

# XML Tools and Trends: Why You Should Be Leading the Charge

---

**InterChange**

**2001**

**October 15, 2001**



**© Leonor Ciarlone  
Senior Consultant, CAP Ventures  
and Bob Boeri,  
Information Architect  
Forefield Inc.**

Version October 14, 2001

# Presentation Overview

---

- ◆ **Recap: XML, HTML, and XHTML**
- ◆ **XML Impact from a Developer's Perspective**
- ◆ **XML Impact from an Analyst Perspective**
- ◆ **XML Impact on Your Career**
- ◆ **Learn more**

# What is XML?

## The Details.

---

XML is a **subset** of SGML.

- ◆ A W3C Recommendation as of February, 1998; begun in October 1996.
- ◆ A flexible, platform-independent, vendor neutral, meta-language.
- ◆ Designed to work at Web-speed.
- ◆ Backed by all major vendors and standards groups throughout the world.
- ◆ A lot less complex than SGML.
- ◆ Has spawned dozens of related standards.

# Structured Document Puzzle Pieces

---

**Three fundamentally different,  
but complementary concepts:**

- ❖ **DTD (content model)**
- ❖ **Style (presentation)**
- ❖ **Instance (the content, or document)**



# Understanding the Content Model: Elements, Attributes and Entities

---

**HTML**

**XML**

**Element**

**<TABLE>**

**<Warning>**

**Attribute**

**color=fuschia**

**type=critical**

**Entity**

**&quote  
(“)**

**boilerplate, e.g., “pubstatus”  
“Pre-release version”**

**!impact:** Help guide the definition of elements, attribute and entities and you'll gain a leadership role.

# How Does the Content Model (DTD) Relate to the Document?

---

- ◆ An “*instance*” of an XML document associated with a DTD is called a “*valid document*.”
- ◆ A valid XML document specifies its associated DTD within the markup.
- ◆ A parser can validate, or compare, a valid XML document against its DTD.



# How Style Fits In: Making Content Presentable

---

XML separates content from style. XSL is the XML-based language for expressing stylesheets. A style sheet describes:

- ❖ How each element should look.
- ❖ How elements are distinguished from one another.
- ❖ How the entire instance is visually presented once rendered to a printer, a screen, or a file.



**!impact:** Separating style from content enables a manageable, consistent process to apply formatting to each element (tag).

# How Does the Content Model Relate to the Stylesheet?

---

An “*instance*” of an XML document can be associated with an XSL or CSS style sheet. XSLT, a powerful XML-based transformation language, can transform or manipulate XML instances to achieve results such as:

- ◆ **Re-ordering**, can both format and re-sequence content for different users or applications.
- ◆ **Automatic text generation**, including page numbers and section headers.
- ◆ **Dynamic re-formatting**, based on the elements of related documents.
- ◆ **Application Processing**, server- and client side
- ◆ **Conversion**, such as XML to other XML, HTML, PDF, ...

# XML: It All Comes Together

---

The three puzzle pieces of XML:

- ◆ A well-formed instance, or document.
- ◆ A DTD if the document must be valid.
- ◆ One or more XSL stylesheets.



**!impact:** XML enables multi-channel publishing.

# HTML Works for Me.

## Why Isn't it “Good Enough?”

---

Sometimes it is. But:

- ◆ They're not *\*your\** tags.
- ◆ Single-sourcing from HTML is practically impossible.
- ◆ Presentation-based design is *\*not\** the only way to go. (And HTML's presentation isn't that good.)
- ◆ Lot's of loopholes!
- ◆ HTML linking won't get better than it is.
- ◆ HTML will never be good for extremely long pages.
- ◆ HTML is now obsolete; XHTML is now the “HTML” standard.

