

Fall 2006 OCT Lecture 1

“An Introduction to XML” or “Why I still find computers exciting”

Michael McCarthy
Associate Teaching Professor
Carnegie Mellon University

XML On The World Wide Web

- What is XML?
- What is XML's Primary purpose?
- XML in 10 points (from the W3C)
- Example Documents
- Predictions

What is XML?

- ❑ XML is an acronym that stands for the eXtensible Markup Language.
- ❑ It is a flexible framework which can be used to create new customized markup languages.
- ❑ HTML, for example, is a markup language with a fixed tag set. <P> <H1>
- ❑ XML, on the other hand, does not define any particular set of tags. It allows you to create your own tag set.

What is XML's primary purpose?

- ❑ **Separating data from presentation**

The range of internet capable devices is growing- WAP phones, pagers, TV sets, web browsers, in car computers, have different presentation requirements.

- ❑ **Transmitting data between applications**
XML allows systems to interoperate.

XML in 10 Points

XML is for structuring data

XML looks a bit like HTML

XML is text, but it isn't meant to be read

XML is verbose by design

XML is a family of technologies

(Xlink, Xpointer, CSS, XSLT, DOM, SAX...)

XML is new, but not that new

XML in 10 Points

XML leads HTML to XHTML

XML is modular

XML is the basis for RDF and the Semantic Web

XML is license-free, platform-independent and well-supported



Example XML Documents

An Interest Rate Swap in XML

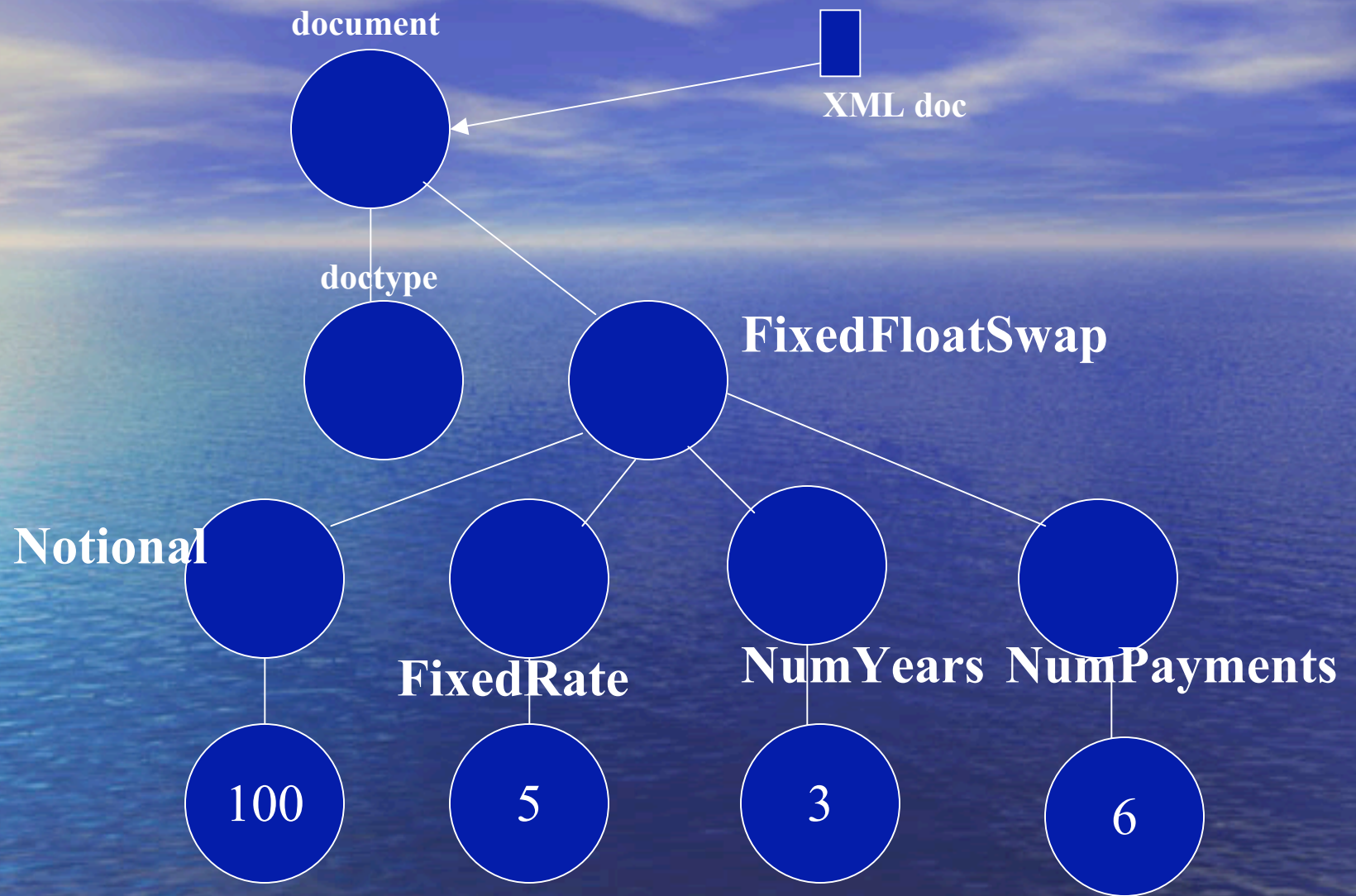
```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE FixedFloatSwap SYSTEM
    "FixedFloatSwap.dtd">
<FixedFloatSwap>
  <Notional currency="pounds">100</Notional>
  <Fixed_Rate>5</Fixed_Rate>
  <NumYears>3</NumYears>
  <NumPayments>6</NumPayments>
</FixedFloatSwap>
```


