
| <xs:element name="from" type="xs:string"/>
| <xs:element name="heading" type="xs:string"/>
| <xs:element name="body" type="xs:string"/>
| <xs:element name="body" type="xs:string"/>
| </xs:complexType>
| </xs:complexType>
| </xs:complexType>
| </xs:schema>
```







□ Đoạn xmlns:xs=http://www.w3.org/2001/XMLSchema chỉ ra rằng các phần tử và kiểu dữ liệu sử dụng trong lược đồ lấy từ không gian tên http://www.w3.org/2001/XMLSchema (chúng nên gắn kèm tên xs)
□ targetNamespace=http://www.w3schools.com chỉ ra rằng các phần tử xác định bởi lược đồ này (note, to, from, heading, body) lấy từ không gian tên http://www.w3schools.com □ xmlns=http://www.w3schools.com : không gian tên mặc định

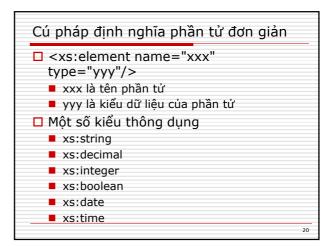
Tham khảo tới một lược đồ trong tài liệu XML:

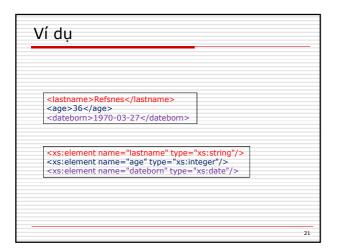
<note xmlns="http://www.w3schools.com"</p>
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3schools.com note.xsd">
<to>Tooe</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>

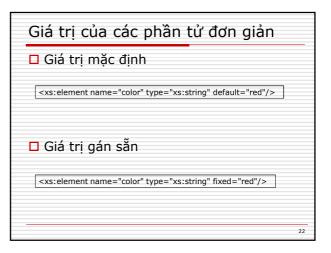
2. Các kiểu đơn giản

Phần tử XSD
Thuộc tính XSD
Giới hạn XSD

### 2.1. Phần tử đơn giản XSD ☐ Phần tử đơn giản là phần tử chỉ chứa text, không chứa phần tử hay thuộc tính khác. ☐ Text có thể là nhiều kiểu khác nhau: boolean, string, date,...







2.2. Thuộc tính XSD

Nếu một phần tử có các thuộc tính, thì phần tử đó có kiểu phức tạp.

Cú pháp định nghĩa thuộc tính

(xs:attribute name="xxx" type="yyy"/>

Một số kiểu thông dụng

xs:string

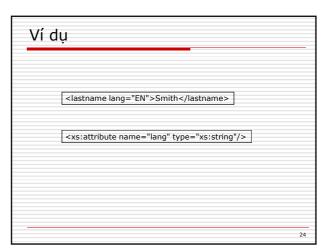
xs:decimal

xs:integer

xs:boolean

xs:date

xs:time



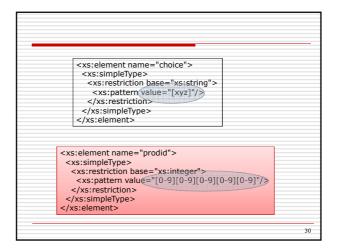


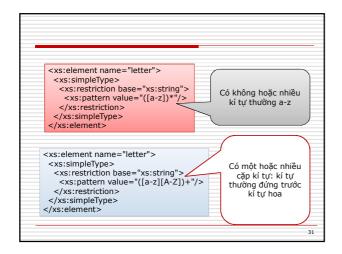
```
Giới hạn trên tập các giá trị

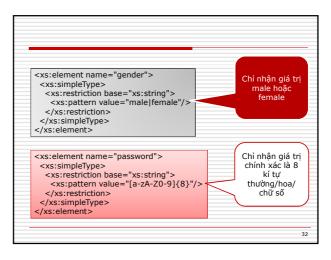
các giá trị có thể chấp nhận: Audi, Golf, BMW