# XHTML:

## The Bridge from HTML to XML

---

Get Ready for XHTML By Cleaning Up Your HTML:

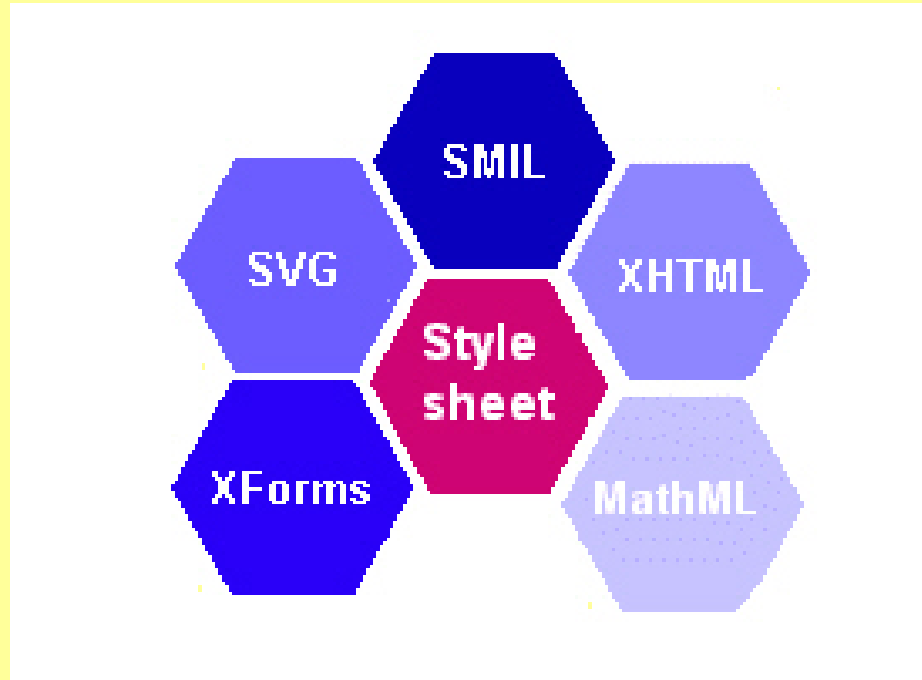
- ❖ All HTML must be lower-case
- ❖ All attribute values must be quoted.
- ❖ All non-empty tags must be terminated: `<p>...</p>`
- ❖ Elements must nest, not overlap: `<p> .. <b>....</b>...</p>`
- ❖ Empty elements have modified termination: `<br />`

**!impact:** As browsers and new kinds of hardware emerge (e.g., Palmtops, wireless), you'll be ready. As XML usage grows, you can more easily upgrade to XML.

# XHTML:

## Works with a cluster of standards

---



**!impact:** XML, via XHTML, is now touching all areas of publishing: Graphics, multimedia, vocabularies, structured data.

# What Makes XML Different From HTML?

---

- ◆ **Extensible:** define your own tag set.
- ◆ **Intelligent:** semantic and structured.
- ◆ **Flexible:** design documents with simple or complex structures.
- ◆ **Multi-faceted:** suited for re-purposing, re-use and programmability.
- ◆ **Maintainable:** you define the tag set, you maintain the tag set. (As opposed to Microsoft, Netscape, or even the W3C.)

# XML-Driven Standards

---

- ◆ 1998-99: Styling (CSS), applications (DOM), graphics and multimedia (SVG, SMIL), meaning (RDF), wireless (WML).
- ◆ 2000: XHTML, DOM, accessibility, querying
- ◆ 2001: Richer modeling (Schemas), meaning (“semantic web”), linking, graphics and multimedia.
- ◆ Long lead time from standards to tools to widespread acceptance.

**!impact:** Although the process is slow, text and multimedia content are becoming just more data types. Understanding XML becomes critical.

