

Understanding XML

Hans C. Arents
senior IT market analyst

I.T. Works
"Guiding the IT Professional"

Innovation Center, Technologiepark 3, B-9052 Gent (Belgium), Tel: +32 (0)9 241 56 21 - Fax: +32 (0)9 241 56 56

E-mail: hca@itworks.be - Site: <http://www.itworks.be/> - Home: <http://www.arents.be/>

Understanding XML

n Part 1: **XML intro**

13u30 – 15u00

coffee break

n Part 2: XML standards

15u30 – 17u30

questions & answers

XML intro

Hans C. Arents

senior IT market analyst

I.T. Works

"Guiding the IT Professional"

Innovation Center, Technologiepark 3, B-9052 Gent (Belgium), Tel: +32 (0)9 241 56 21 - Fax: +32 (0)9 241 56 56

E-mail: hca@itworks.be - Site: <http://www.itworks.be/> - Home: <http://www.arents.be/>

An introduction to XML

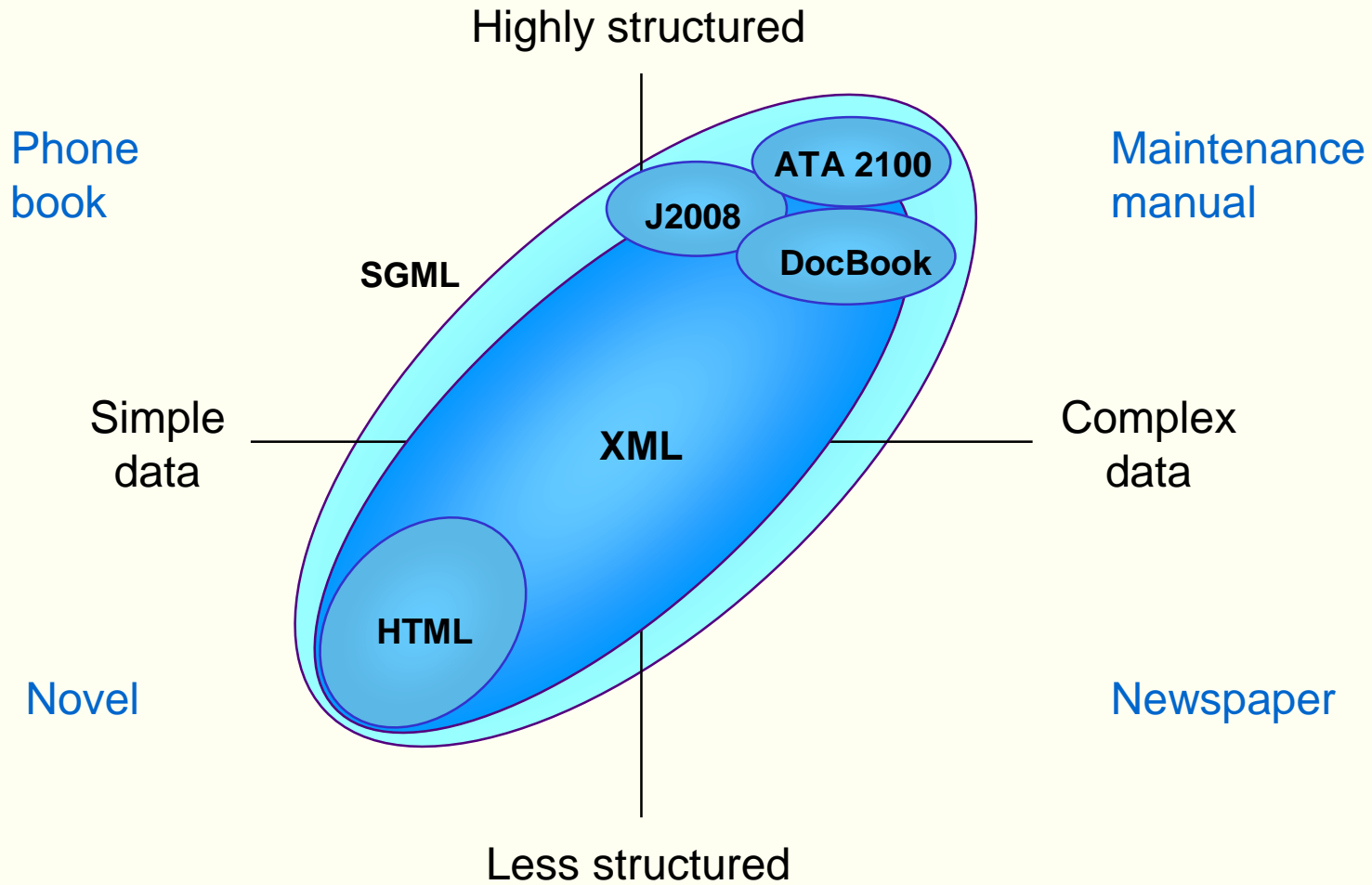
- n** The need for XML
 - XML compared to HTML / SGML
- n** Key concepts behind XML
 - XML is SGML
 - XML is about structure
 - XML is about meaning
 - XML is about validation
- n** XML is about documents/data
 - XML as a document format
 - XML as a data format
 - XML as a metadata format
- n** XML is the universal Web format

The need for XML

n The need to go beyond HTML

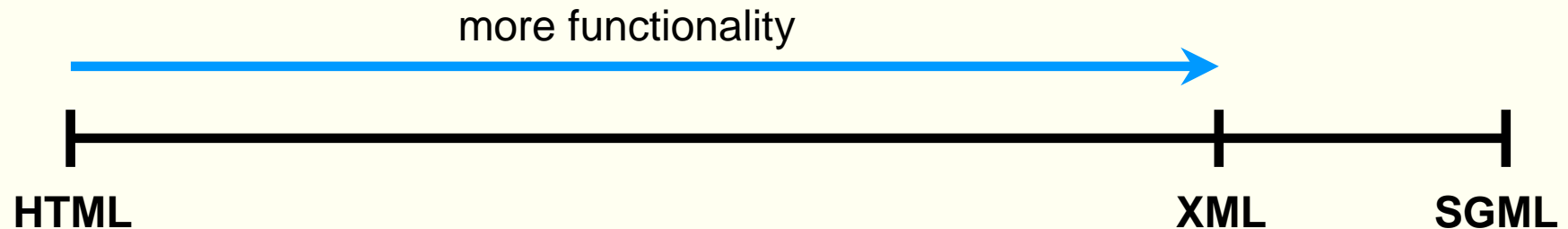
- û fixed, simplistic, unstructured, unmanageable
 - û hardwired for presentation, not designed for data
 - suggested solution: **SGML** (**S**tandard **G**eneralized **M**arkup **L**anguage)
 - ü extensible and structured by definition
 - ü has all the necessary mechanisms to address HTML's shortcomings
 - û big and (in certain spots very) complicated
 - û too hard to use for Web developers, not enough benefits for Web users
 - real solution:
 - throw the hard parts of SGML away
 - optimize SGML for Web creation and delivery à XML is SGML – –
 - learn from the lessons taught by HTML
 - enrich SGML by looking at what's good in HTML à XML is not HTML + +
- find a sexy name: **XML** (**E**xensible **M**arkup **L**anguage)

XML compared to HTML / SGML

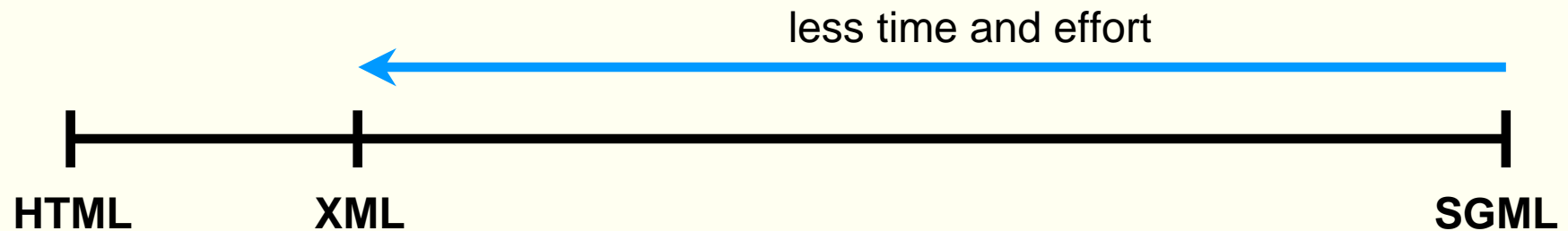


XML compared to HTML / SGML

n 80% of the SGML functionality ...