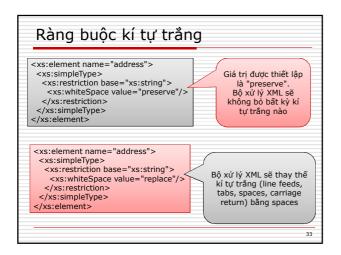
<xs:slement name="car">
<xs:simpleType>
<xs:enumeration value="Audi"/>
<xs:enumeration value="Golf"/>
<xs:enumeration value="BMW"/>
</xs:restriction>
</xs:simpleType>
</xs:element name="car" type="carType"/>
<xs:enumeration value="Golf"/>
<xs:enumeration value="dolf"/>
<xs:enumeration value="audi"/>
<xs:enumeration value="dolf"/>
<xs:enumeration value="dolf"/>
<xs:enumeration value="dolf"/>
<xs:enumeration value="BMW"/>
<xs:enumeration value="BMW"/>
</xs:enumeration value="BMW"/>
</xs:restriction>
</xs:simpleType>
```

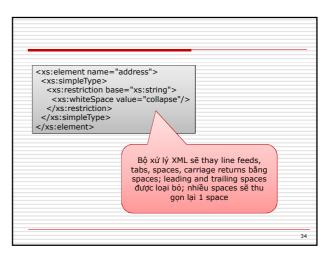
```
phần tử
                            <xs:element name="initials">
                             <xs:simpleType>
<xs:restriction base="xs:string">
<xs:pattern value="[A-Z][A-Z][A-Z]"/>
    initials chi
    chấp nhận
    giá trị là 3 kí
                               </xs:restriction>
                             </xs:simpleType>
    tự hoa từ A-Z
                            </xs:element>
■ phần tử
   initials chi
                      <xs:element name="initials">
   chấp nhận
                       <xs:simpleType>
                         <xs:restriction base="xs:string">
  <xs:pattern value="[a-zA-Z][a-zA-Z][a-zA-Z]"/>
   giá trị là 3 kí
   tự hoa hoặc
                         </xs:restriction>
   thường từ a-
                      </xs:element>
   z hoặc A-Z
```

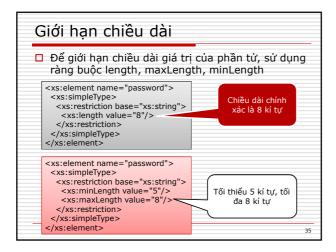


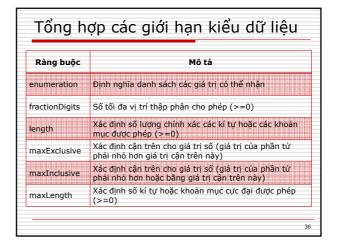




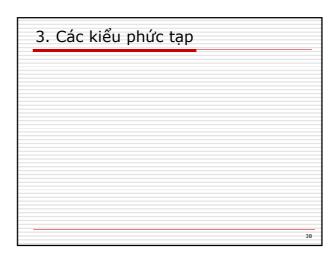


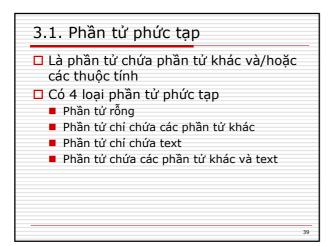


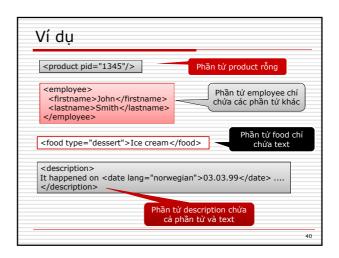












Dịnh nghĩa phần tử phức tạp

Quan sát ví dụ:

<p

```
Dinh nghĩa phần tử phức tạp (tiếp)