# Presentation Overview

---

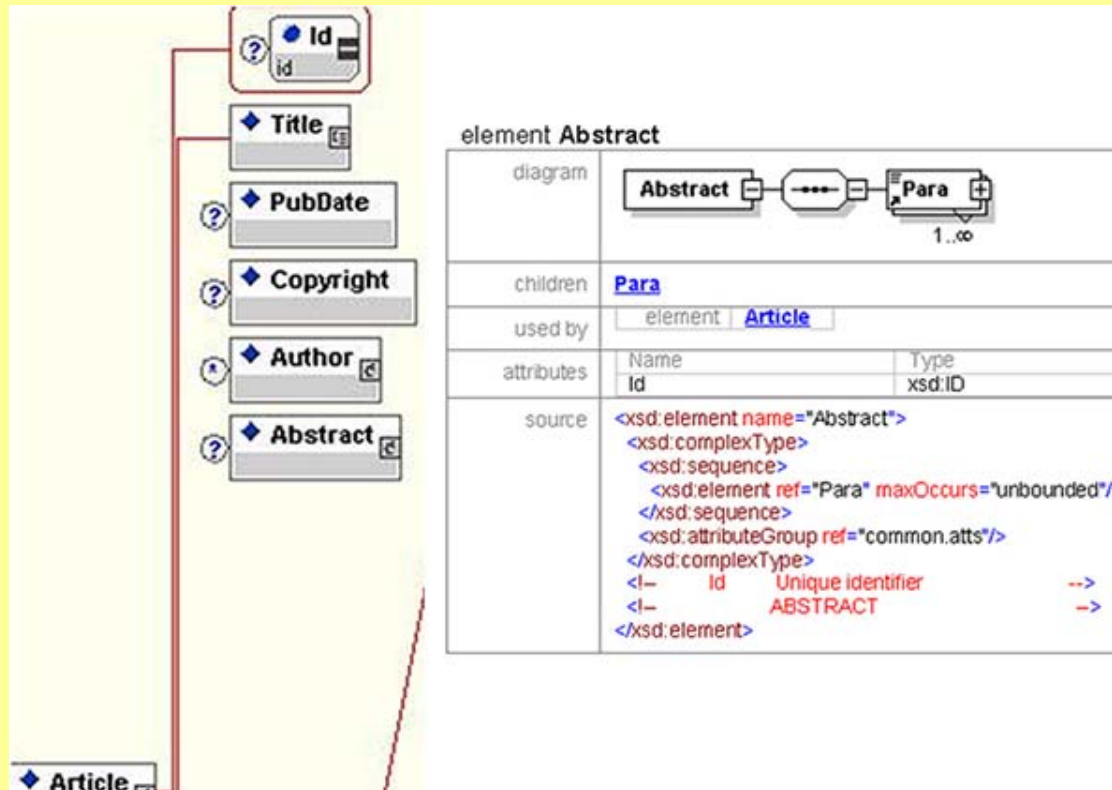
- ◆ **Recap: XML, HTML, and XHTML**
- ◆ **XML Impact from a Developer's Perspective**
- ◆ **XML Impact from an Analyst Perspective**
- ◆ **XML Impact on Your Career**
- ◆ **Learn more**

# XML Tools and Suites

---

- ◆ **Information Modeling:** Tibco's Turbo XML, XML Spy
- ◆ **Native XML Authoring:** SoftQuad XMetaL, Arbortext Epic, Turbo, Spy.
- ◆ **Hybrid XML Authoring:** FrameMaker+SGML, many Word add-on products.
- ◆ **XSLT Development:** Spy, Transform, Stylus
- ◆ **XML Parsing and XSL Rendering:** MSXML 3 and higher (IE 6), Exoterica.
- ◆ **SVG Support:** Adobe Illustrator, CorelDraw, Real (soon)
- ◆ **SMIL Authoring and Delivery:** RealPlayer and RealSlideShow, RealProducer, GRiNS

# XML Modeling: Spy and Turbo



- ◆ Graphical modeling (Turbo)
- ◆ Dictionary-like modeling (Spy)


# XSLT Development: Excelon Stylus

The screenshot displays the Excelon Stylus 3 development tool interface. The top window shows the XSLT code for a video listing page. The code includes a header with a meta tag for content type, a title 'VideoCenter - Video Listing', and a CSS style block for the body. The body content is partially visible, showing a table structure. The right-hand pane shows a tree view of the resulting XML structure, with elements like 'video\_template', 'title', 'genre', 'choice', 'rating', 'user\_rating', 'summary', 'details', 'year', 'director', 'studio', and 'runtime'. Below the code editor, the 'Preview' section shows the rendered HTML output. The preview includes the 'VideoCenter' logo, a 'Video Listing' heading, and a table with columns for Title, Summary, Genre, Format, and Price. The first row of the table lists 'Enemy of the State' with a summary, genre of 'action', and two price options for different formats: \$16.99 for VHS and \$29.99 for DVD.

```
<html>
<head>
<meta http-equiv="Content-Type"
content="text/html; charset=UTF-8" />
<title>VideoCenter - Video Listing</title>
<style type="text/css">
<![CDATA[
    body
    { color: #ffffff;
background: #000000;
font-family: Helvetica, Aria...
```