A DTD for Interest Rate Swap

```
<?xml version="1.0" encoding="utf-8"?>  
<!ELEMENT FixedFloatSwap ( Notional, Fixed_Rate,  
                             NumYears, NumPayments ) >  
  
<!ELEMENT Notional (#PCDATA)>  
<!ATTLIST Notional currency (dollars | pounds)  
                #REQUIRED>  
  
<!ELEMENT Fixed_Rate (#PCDATA) >  
<!ELEMENT NumYears (#PCDATA) >  
<!ELEMENT NumPayments (#PCDATA) >
```

Document Type Definitions

- DTD's are used to define a class of document
- DTD's specify a grammar
- Valid documents conform to the grammar specified in the DTD.
- Automatic validation is provided by parsers from Sun, Microsoft, IBM, Oracle and others.
- An application programmer can program against these documents by referring to their DTDs.
- XSDL (XML Schema Definition Language) is an improvement



The document as seen by the DOM programmer

Real World XML Languages

MATHML (1998)

- The first specialized XML Vocabulary

MathML

Working Group Members at W3C:

- Wolfram Research
(Mathematica 3.0)
- Waterloo Maple (Maple V7.0)

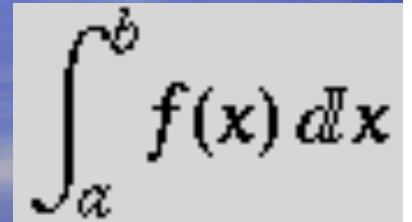
MathML Example

```
<math>
  <apply>
    <int/>
    <bvar>
      <ci> x
    </ci>
    </bvar>
    <lowlimit>
      <ci>
        a
      </ci>
    </lowlimit>
```

```

  </apply>
</math>

<uplimit>
  <ci>
    b
  </ci>
</uplimit>
<apply>
  <ci type="fn"> f
  </ci>
  <ci>
    x
  </ci>
</apply>
```


$$\int_a^b f(x) dx$$

Chemical Markup Language (CML)

Tackling the problems associated
with the exchange of chemical
information since 1995

A CML Example Document

```
<cml>
<molecule>
<bibliography>
<person>
<string title="lastname">PROUT</string>
<string title="initials">C.K</string>
</person>
<person>
<string title="lastname">ARMSTRONG</string>
<string title="initials">R.A</string>
</person>
<person>
```