□ Cách 2: Định nghĩa thông qua tên kiểu phức tạp

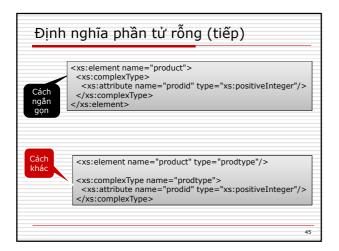
<pr
```

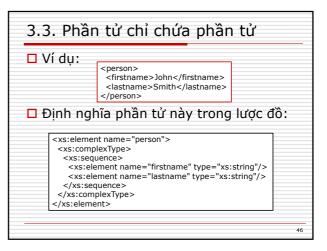
```
Có thể định nghĩa 1 kiểu phức tạp dựa trên 1 kiểu phức tạp
```

```
3.2. Phần tử rỗng:

| Phần tử rỗng:
| Sproduct prodid="1345"/>
| Dịnh nghĩa phần tử rỗng:

| (xs:element name="product">
| (xs:complexType>
| (xs:complexContent>
| (xs:restriction base="xs:integer">
| (xs:restriction base="xs:integer">
| (xs:complexContent>
| (xs:
```

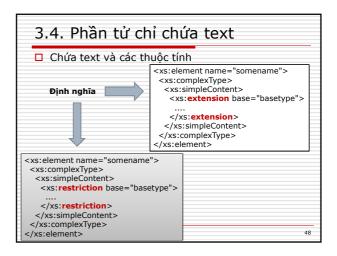


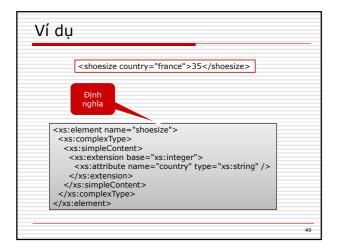


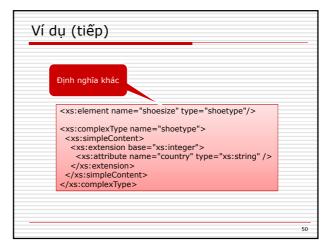
```
Phần tử chỉ chứa phần tử (tiếp)

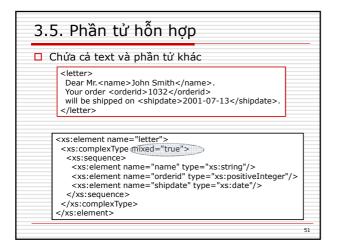
Dịnh nghĩa cách khác:

| (xs:element name="person" type="persontype"/>
| (xs:complexType name="persontype">
| (xs:sequence>
| (xs:element name="firstname" type="xs:string"/>
| (xs:element name="lastname" type="xs:string"/>
| (xs:element name="lastname" type="xs:string"/>
| (xs:sequence>
| (xs:complexType>
```











3.6. Indicator

Có 7 indicator:

Indicator trình tự

All
Choice
Sequence
Indicator xuất hiện
maxOccurs
minOccurs
Indicator nhóm
Group name
attributeGroup name

```
3.6.1. Indicator trình tự (tiếp)

Indicator Sequence: các phần tử con phải xuất hiện theo đúng trật tự

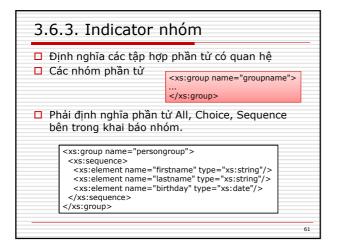
(xs:element name="person">
(xs:complexType>
(xs:sequence)
(xs:element name="firstname" type="xs:string"/>
(xs:element name="lastname" type="xs:string"/>
(xs:sequence>
(xs:complexType>
(xs:complexType>
(xs:element)
```

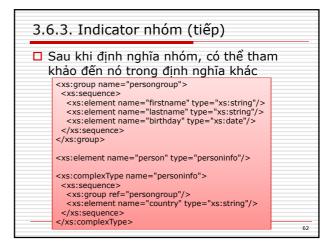
```
VÍ dụ: Myfamily.xml