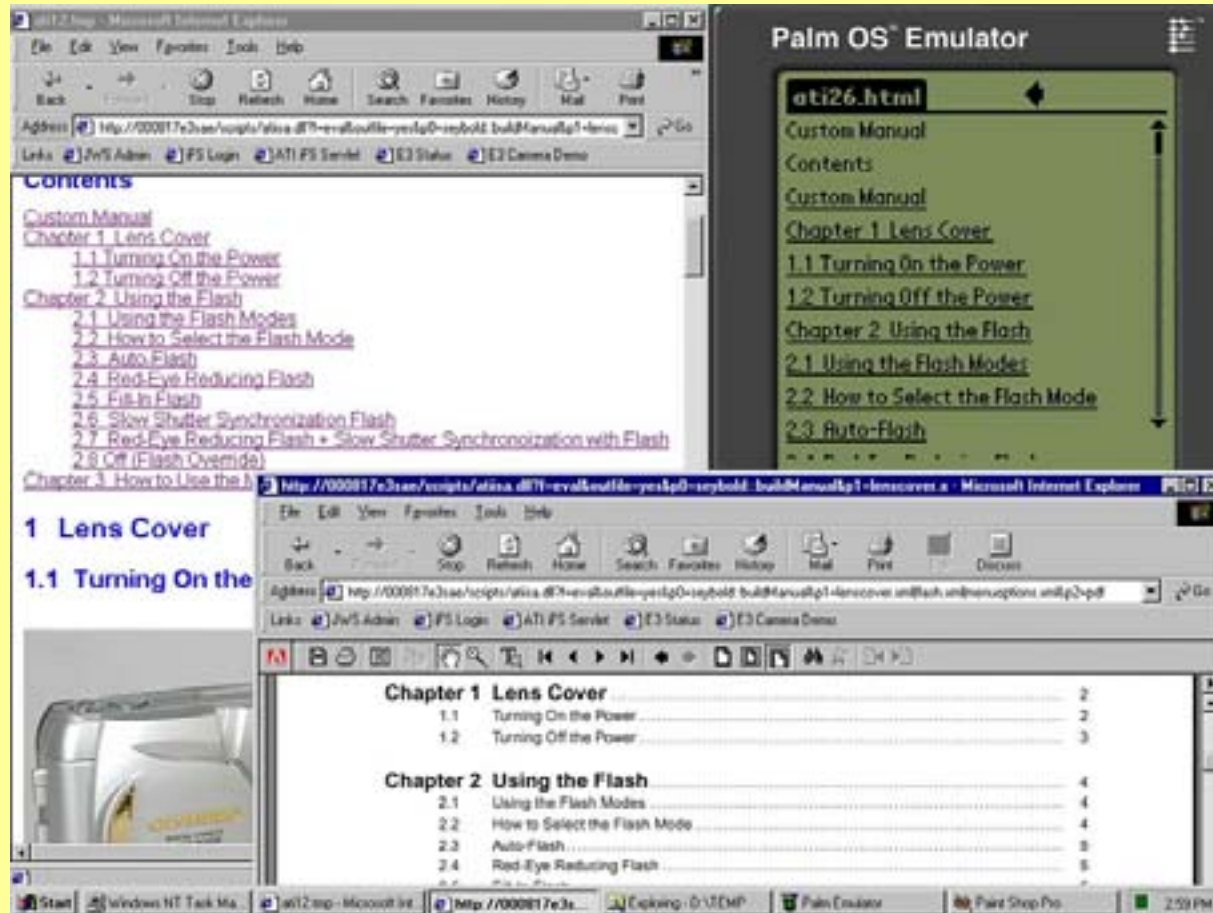
VideoCenter

Video Listing

Title	Summary	Genre	Format	Price
Enemy of the State	 <p>After a chance meeting with an old pal, Robert Deal finds himself in possession of a disk that contains evidence of an</p>	action	VHS DVD	\$16.99 \$29.99