n ... for 20% of the cost in time and effort

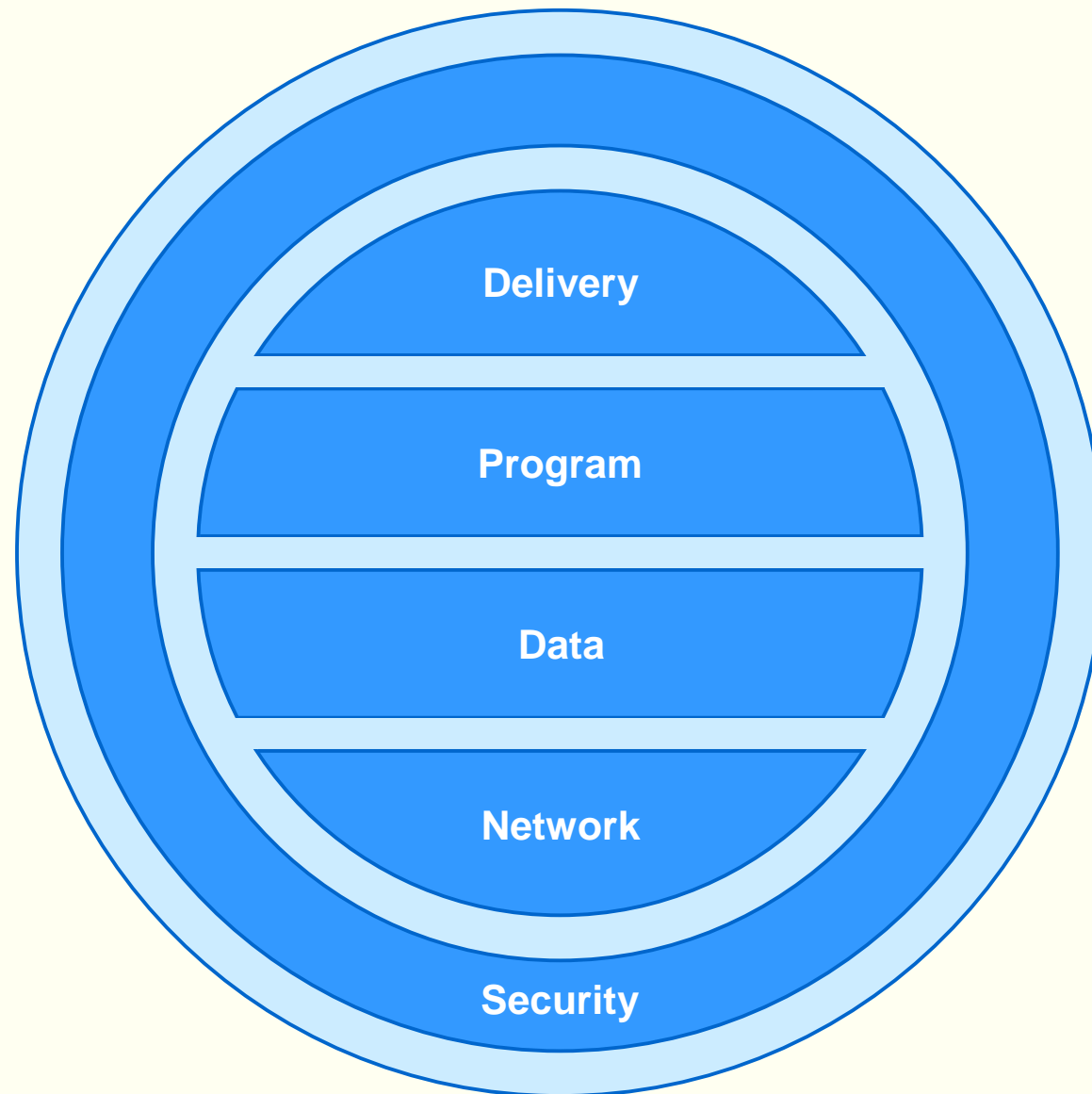


XML history

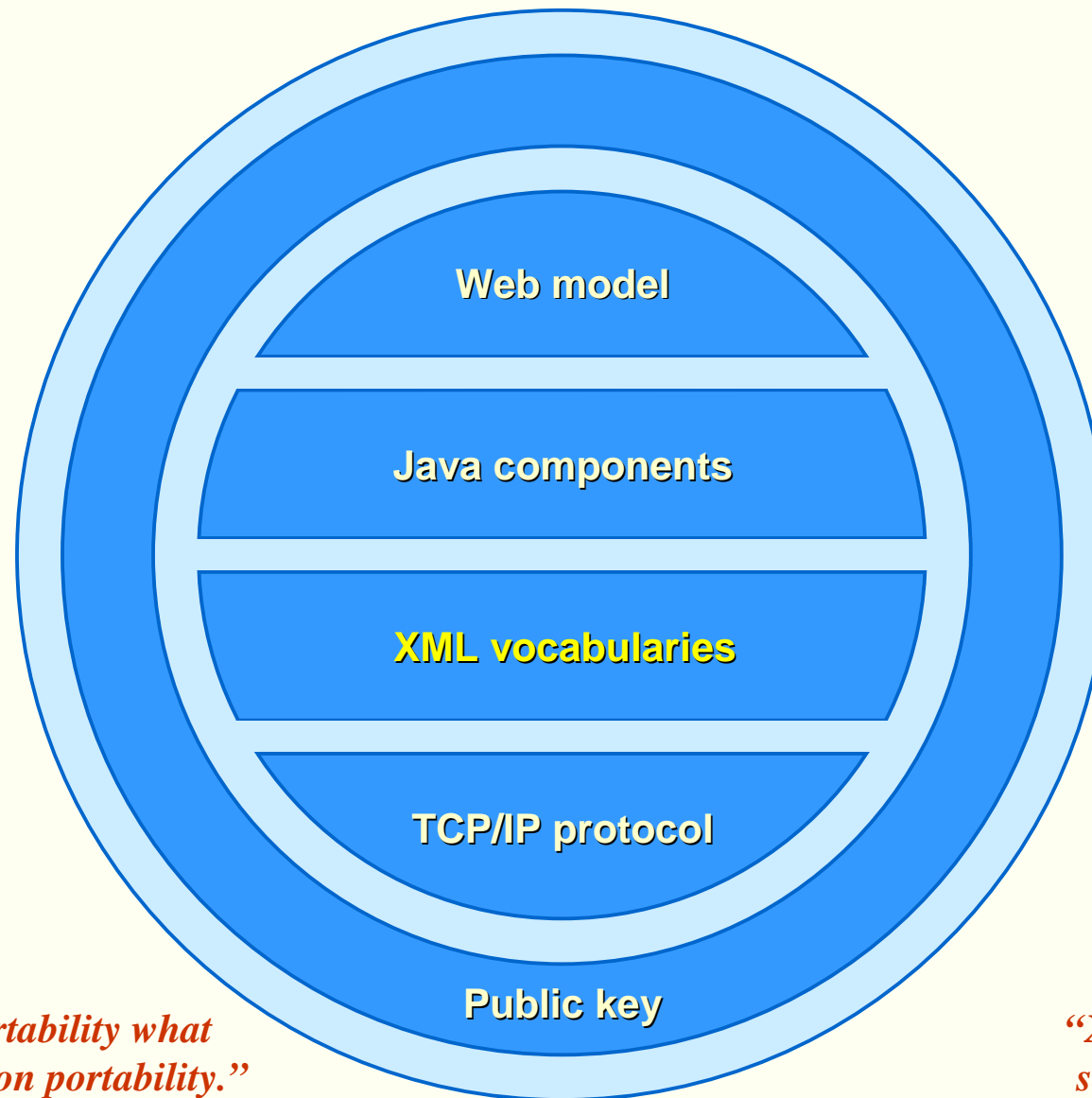
n Milestones:

- Jul '96 W3C XML Working Group is formed
- Nov '96 First draft of XML standard published
- Mar '97 Microsoft announces CDF push format
- Oct '97 Microsoft ships IE 4.0 with 2 built-in XML parsers
- Feb '98 Final XML 1.0 standard approved by W3C**
- Q3 '98 Availability of commercial XML tools and technologies
- Q4 '98 Announcements of XML support: IBM, Sun, Oracle, SAP, ...
- Jan '99 Microsoft announces XML native data format in Office 2000
- Apr '99 Microsoft ships IE 5.0 with (not quite) complete XML support
- Dec '99 Netscape promises to ship Navigator 6.0 with XML support
- in 2000 Widespread adoption of XML - Core of MS .NET framework
- for Internet-enabled document *and* data applications

XML: the missing piece of the puzzle



XML: the missing piece of the puzzle



*“XML is to data portability what
Java is to application portability.”*

*“XML gives Java
something to do”*

Key concept: XML is SGML

n XML looks a lot like HTML, but really behaves like SGML

- focus on information *structure* and *meaning* instead of *presentation*
- for simple markup languages, but adhering to strong conventions
- obscure and never-used SGML features have been left behind

n XML is an **open standard**

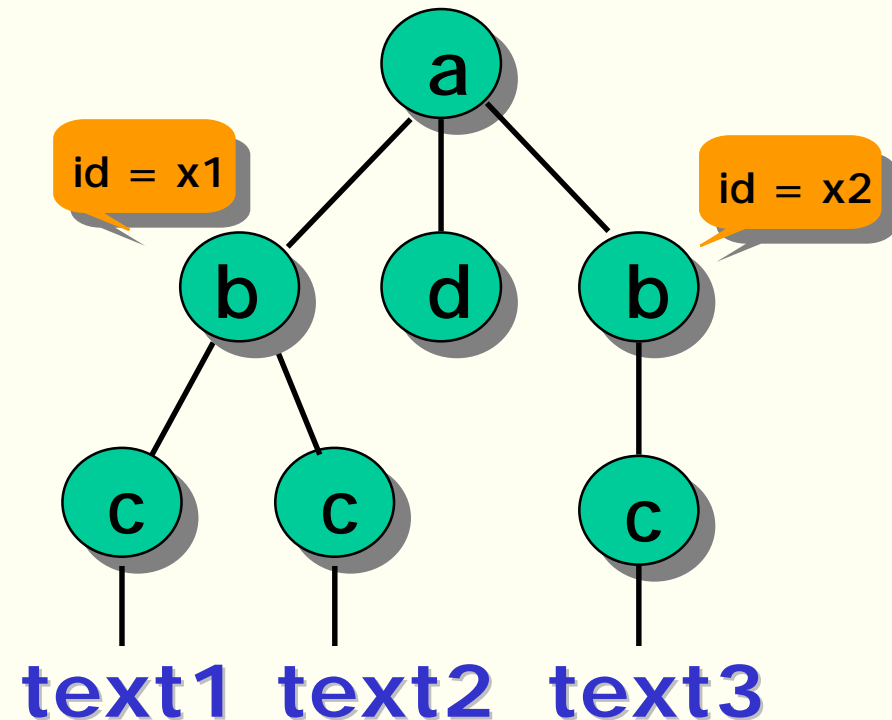
- not a proprietary standard
- not an international standard (ISO)
- but a de facto consensus standard (W3C)

n XML is **extensible**

- not a single markup language,
but an infinite variety of markup languages
- tags can be added to an XML-based markup language without necessarily breaking the applications that use that markup language

Key concept: XML is about structure

```
<?xml version="1.0"?>  
  
<a>  
  <b id="x1">  
    <c>text1</c>  
    <c>text2</c>  
  </b>  
  <d att="xyz"/>  
  <b id="x2">  
    <c>text3</c>  
  </b>  
</a>
```

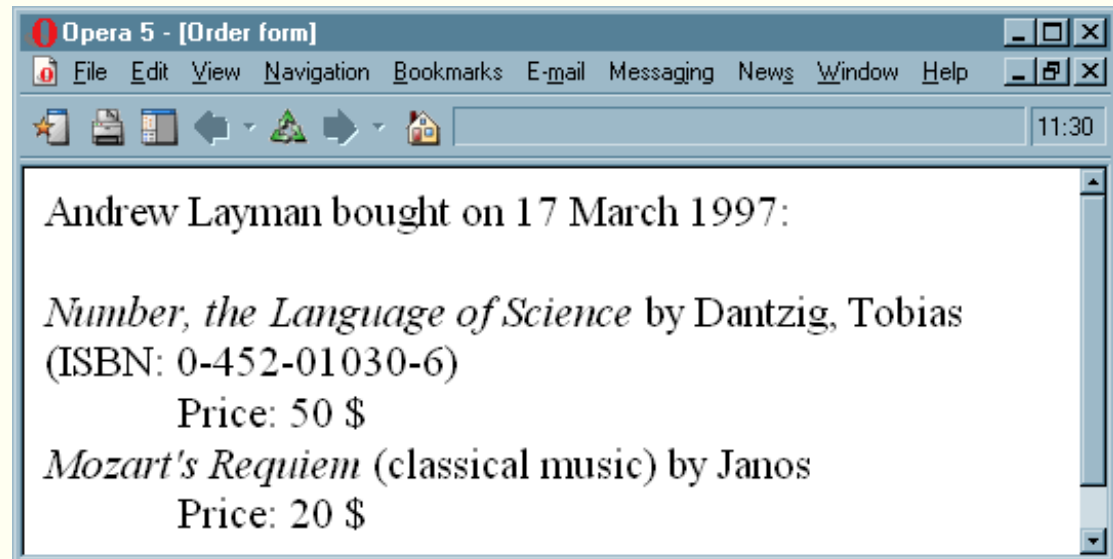


data markup
“transport format”
well-formed document

data structure
“processing format”
tree of elements/attributes

Key concept: XML is about meaning

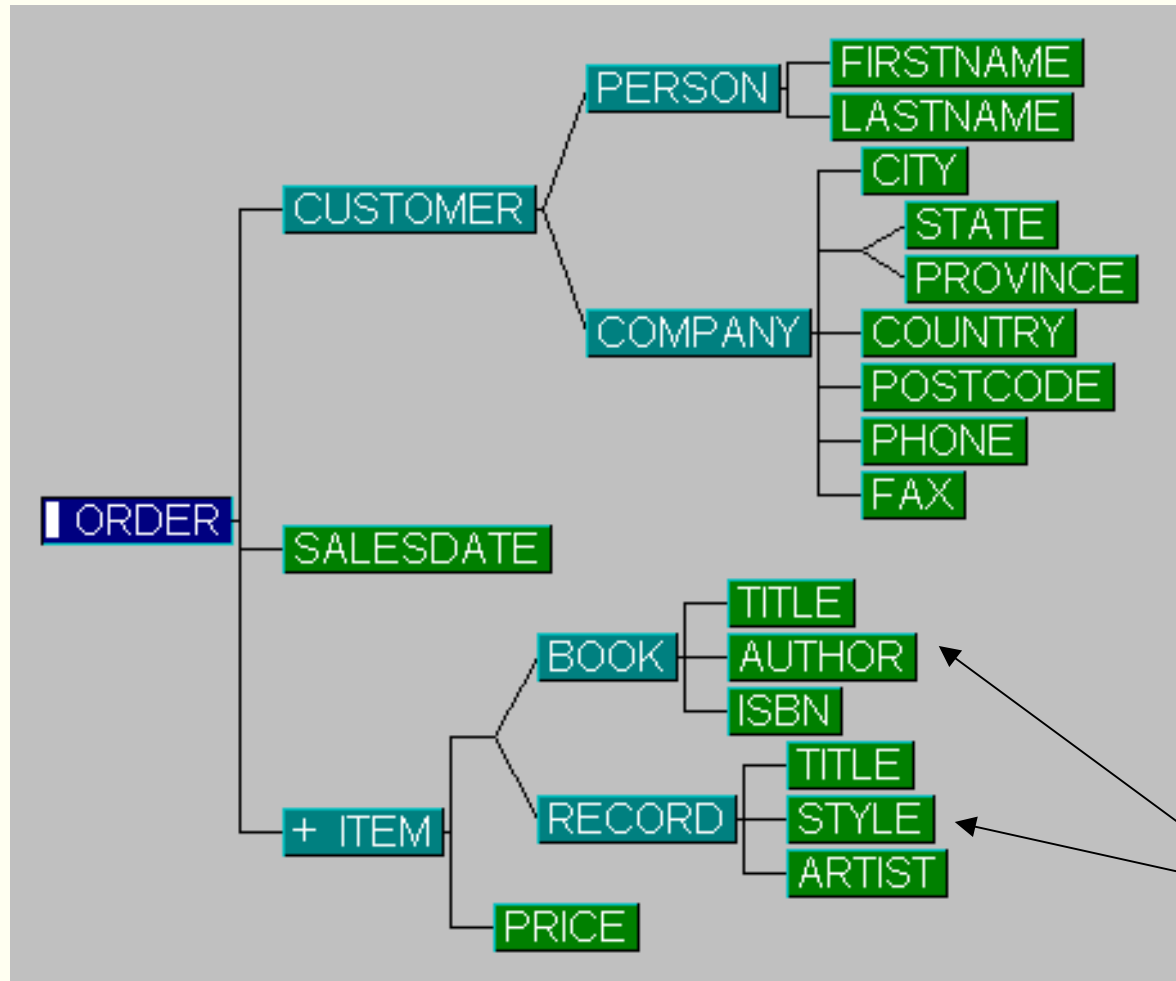
```
<html>
  <head>
    <title>Order form</title>
  </head>
  <body>
    <p>Andrew Layman bought on 17 March 1997:</p>
    <dl>
      <dt><i>Number, the Language of Science</i>
        by Dantzig, Tobias (ISBN: 0-452-01030-6)</dt>
      <dd>Price: 50 $</dd>
      <dt><i>Mozart's Requiem</i>
        (classical music) by Janos</dt>
      <dd>Price: 20 $</dd>
    </dl>
  </body>
</html>
```