<p
```

```
family.xsd

<pre
```





```
3.6.3. Indicator nhóm (tiếp)

Các nhóm thuộc tính:

dịnh nghĩa với khai báo như sau:

(xs:attributeGroup name="groupname">
...
(/xs:attributeGroup>

ví dụ:

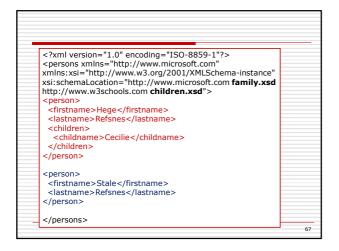
(xs:attributeGroup name="personattrgroup">
(xs:attributeGroup name="firstname" type="xs:string"/>
(xs:attribute name="lastname" type="xs:string"/>
(xs:attribute name="birthday" type="xs:date"/>
(/xs:attributeGroup>
```

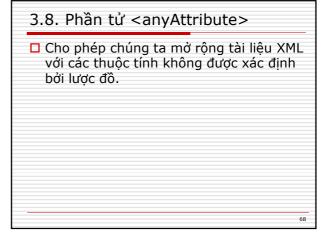
```
Children.xsd

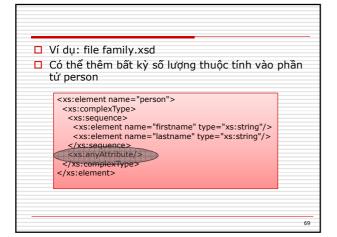
□ Chúng ta muốn mở rộng phần tử person với một phần tử children

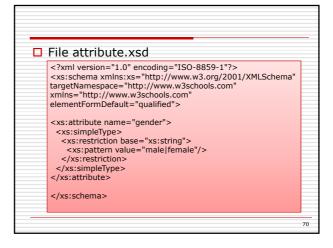
<?xml version="1.0" encoding="ISO-8859-1"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/xMLSchema" targetNamespace="http://www.w3schools.com" xmlns="http://www.w3schools.com" elementFormDefault="qualified">

<xs:element name="children">
<xs:complexType>
<xs:sequence>
<xs:sequence>
</xs:complexType>
```



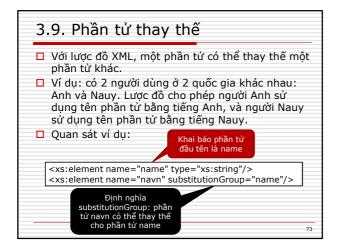


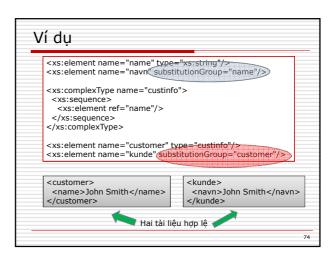


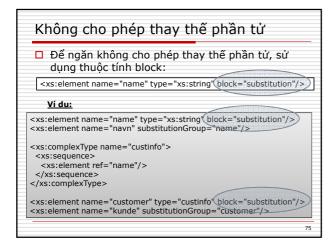


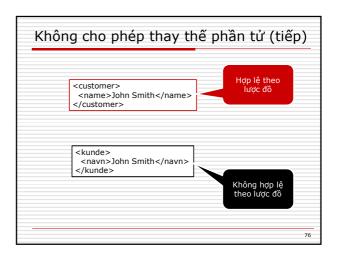
File Myfamily.xml sau sử dụng các thành phần từ 2 file family.xsd và attribute.xsd:

□ Như vậy, phần tử <any> và <anyAttribute> được sử dụng để làm cho tài liệu có thể mở rộng. Chúng cho phép thêm các phần tử chưa được khai báo trong XML schema chính.







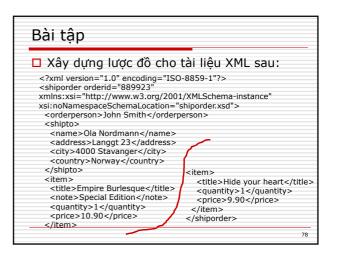


Chú ý

Phần tử đầu và phần tử thay thế phải có cùng kiểu

Tất cả các phần tử đầu và phần tử thay thế nó phải được khai báo là phần tử toàn cục.

Phần tử toàn cục là phần tử con trực tiếp của phần tử <schema>. Các phần tử lồng trong phần tử khác là các phần tử cục bộ.