- ◆ Excelon Stylus 3-widowed development tool
- ◆ XML to HTML, XML to XML

# Native XML Authoring: Arbortext Epic



- ◆ Multi-channel output via XSLT (e.g., HTML, PDF, and Palm)

# Converting Proprietary Content to XML

---

Conversions options and tools are available, but none are “out of the box” solutions.

- ◆ **HTML Conversion:** Tidy, Perl. 80% solutions!
- ◆ **IDEs:** Spy and Turbo also offer conversion aids from Word, Excel, Access, etc.
- ◆ **Word add-ons:** WorX SE (Hypervision), S4Text (i4i), eXstyles (Inera), Xfinity Author wX (B-Bop)
- ◆ **Unstructured to structured conversion:** RTF2XML, RTF2FO (Novasoft), Logictran RTF Converter (Logictran), Outside In (Stellent), Xtent (Liquent), xPower Transform Suite (BDI Systems), Text Café (Texterity), Omnimark

# XML Tools Trends Today

---

- ◆ **Price:** Tibco's XML Authority increased \$50 to \$200, XML Spy about \$250, and pricing favors buying one of their bundled packages. SoftQuad's XMetaL version 2, still \$499. Excelon Stylus, \$200.
- ◆ **Integration:** Tibco Turbo and Canon; Spy "IDE". Stylus "studio."
- ◆ **Name Recognition:** Fewer and fewer "brand name" vendors.
- ◆ **Microsoft Factor:** Microsoft offers surprisingly quality upgrades, usually free. Internet Explorer provides increasing support for XML, and MSXML (on the client or the client or server) provides free parsing and XSLT processing.

# Lessons Learned On the Job

---

- ◆ **Remember your experiences in software development organizations.**
  - ❖ **Developing models, style sheets or transformation files is software development.**
    1. **Manage versions**
    2. **Document changes**
    3. **Develop individual and integrated tests**
- ◆ **We're in the early stages.**
  - ❖ **Think "Help development in the early 90s."**
  - ❖ **XML tools are new, somewhat immature and inexpensive (but growing expensive).**

# Lessons Learned On the Job

---

- ◆ **XML is complex!**
  - ❖ XML started out as a 30 page spec.
  - ❖ My favorite XML text is the 1015 page XML Bible.
  - ❖ XSL started out as a much more complex spec, and now is three (XSL, XSLT, XPATH).
  - ❖ My favorite XSLT book is the XSLT Programmer's Reference. It is 780 pages long.
- ◆ **You'll need a content management strategy.**
  - ❖ As with other software development, version management.
  - ❖ XML is content-oriented, not file-oriented.

# Lessons Learned On the Job

---

- ◆ **Expect culture shock**
  - ❖ **Apply past experiences, but you'll also be in new territory.**
  - ❖ **Expect resistance from all quarters.**
  - ❖ **Expect XML projects snags .**
  - ❖ **Find yourself a mentor (and discussion group) to share anguish, questions, and insights.**
- ◆ **Persist!**
  - ❖ **XML really is the wave of future content.**
  - ❖ **XML will pay dividends.**
  - ❖ **XML evangelists can become martyrs.**

# Presentation Overview

---

- ◆ **Recap: XML, HTML, and XHTML**
- ◆ **XML Impact from a Developer's Perspective**
- ◆ **XML Impact from an Analyst Perspective**
- ◆ **XML Impact on Your Career**
- ◆ **Learn more**