Key concept: XML is about meaning

```
<?xml version="1.0" standalone="yes"?>
<order>
  <customer>
    <person>
      <firstname>Andrew</firstname><lastname>Layman</lastname>
    </person>
  </customer>
  <salesdate>19970317</salesdate>
  <item>
    <book>
      <title>Number, the Language of Science</title>
      <author>Dantzig, Tobias</author>
      <isbn>0-452-01030-6</isbn>
    </book>
    <price currency="dollar">50</price>
  </item>
  <item>
    <record>
      <title><composer>Mozart</composer>'s Requiem</title>
      <style>classical music</style>
      <artist>Janos</artist>
    </record>
    <price currency="dollar">20</price>
  </item>
</order>
```

Key concept: XML is about validation



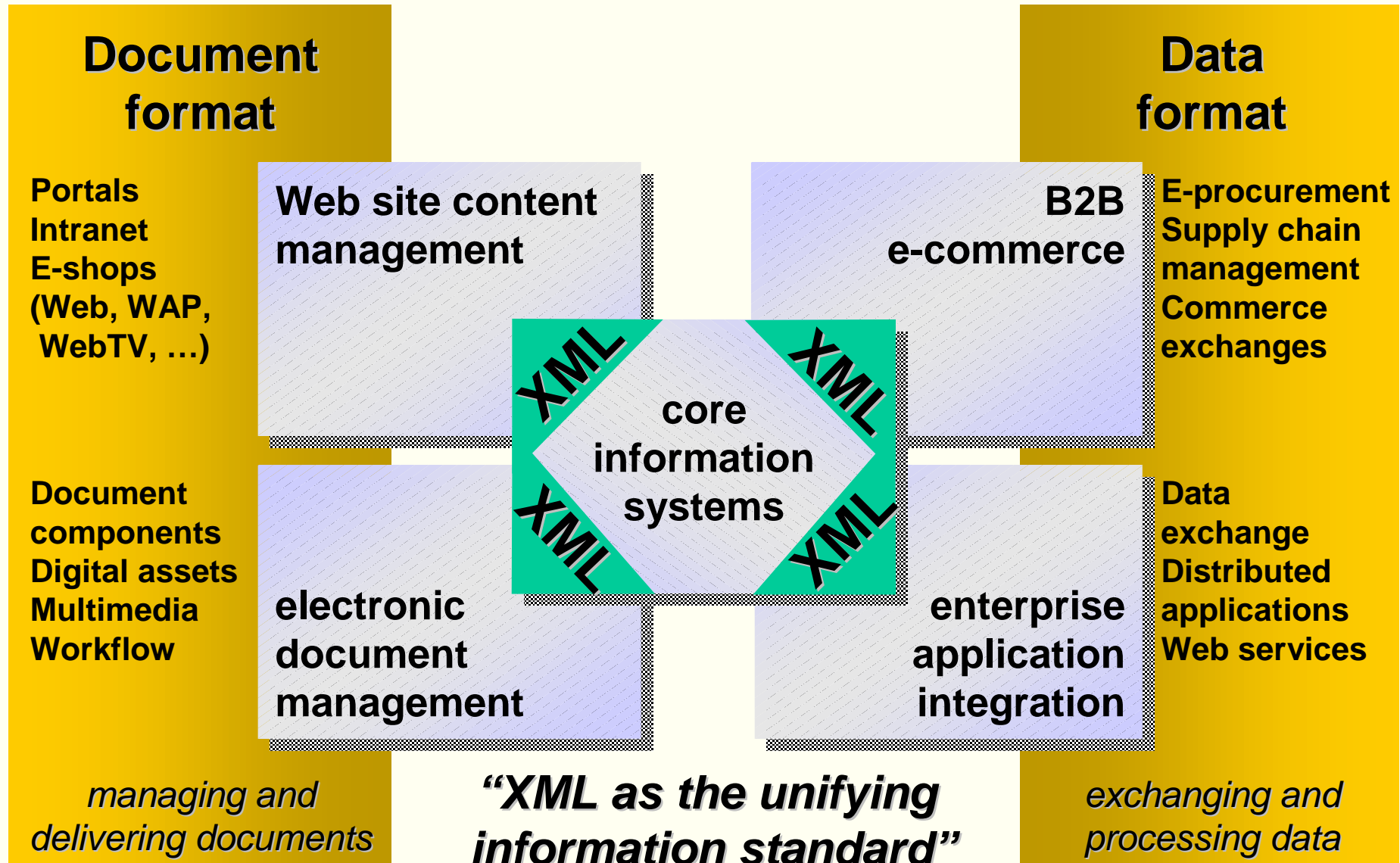
information =
content + **structure**
+ **datatypes**

expressed by
a **schema**
(DTD or
XML Schema)

verified/enforced
by a **parser**

reusable
information
objects

Key concept: XML is about documents/data



XML as a document format

n In active use:

- XHTML: **E**xtensible **H**ypertext **M**arkup **L**anguage
 - reformulating HTML to ensure its survival in an XML world
- OFX: **O**pen **F**inancial **E**xchange
 - for exchanging financial data and transaction instructions between customers and their financial institutions (Microsoft Money, Quicken)
 - è IFX (**I**nteractive **F**inancial **E**xchange) for the financial services industry
- XFA: **X**ML **F**orms **A**rchitecture è XForms
XFDL: **E**xtensible **F**orms **D**escription **L**anguage
 - for the presentation, signing, transfer and processing of electronic forms
- Open **e**Book Publication Structure
 - for creating and managing reading material for portable e-books

n Under development:

- MathML: **M**athematical **M**arkup **L**anguage
 - for describing mathematical formulas (both presentation and content)

XHTML

n Extensible Hypertext Markup Language 1.0

n What?

- reformulation of HTML 4.01 as an application of XML
- comes in 3 flavors: strict / transitional / frameset
- W3C Recommendation (January 26, 2000)



n Extensible Hypertext Markup Language 1.1

n What?

- modular redefinition of XHTML (no deprecated elements, no frames)
 - e.g. XHTML Basic: text, images, basic forms and tables
- W3C Proposed Recommendation (April 06, 2001)

n Why?

- there will be no HTML 5.0 è subset/extend XHTML e.g. extended forms
- by 2002: 75% of all Internet access using Webphones, WebTV's, ...