```
Cách 1:

| File shiporder.xsd:
| <?xml version="1.0" encoding="ISO-8859-1" ?>
| <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">|
| <xs:element name="shiporder">|
| <xs:complexType>
| <xs:sequence>
| <xs:element name="orderperson" type="xs:string"/>|
| <xs:element name="shipto">|
| <xs:complexType>
| <xs:complexType>
| <xs:complexType>
| <xs:element name="rame" type="xs:string"/>|
| <xs:element name="name" type="xs:string"/>|
| <xs:element name="address" type="xs:string"/>|
| <xs:element name="country" type="xs:string"/>|
| <xs:element name="country" type="xs:string"/>|
| <xs:element name="country" type="xs:string"/>|
| <xs:element name="country" type="xs:string"/>|
| <xs:complexType>|
| <xs:element>|
```

```
Cách 2:

<!-- definition of complex elements -->
<xs: element name="shipto">
<xs: complexType>
<xs: complexType>
<xs: element ref="name"/>
<xs: element ref="address"/>
<xs: element ref="city"/>
<xs: element ref="country"/>
<xs: element ref="country"/>
<xs: complexType>
</xs: complexType>
</xs: element name="item">
<xs: complexType>
<xs: element ref="citite"/>
<xs: element ref="title"/>
<xs: element ref="note" minOccurs="0"/>
<xs: element ref="quantity"/>
<xs: element ref="price"/>
<xs: element ref="price"/>
<xs: element ref="price"/>
</xs: complexType>
</xs: element ref="price"/>
<xs: element ref="price"/>
</xs: element ref="price"/>
</xs
```

```
Cách 3: định nghĩa kiểu

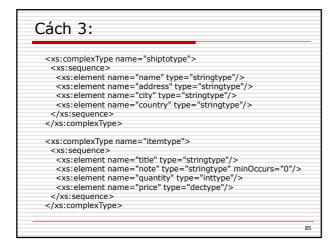
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

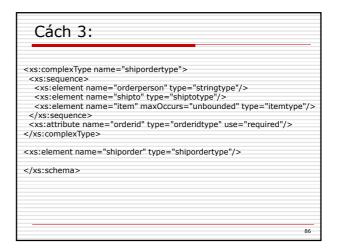
<xs:simpleType name="stringtype">
<xs:restriction base="xs:string"/>
</xs:simpleType>

<xs:restriction base="xs:positiveInteger"/>
</xs:simpleType>

<xs:restriction base="xs:positiveInteger"/>
</xs:simpleType name="dectype">
<xs:restriction base="xs:decimal"/>
</xs:simpleType>