# Understanding Business Trends

---

- ◆ **The Web changes how organizations develop and market products**
  - ❖ Increased competition
  - ❖ Increased product complexity
  - ❖ Shorter product lifecycles
  - ❖ Movement towards a global marketplace
- ◆ **The Web changes how organizations conduct business**
  - ❖ Geographically dispersed and remote employees
  - ❖ Shifting perceptions: information management as primary business driver rather than *supporting* business function
  - ❖ Simultaneous evolution of customer requirements, processes and technology
  - ❖ Time to market is *critical* in order to survive and thrive

# Trends Affect the Role of Content

---

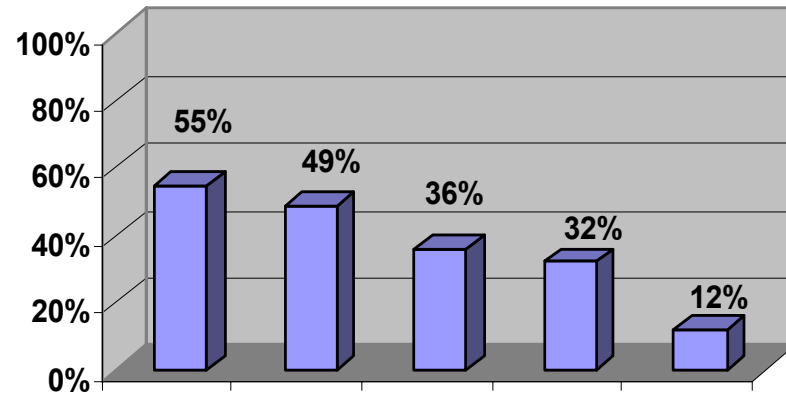
- ◆ **The characteristics associated with content have changed**
  - ❖ *Volume* expands
  - ❖ *Velocity* increases
  - ❖ *Definition* expands
- ◆ **Customer/user expectations about information access have changed**
- ◆ **Competitive positioning across numerous verticals has changed to include information as an “edge”**

**!impact:** The pursuit of “intellectual capital” is back in vogue.

# A Resurgence of Knowledge Management

- Key Focus on Productivity
- Capture Corporate Knowledge (/Assets)
- Leverage Expertise
  - Human Resources
  - Know-How
- Keep Knowledge From “Walking Out the Door”
- Serendipity: Discovering Information

Business Drivers for Dynamic Content Applications



# As a Result:

---

Organizations embrace the “notion” that content-driven systems and applications are *essential* to successful eBusiness or eCommerce initiatives.

**!impact:** Content becomes a corporate asset.

# What is an Asset?

---

**Asset:** The entire property of a person or company that may be used to pay debts. 2. Advantage Resource: Merriam-Webster Dictionary

## **Web Content:**

- HTML
- XML
- Graphics/Images
- PDF
- Metadata
- Source Code/Scripts

## **Rich Media:**

- Audio
- Streaming Video
- Animation

## **Documents:**

- Word Processing
- DTP
- PDF
- Images
- Reports

# Investment in Content is Now Key

---

**Organizations are focusing on technologies that enable:**

- ❖ **Ease of “knowledge capture”**
- ❖ **Collaboration**
- ❖ **Productivity increases and process control, e.g. effective creation, management, and publishing**
- ❖ **Information re-use and re-purposing**
- ❖ **Solidified views of and access to disparate information resources**
- ❖ **Personalization**
- ❖ **Multi-channel publishing**
- ❖ **User or customer-driven self-service**

# Issues With Getting to Content-Driven Applications

- ◆ **Integration** with overall Web and technology architecture
- ◆ **Conversion** and normalization of content from internal and external sources
- ◆ **Perception**; getting contributors to think units, not documents
- ◆ **Personalization**: supporting the needs of customers whose expectations change based on role and context