XHTML

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">

<head>
<meta http-equiv="author" name="Author" content="Hans C. Arents" />
<title>A simple document</title>
</head>

<body>
<h1>A document you may want to read</h1>
<p align="left">A paragraph with
an image 
you may want to briefly look at.</p>
<hr />
<p>Another paragraph with some text
you may want to briefly read.</p>
</body>

</html>
```

XFDL

n Extensible Forms Description Language

n What?

- format for representing complex business fill-in forms
- W3C Note (September 02, 1998)

n Why?


- forms are the primary interfaces for doing e-business
 - simple and elegant method for customers to provide structured data to information systems
 - legal proof that a commercial transaction took place
- need for built-in data validation, security and auditability
 - integrated computations, input validation and help
 - multiple overlapping digital signatures
 - legally binding audit trails

XFDL

```
<?xml version="1.0"?>
<XFDL version="4.0.1">
  <vfd_title>REG 343</vfd_title>
  <page sid="PAGE1">
    <vfd_pagesize>letter</vfd_pagesize>
    <vfd_pagedpi>120</vfd_pagedpi>
    <vfd_printsize>8.0;10.5</vfd_printsize>
    <printsettings content="array">
      <borderwidth>0</borderwidth>
    </printsettings>
    <bgcolor content="array">
      <ae>239</ae><ae>239</ae><ae>239</ae>
    </bgcolor>
    <label sid="LABEL1">
      <itemlocation content="array">
        <ae content="array"><ae>absolute</ae><ae>253</ae><ae>52</ae></ae>
      </itemlocation>
      <value>APPLICATION FOR ORIGINAL REGISTRATION</value>
      <fontinfo content="array">
        <ae>Helvetica</ae><ae>14</ae><ae>bold</ae>
      </fontinfo>
      <justify>center</justify>
    </label>
  </page>
  ...
</XFDL>
```

Netscape
File Edit View Go Communicator Help

Submit Save Print Cancel Next Page



APPLICATION FOR ORIGINAL REGISTRATION

ALL applicants must complete SECTIONS A, B, E, F, AND G to register a vehicle.
If eligible for an exempt registration or an exempt license plate also complete SECTION C OR D.

Section A Vehicle Description	VEHICLE IDENTIFICATION NUMBER		MOTORCYCLE ENGINE NUMBER		
	CALIFORNIA LICENSE NUMBER	VEHICLE MAKE	YEAR MODEL	BODY TYPE	FUEL
		GM	1997	Pick-up	Diesel
	TYPE OF VEHICLE (CHECK ONE BOX)				
	<input checked="" type="checkbox"/> Auto	<input type="checkbox"/> Commercial (including pickup or truck)	<input type="checkbox"/> Trailer	<input type="checkbox"/> Commercial Trailer	
	<input type="checkbox"/> Motorcycle	<input type="checkbox"/> Off Highway			
	COMMERCIAL VEHICLE OR TRAILER	TRAILER COACH ONLY		CAMP TRAILER ONLY	
	Number of Axles <input type="text" value="2"/>	Length <input type="text"/>	(Inches)	Length <input type="text"/>	(Inches)
	Unladen Weight <input type="text" value="1200"/>	Width <input type="text"/>	(Inches)	Width <input type="text"/>	(Inches)
		Stored at: <input type="text"/>			

Section B Registration Questionnaire	Vehicle Information	
	The total cost or value of the vehicle must include the cost of the basic vehicle, value of any trade-in, and all accessories and leased equipment permanently attached.	
	NOTE: Cost does NOT include sales tax, insurance, finance charges, or extended warranty charges.	
	1. Complete either A or B:	DATE COST
A. Cost of vehicle purchased or acquired as a complete vehicle	32-05-99 <input style="background-color: yellow;" type="text"/>	
Cost of additions and/or alterations (include cost of installation)	This entry is invalid. Please try again.	
Total cost of vehicle	<input type="text" value="\$0.00"/>	

Form Ready

XML as a data format

n In active use:

- OSD: **O**pen **S**oftware **D**escription Format
 - for describing software packages and their dependencies to support automated software distribution (Marimba, Microsoft)
- WML: **W**ireless **M**arkup **L**anguage
 - for use in specifying content and user interface for narrowband devices, including personal digital assistants, cellular phones and pagers
- XML for B2B e-commerce
 - CBL: **C**ommon **B**usiness **L**anguage
 - cXML: **c**ommerce **X**ML
 - ebXML: **e**lectronic **b**usiness **X**ML

n Under development:

- SOAP: **S**imple **O**bject **A**ccess **P**rotocol
 - for communication over the Internet between software objects developed using different languages and different programming models



n Wireless Markup Language



n What?

- part of WAP (Wireless Application Protocol)
- for specifying content and user interface for narrowband wireless devices

n Why?

- wireless networks and devices have specific needs and requirements not addressed by existing Internet technologies
- WAP uses plain Web HTTP 1.1 servers
- WML + WMLScript support
 - user interface metaphor: cards with hyperlinks
 - state management, input validation, extensions to enable pushed content, user profiling, ...

WML

Navigation

```
<WML>
  <CARD>
    <DO TYPE="ACCEPT" LABEL="NEXT">
      <GO URL="#eCard"/>
    </DO>
    Acme Inc.<BR/>Directory
  </CARD>
```

Card

Variables

```
<CARD NAME="eCard">
  <DO TYPE="ACCEPT">
    <GO URL="/send?how=${type}"/>
  </DO>
```

Deck

Input Elements

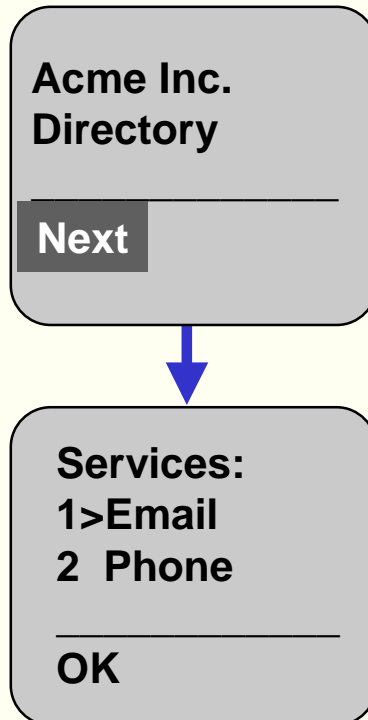
```
Services:
<SELECT KEY="type">
  <OPTION VALUE="e">Email</OPTION>
  <OPTION VALUE="p">Phone</OPTION>
</SELECT>
```

```
</CARD>
</WML>
```

WML

```
<WML>
  <CARD>
    <DO TYPE="ACCEPT" LABEL="Next">
      <GO URL="#eCard"/>
    </DO>
    Acme Inc.<BR/>Directory
  </CARD>

  <CARD NAME="eCard">
    <DO TYPE="ACCEPT">
      <GO URL="/send?how=$(type)"/>
    </DO>
    Services:
    <SELECT KEY="type">
      <OPTION VALUE="e">Email</OPTION>
      <OPTION VALUE="p">Phone</OPTION>
    </SELECT>
  </CARD>
</WML>
```



SOAP

n Simple Object Access Protocol

n What?

- an open, extensible way for applications to communicate over the Web using XML-based messages, regardless of what operating system, object model, or programming language they use
- Microsoft initiative, but has been submitted as a potential standard to the IETF (Internet Engineering Task Force)

n Why?

- to define a distributed object protocol that allows RPCs (remote procedure calls) between clients and servers
 - can pass through Internet firewalls
 - not tied to a specific component technology
- exchange of request/response messages as serialized XML
 - alternative for DCOM and CORBA protocols / integral part of Visual Studio 7

SOAP

POST /StockQuote HTTP/1.1

Host: www.stockquoteserver.com

Content-Type: text/xml

Content-Length: nnnn

SOAPMethodName: urn:schemas-quote-com#GetLastTradePrice

```
<SOAP:Envelope xmlns:SOAP="urn:schemas-xmlsoap-org:soap.v1">
  <SOAP:Body>
    <m:GetLastTradePrice
      xmlns:m="urn:schemas-quote-com">
      <symbol>DIS</symbol>
    </m:GetLastTradePrice>
  </SOAP:Body>
</SOAP:Envelope>
```

REQUEST

HTTP/1.1 200 OK

Content-Type: text/xml

Content-Length: nnnn