<xs:restriction base="xs:decimal"/>
</xs:restriction base="xs:string">
<xs:restriction base="xs:string">
<xs:restriction base="xs:string">
<xs:restriction base="xs:string">
<xs:restriction base="xs:string">
<xs:restriction>
</xs:restriction>
</xs:simpleType>
```





- 4. Các kiểu dữ liệu

  Kiểu chuỗi

  Kiểu Date

  Kiểu số
- 4.1. Kiểu chuỗi

  Chứa các kí tự, kí tự xuống dòng, kí tự về đầu dòng, kí tự tab

  Ví dụ:

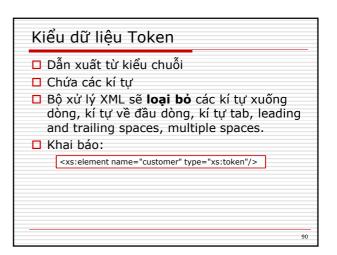
  khai báo

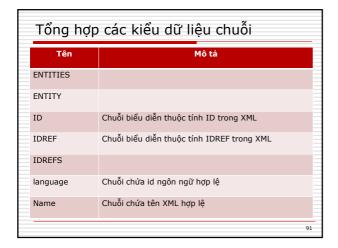
  xs:element name="customer" type="xs:string"/>

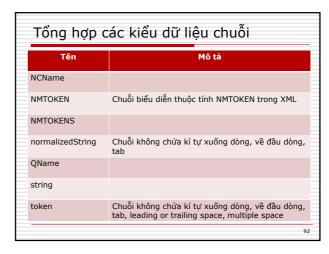
  phần tử trong tài liệu XML có thể:

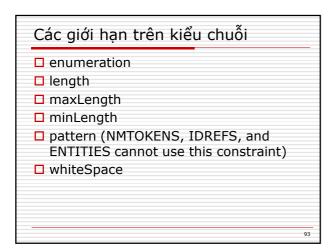
  xcustomer> John Smith </customer>

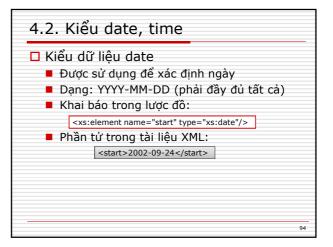
  Khi sử dụng kiểu chuỗi, bộ xử lý XML sẽ không biến đổi giá trị!







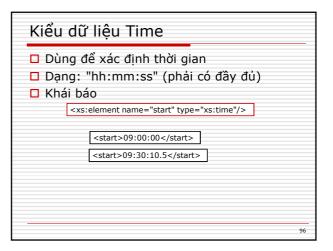


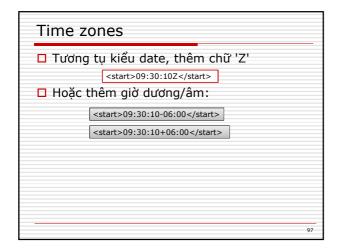


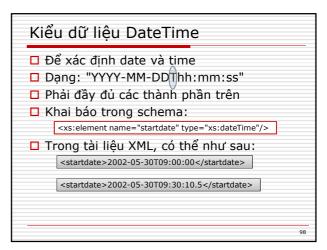
Time zones

□ Để xác định vùng thời gian theo UTC, thêm chữ Z sau date

| <start>2002-09-24Z</start>
| Xác định offset: thêm thời gian dương/âm
| <start>2002-09-24-06:00</start>
| <start>2002-09-24+06:00</start>







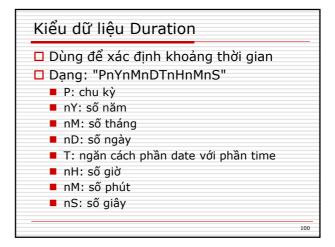
Time zones

Tương tự, thêm Z vào đẳng sau hoặc thêm giờ dương/âm

<startdate>2002-05-30T09:30:10Z</startdate>

<startdate>2002-05-30T09:30:10-06:00</startdate>