*The predominance of Website content is “locked” in unstructured HTML pages*

# As a Result:

---

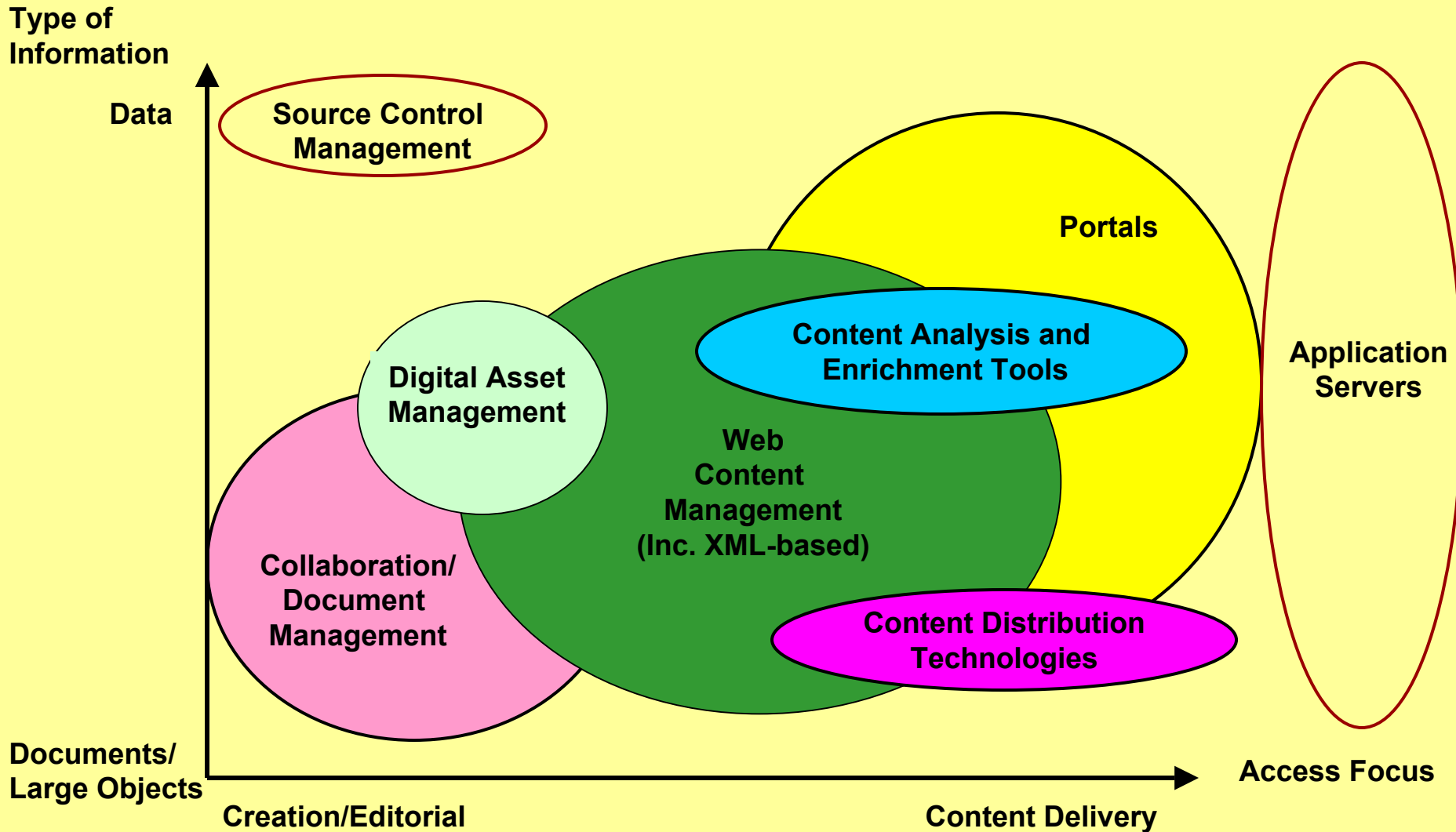
Vendors have evolved in response to business trends, market drivers, and the overall need for organizations to “get to the Web”

# The Language of The Industry

- **E-content** (“content-rich”, “intelligent content”, “dynamic content”)
- **E-business, E-commerce**
- **E-publishing**
- **E-catalogs**
- **Web Publishing**
- **Content Management**
- **Document Management**
- **Portals**
- **Digital Asset Management**
- **Information Commerce**
- **B2B, B2C Applications**
- **Customer Relationship Management (CRM)**
- **Knowledge Management**
- **Electronic Delivery Servers**
- **Content Aggregation Tools**
- **Categorization**
- **Content Profiling**
- **Personalization**
- **Digital Rights Management**