Crystal.xml from
CML's web site


```
<string title="lastname">CARRUTHERS</string>
<string title="initials">J.R</string>
</person>
<person>
<string title="lastname">FORREST</string>
<string title="initials">J.G</string>
</person>
<person>
<string title="lastname">MURRAY-RUST</string>
<string title="initials">P</string>
</person>
<string builtin="JOUR">J.CHEM.SOC.A</string>
</bibliography>
```

```
<molecule class="2d">
<formula><string title="stoichiometry">
    C16 H18 CU1 O8</string></formula>
<atomArray>
<stringArray builtin="elementType">CU O O O O O O C
    C C C C C C C C C C
    O O C C C C C C C
</stringArray>
<stringArray builtin="atomId">a1 a2 a3 a4 a5 a6 a7 a8 a9
    a10 a11 a12 a13 a14 a15
    a16 a17 a18 a19 a20 a21
    a22 Long Document Truncated
```




The Financial Product Markup Language
(FpML)

Enables e-commerce activities in
the field of financial derivatives.

An FpML Example Document

```
<?xml version = "1.0"?>
```

```
<!DOCTYPE FpML PUBLIC "-//FpML//DTD Financial  
product Markup Language 1-0//EN"  
"fpml-dtd-1-0-2001-05-14.dtd" >
```

```
<FpML version = "1-0"
```

```
businessCenterSchemeDefault =
```

```
"http://www.fpml.org/spec/2000/business-center-1-0"
```

```
businessDayConventionSchemeDefault =
```

```
"http://www.fpml.org/spec/2000/business-day-convention-1-0"
```



```
<trade>
  <tradeHeader>
    <partyTradeIdentifier>
      <partyReference href = "#CHASE" />
      <tradeId tradeIdScheme =
        "http://www.chase.com/swaps/trade-id">
        TW9235
      </tradeId>
    </partyTradeIdentifier>
```

Document truncated - about 10 pages
of text

XBRL

Extensible Business Reporting Language (XBRL) is an open specification which uses XML-based data tags to describe **financial statements** for both public and private companies.

XBRL benefits all members of the financial information supply chain. (From XBRL.org)

Taken from Merk's Balance Sheet

```
<group type='ci:currentAssets.receivablesNet'>
  <label href='xpointer(..)' xml:lang='en'>Accounts
    receivable</label>
  <item period='1999-12-31'>4089.0</item>
  <item period='1998-12-31'>3374.1</item>
</group>
<group type='ci:currentAssets.inventoriesNet'>
  <label href='xpointer(..)' xml:lang='en'>Inventories</label>
  <item period='1999-12-31'>2846.9</item>
  <item period='1998-12-31'>2623.9</item>
</group>
```

The Web Ontology Language OWL

All dolphins are mammals.
Flipper is a dolphin.

=====

So...

The Web Ontology Language

OWL

```
<rdf:Description rdf:ID="Dolphin">  
  <rdf:type  
    rdf:resource= "http://www.w3.org/2000/01/rdf-schema#Class"/>  
</rdf:Description>
```

```
<rdf:Description rdf:about="#Flipper">  
  <rdf:type rdf:resource="animal-schema.rdf#Dolphin"/>  
</rdf:Description>
```

```
<owl:Class rdf:ID="Dolphin">  
  <rdfs:subClassOf rdf:resource = "#Mammal" />  
</owl:Class>
```

SOAP

The Simple Object Access Protocol is a lightweight protocol for the exchange of information in a decentralized, distributed environment.

Example SOAP Message

```
POST /StockQuote HTTP/1.1
Host: www.stockquoteserver.com
Content-Type: text/xml; charset="utf-8"
Content-Length: nnnn
SOAPAction: "Some-URI"
```

HTTP

XML (SOAP)

```
<SOAP-ENV:Envelope
  xmlns:SOAP-ENV=
    "http://schemas.xmlsoap.org/soap/envelope/"
  SOAP-ENV:encodingStyle=
    "http://schemas.xmlsoap.org/soap/encoding/">
  <SOAP-ENV:Body>
    <m:GetLastTradePrice xmlns:m="Some-URI">
      <symbol>DIS</symbol>
```