<startdate>2002-05-30T09:30:10+06:00</startdate>



Kiểu dữ liệu Duration (tiếp)

Khai báo trong lược đồ:

(xs:element name="period" type="xs:duration"/>

Phần tử trong tài liệu

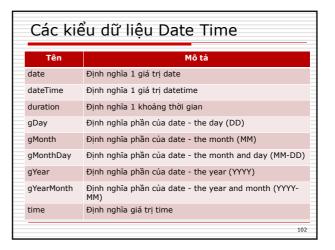
Chu kỳ 5 năm: (speriod>P5Y</period>
Chu kỳ 5 năm 2 tháng 10 ngày:
(speriod>P5Y2M10D</period>

Chu kỳ 5 năm, 2 tháng, 10 ngày 15 giờ:
(speriod>P5Y2M10DT15H</period>

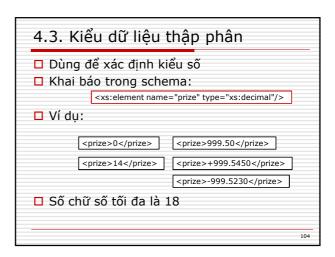
Chu kỳ 15 giờ: (speriod>P15H</period>

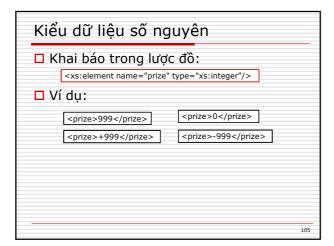
Thêm dấu trừ: chu kỳ âm

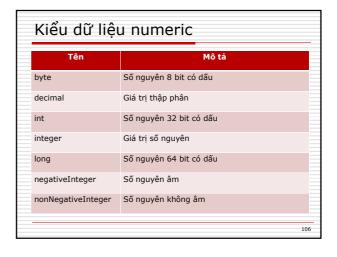
(speriod>-P10D</period>

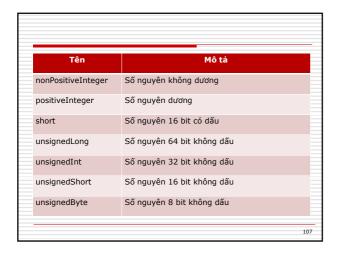


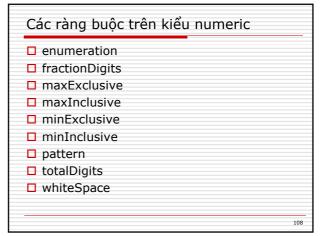




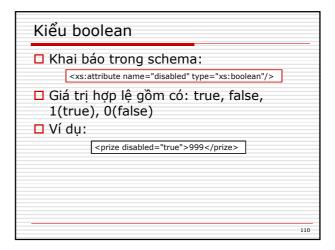


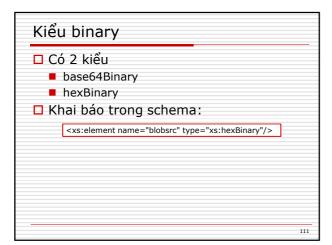


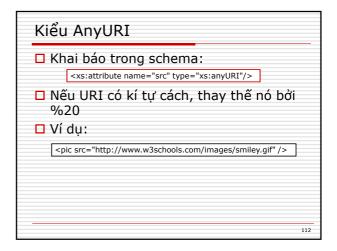




□ Boolean	
☐ Binary ☐ AnyURI	
	109







Một số kiểu

anyURI
base64Binary
boolean
double
float
hexBinary
NOTATION
QName

Các giới hạn trên những kiểu này

nenumeration (kiểu boolean không sử dụng ràng buộc này)

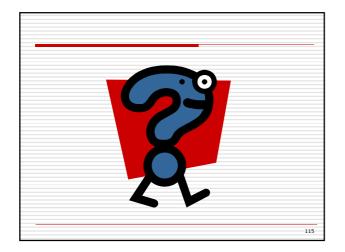
length (kiểu boolean không sử dụng ràng buộc này)

maxLength (kiểu boolean không sử dụng ràng buộc này)

minLength (kiểu boolean không sử dụng ràng buộc này)

pattern

whiteSpace



## Lời hay ý đẹp "Hãy lấy sự khôn ngoan làm hành trang cho bạn trong hành trình của cuộc đời, vì đó là sự nâng đỡ đáng tin cậy hơn mọi tài sản khác" G. Bizet