```
<SOAP:Envelope xmlns:SOAP="urn:schemas-xmlsoap-org:soap.v1">
  <SOAP:Body>
    <m:GetLastTradePriceResponse xmlns:m="urn:schemas-quote-com">
      <return>34.5</return>
    </m:GetLastTradePriceResponse>
  </SOAP:Body>
</SOAP:Envelope>
```

RESPONSE

XML for B2B e-commerce

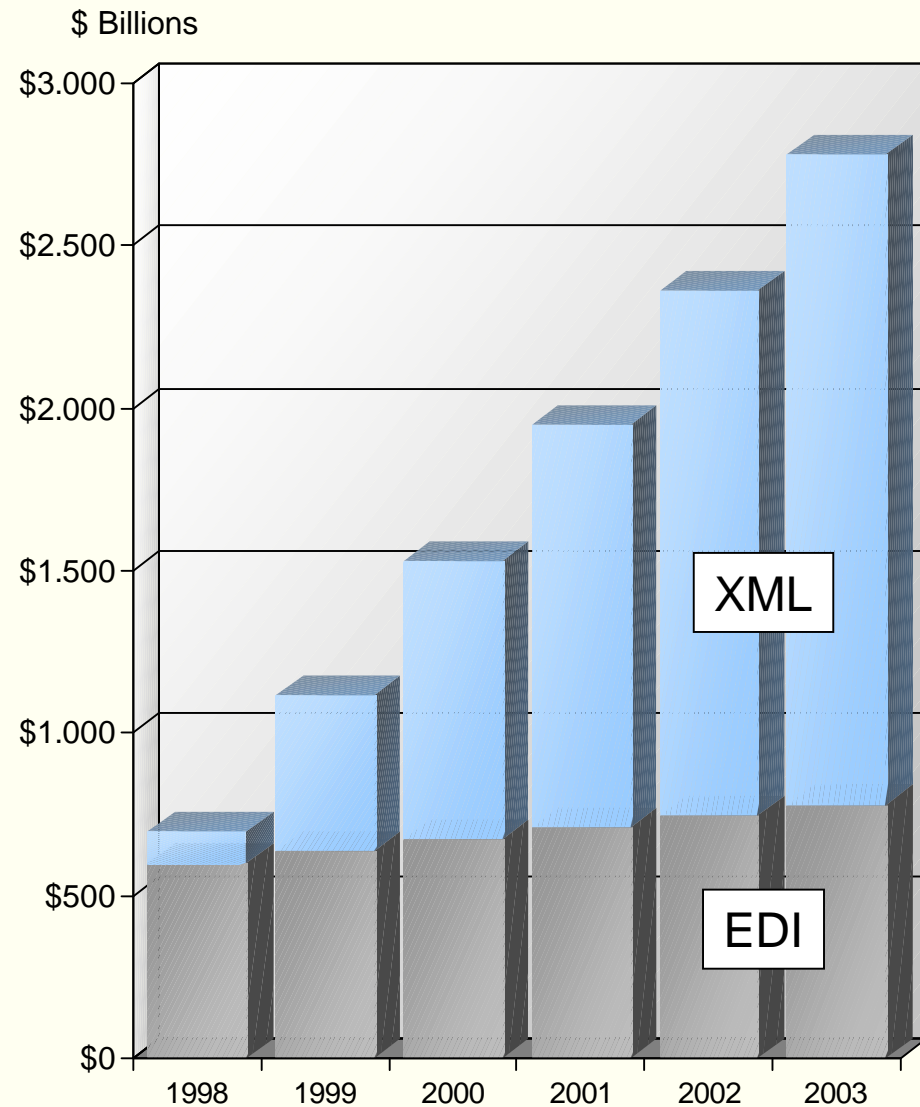
n 24% of all B2B purchasing will be online by 2003 (up from 7% today)

n EDI:

- û complex and expensive
- û for big companies
- û batch oriented, data only
- û not evolving at Internet speed

n XML:

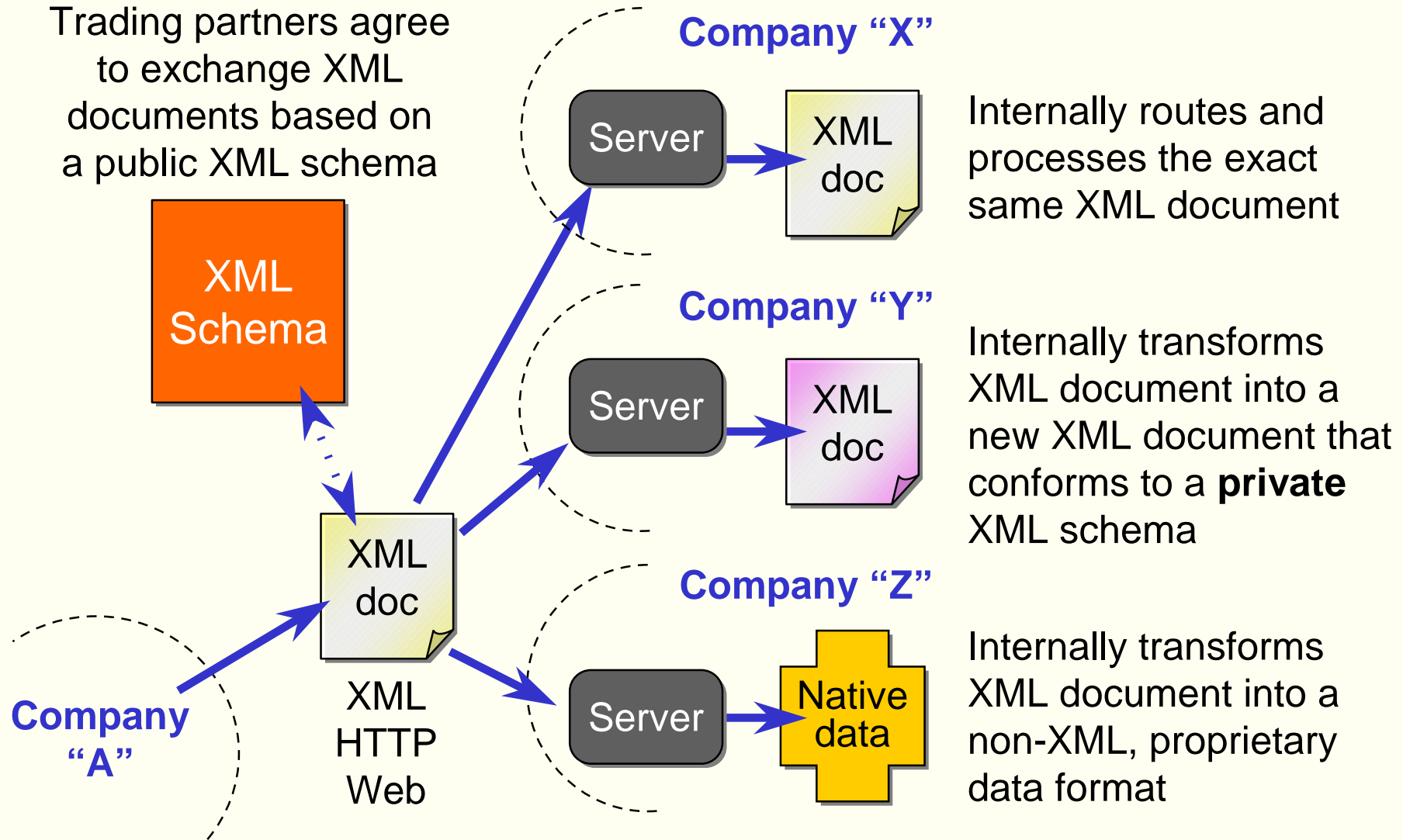
- ü cheap and simple
- ü for small companies
- ü interactive and real-time
- ü evolving at Internet speed



Source: The Boston Consulting Group (BCG)

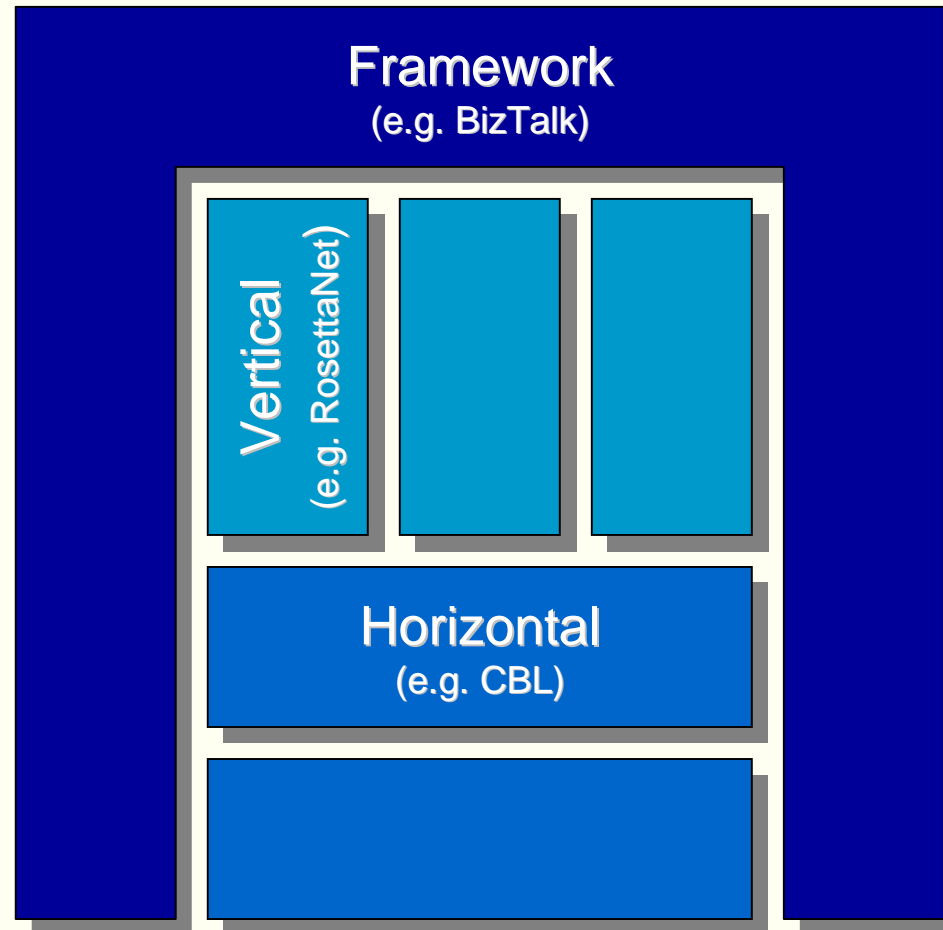
XML for B2B e-commerce

Trading partners agree to exchange XML documents based on a public XML schema



XML for B2B e-commerce

- n Major initiatives to define *the* XML vocabulary for an industry or group of industries
- n Vertical initiatives
 - sector-specific
- n Horizontal initiatives
 - cross-sector
- n Framework initiatives
 - cross-everything



XML as a metadata format

n In active use:

- CDF: **C**hannel **D**efinition **F**ormat
 - download directives for push channels (Microsoft Internet Explorer)
- RDF: **R**esource **D**escription **F**ramework
 - metadata about Web pages/program objects (Netscape Navigator)
- ICE: **I**nformation & **C**ontent **E**xchange
 - data format to facilitate the process of automatically exchanging, updating, supplying, and controlling electronic information assets
- P3P: **P**latform for **P**rivacy **P**references
 - enable Web sites to express their privacy practices and enable users to exercise preferences over those practices

n Under development:

- PICS-NG: **P**latform for **I**nternet **C**ontent **S**election - **N**ext **G**eneration
 - rating for the content of Web site pages

CDF

n Channel Definition Format

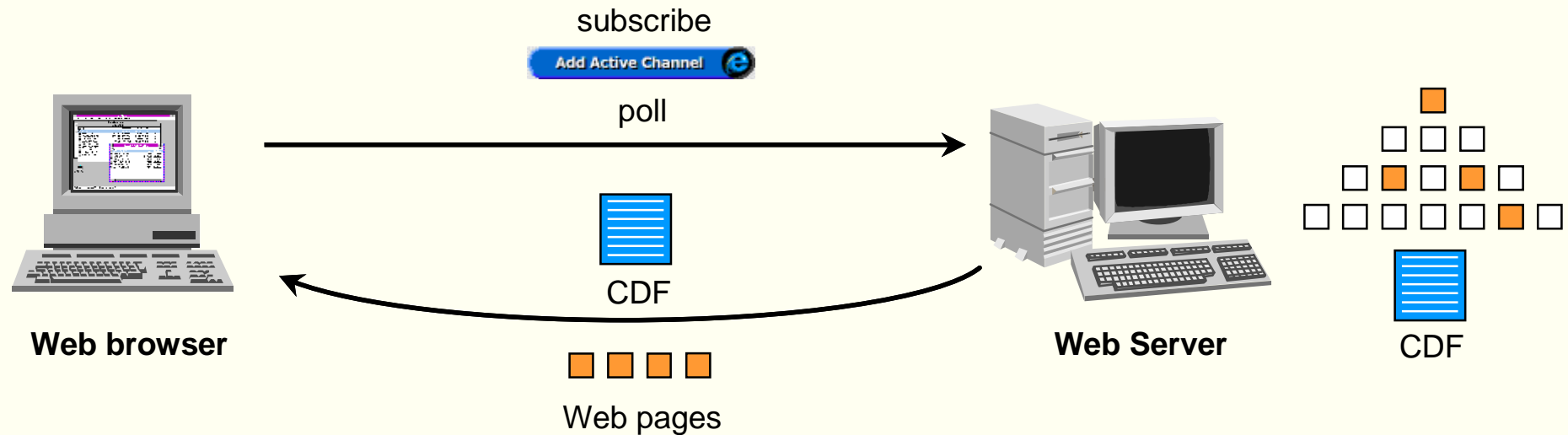
n What?

- offer frequently updated collections of information, or channels, from any Web server for automatic delivery to any Web client
- W3C Note (March 09, 1997)

n Why?

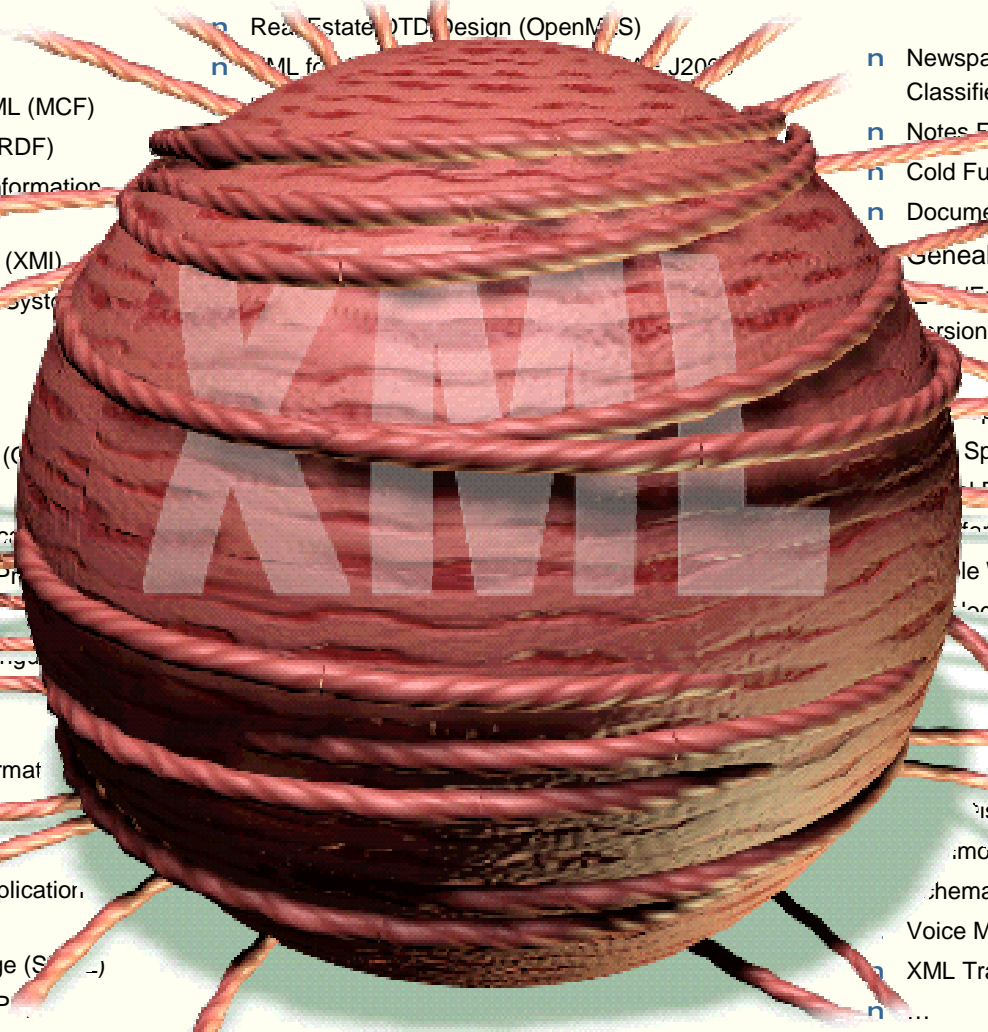
- broadcast personalized, frequently updated news to Web users
- Microsoft initiative to support "Webcasting" or "push" technology
- very first example of real business use of XML on the Internet
- heavily hyped in IE 4.0, almost disappeared in IE 5.0

CDF for information push



- ☒ client subscribes to Web site
 - server sends CDF file with “smart pull” directives
- ☒ client polls Web site for push of changed content
 - server sends changed pages according to CDF directives

XML is the ASCII of the 21st century

- 
- n Channel Definition Format (CDF)
 - n Web Collections using XML
 - n Meta Content Framework Using XML (MCF)
 - n Resource Description Framework (RDF)
 - n The Australia New Zealand Land Information Council (ANZLIC) - Meta
 - n XML Metadata Interchange Format (XMI)
 - n Educom Instructional Management System Project (IMS)
 - n Structured Graph Format (SGF)
 - n Web Standards Project (WSP)
 - n Open Software Description Format (OSDF)
 - n Extensible Log Format (XLF)
 - n Wireless Markup Language Specification
 - n HTTP Distribution and Application Protocol
 - n Chemical Markup Language
 - n Bioinformatic Sequence Markup Language (BSML)
 - n Virtual Hyperglossary (VHG)
 - n Weather Observation Definition Format
 - n Open Financial Exchange (OFX)
 - n Open Trading Protocol (OTP)