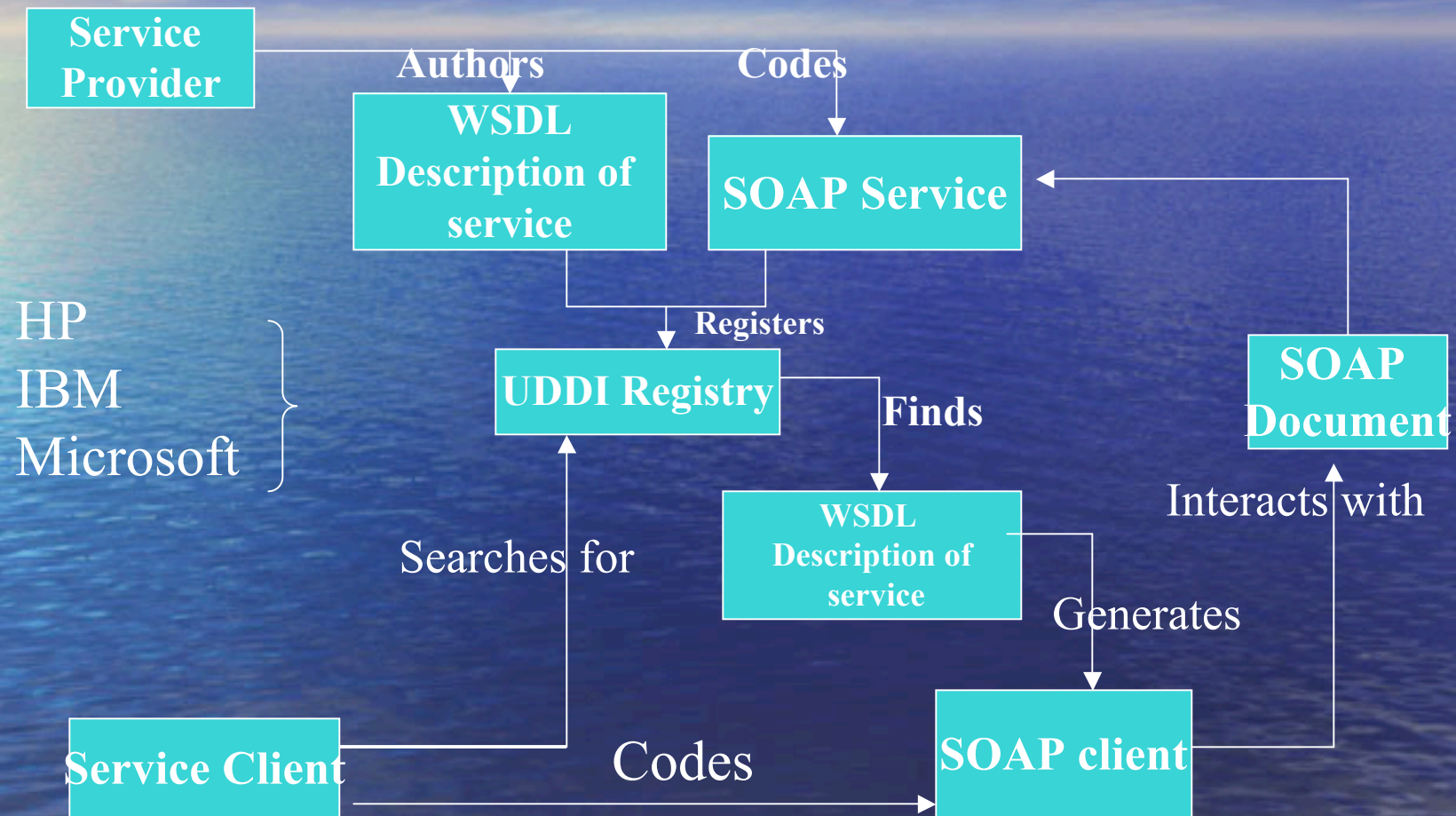
UDDI and Web Services

UDDI Universal Description, Discovery and Integration

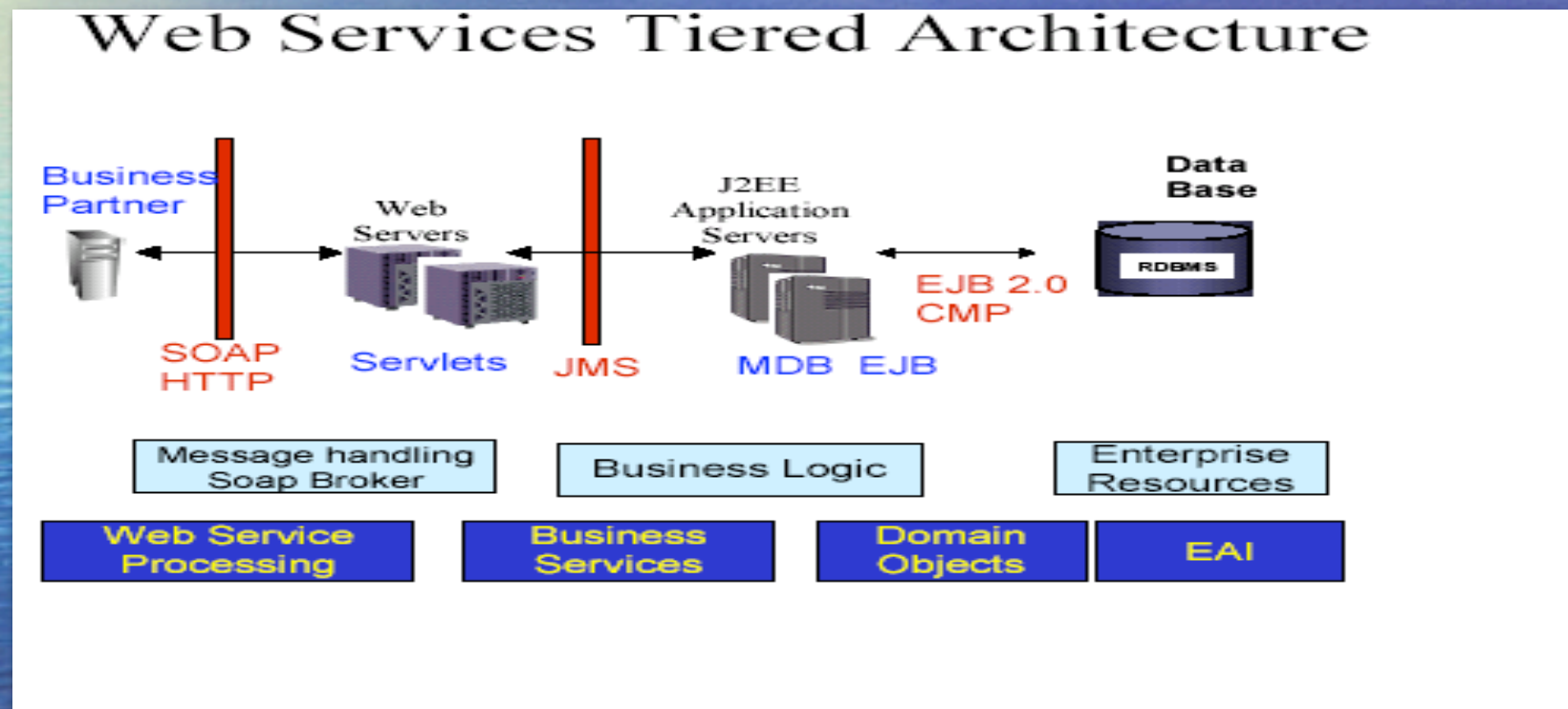
Discover businesses worldwide that offer the exact products and services that you need.
Register the products and services of your own business for others to discover.
(From UDDI.org)

This slide adapted from
Page 353 of “Java and XML”
By McLaughlin

UDDI

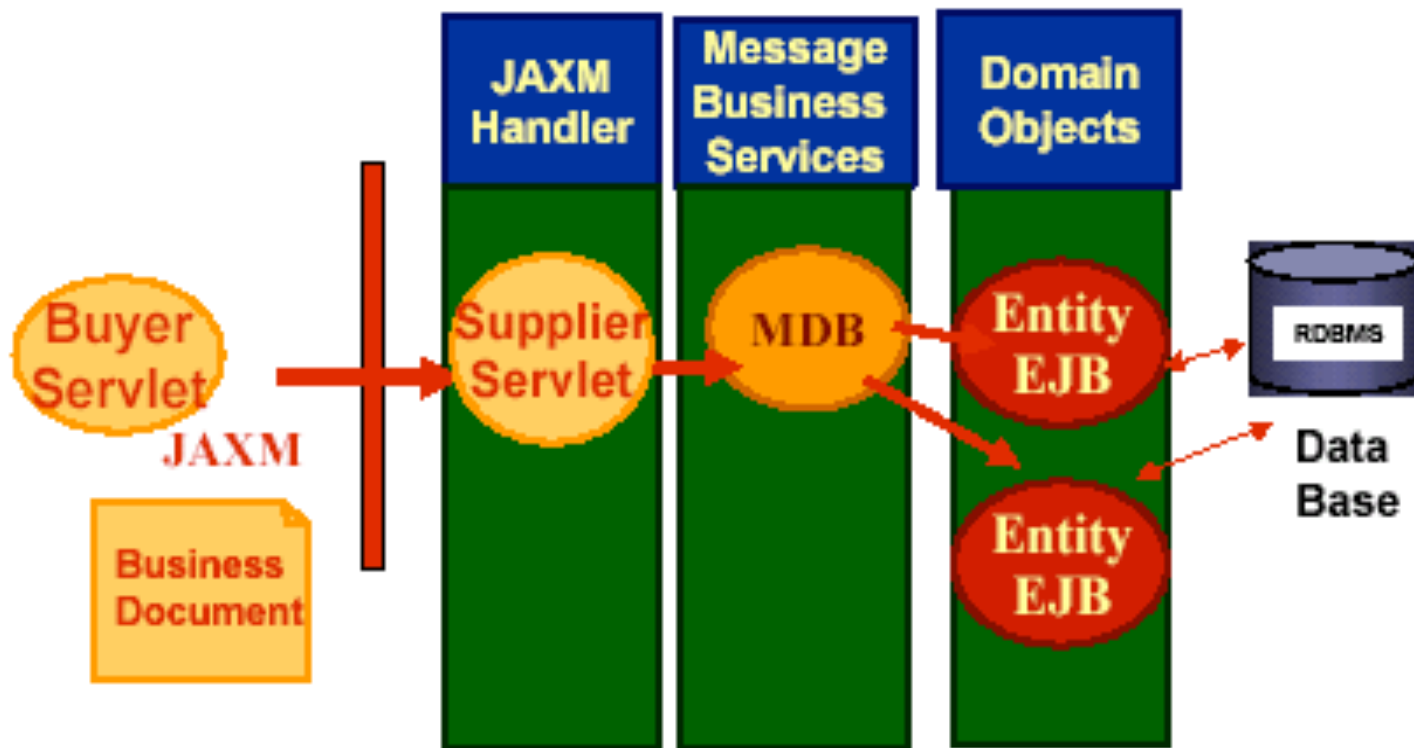


SOAP and J2EE



Slide from JAXM/JMS tutorial at Sun Microsystems

JAXM Web Services & J2EE



Slide from JAXM/JMS tutorial at Sun Microsystems

Processing XML

Many technologies exist (and many are being developed) for processing XML documents.

A few of these are :

- DOM Document Object Model

- SAX The Simple API for XML

- XSLT The extensible Stylesheet Language for transformation

- JAXB Java API for XML Binding

- JAXM Java API for XML Messaging (SOAP)

- Jena from HP

Safe Predictions

- This standard approach to data representation will lead to the development of new and interesting software.
- The application programmer will have the internet as a data source and an API. Prior to the XML revolution, the data sources and API's were mostly local and proprietary.
- The data sources will be both static and dynamic. (RSS Really Simple Syndication became popular over the summer of 2003.)
- Perhaps, OWL and OWL-S will become major standards

Required Reading...

“XML and the Second-Generation Web”

by Jon Bosak and Tim Bray,

Scientific American

May 1999

“The Semantic Web”

by Tim Berners-Lee, James S. Hendler and Ora Lassila,

Scientific American

May 2001