 - n A Markup Language for the FIX Application, Message Layer (FIXML)
 - n Signed Document Markup Language (SDML)
 - n Bank Internet Payment System (BIP)
 - n Real Estate OTD Design (OpenMIS)
 - n XML for J200
 - n Newspaper Association of America (NAA) - Classified Ads Format
 - n Notes File Format (NFF)
 - n Cold Fusion Markup Language (CFML)
 - n Document Content Description for XML (DCD)
 - n Genealogical Data in XML (GedML)
 - n Extensions for Distributed Authoring and Versioning on the World Wide Web (WEBDAV)
 - n Text Markup Language (TML)
 - n Document Markup Language (DML)
 - n Speech Markup Language (JSML)
 - n Virtual Reality Modeling Language (VRML)
 - n Workflow Management [NIST]
 - n Simple Workflow Access Protocol (SWAP)
 - n Thematic Markup Language (ThML)
 - n Forms Description Language (XFDL)
 - n Business Markup Language (BHTML)
 - n Open Service Protocol (OSP)
 - n Web Distribution and Exchange (WDDX)
 - n Global Business Library (CBL)
 - n Schema for Content-oriented XML (SOX)
 - n Voice Markup Language (VoxML)
 - n XML Transfer Protocol (XMLTP.Org)

For an overview: <http://www.schema.net/>

XML is the universal Web format

Layers of Web-based information exchange

Protocol: set of XML messages

How it is to be used

Vocabulary: DTDs / XML schemas

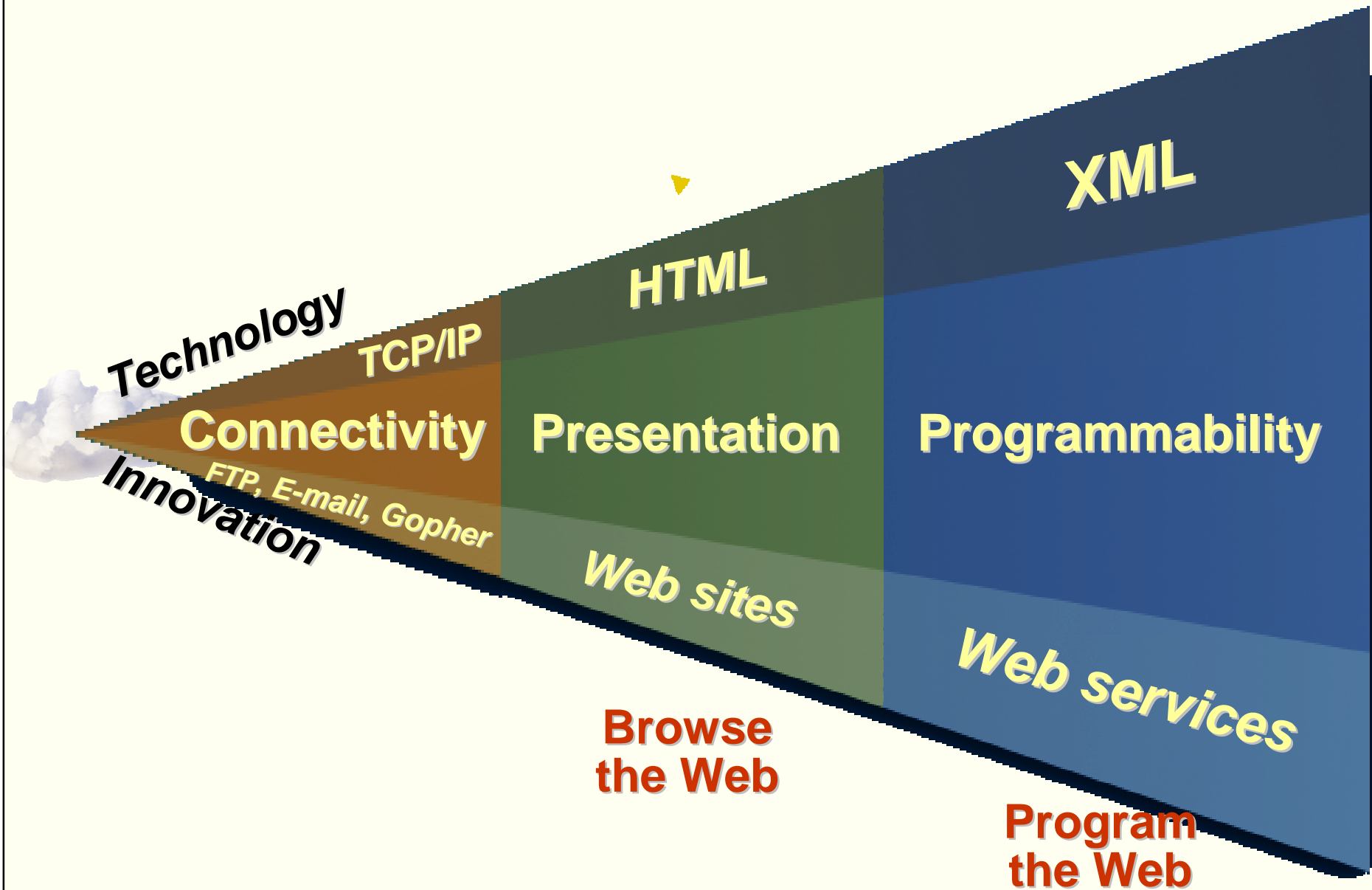
What it means

Format: XML syntax

What it is

Transport: HTTP, SMTP, FTP, ...

The 3rd generation Internet



Conclusions

n The 5 I's of XML

INFORMATION

data and documents

INTEROPERABILITY

sharing information
opening up applications

INTEGRATION

bringing together information
linking together applications

INDEPENDENCE

platform, application,
programming language,
data model, delivery device

INTERNATIONAL

designed to use Unicode