# A Simplified (☺) Technology Landscape



# What Does This Have To Do With XML?

---

**XML has taken a leading role in discussions on how to achieve successful content-driven applications in support of eBusiness or eCommerce initiatives**

- ❖ **XML-based technology standards from the W3C, ISO, OASIS, and other organizations**
- ❖ **XML-based vocabularies from industry-driven organizations**
- ❖ **Numerous acquisitions, partnerships or direct development of XML technologies from dynamic content software vendors**
- ❖ **Continued movement of platform players (e.g. IBM, Microsoft, Oracle, Sun, etc.) into the dynamic content market space**

# Why Has XML Had Such a Big Impact?

---

- ◆ **XML is:**
  - ❖ **Vendor-neutral**
  - ❖ **Extensible**
  - ❖ **“Programmable”**
  - ❖ **Intelligent at discrete levels**
  - ❖ **International“izable”**
- ◆ **XML is a technology well-suited for:**
  - ❖ **Personalization**
  - ❖ **Multi-channel publishing**
  - ❖ **Content re-use and re-purposing**
  - ❖ **Separation of process from content**
  - ❖ **Application to application data exchange**

# How Has XML Influenced Content-Based Systems?

---

- ◆ **Distance learning or “eLearning” market players are offering “Reusable Learning Objects”.**
- ◆ **Digital asset management and eCatalog players are offering “product-driven personalization”**
- ◆ **Portal players are including XML-driven content publishing or management capabilities**
- ◆ **Content Management Players are offering multi-channel publishing solutions. Samples (not an endorsement!)**
  - ❖ **Arbortext**
  - ❖ **Broadvision**
  - ❖ **Crystal Software**
  - ❖ **Documentum**
  - ❖ **Eprise (now divine)**
  - ❖ **Intervoven**
  - ❖ **SiberLogic**
  - ❖ **Percussion**
  - ❖ **Stellent (IntraNet Solutions)**
  - ❖ **Worldweb.net (now Starbase)**
  - ❖ **XyEnterprise**

# XML is Not a Panacea

---

XML does not guarantee successful implementation of principles, philosophies, or methodologies such as “Knowledge Management,” “The Semantic Web” and “Single Sourcing.”

**!impact:** XML enables those ideas and processes to take shape.

# Presentation Overview

---

- ◆ **Recap: XML, HTML, and XHTML**
- ◆ **XML Impact from a Developer's Perspective**
- ◆ **XML Impact from an Analyst Perspective**
- ◆ **XML Impact on Your Career**
- ◆ **Learn more**

# Helping Your Organization Get to Intelligent Content

---

- ◆ **Core requirements are always human-focused.**
- ◆ **Technology --even XML-- is only an enabler.**
- ◆ **Ask yourself:**
  - ◆ **Do content creation standards align with business goals?**
  - ◆ **Do you have a metadata strategy in place?**
  - ◆ **How is information captured and categorized and how can your information lifecycle adapt?**
  - ◆ **How do you account for and promote standards across different levels of content contributors (with competing interests)?**
  - ◆ **How do you get from here (legacy files) to there (structured content), while keeping the business running?**

# The Impact XML Has On Your Job

---

- ◆ **Your role impacts the entire content lifecycle:**
  - ❖ **Creation: Needs analysis**
  - ❖ **Creation: Information analysis**
  - ❖ **Creation: User analysis**
  - ❖ **Management: Information architecture**
  - ❖ **Management: Information qualification**
  - ❖ **Delivery: Information outputs**
  - ❖ **Delivery: Personalization, re-use and re-purposing**
- ◆ **Your career choices can take many paths:**
  - ❖ **Expanded roles in software development companies in various markets (e.g. eLearning, CM, portal, etc.)**
  - ❖ **Content-driven industries such as publishing, financial services, education, etc.**
  - ❖ **IT organizations across multiple verticals**

# Assessing XML and your Career



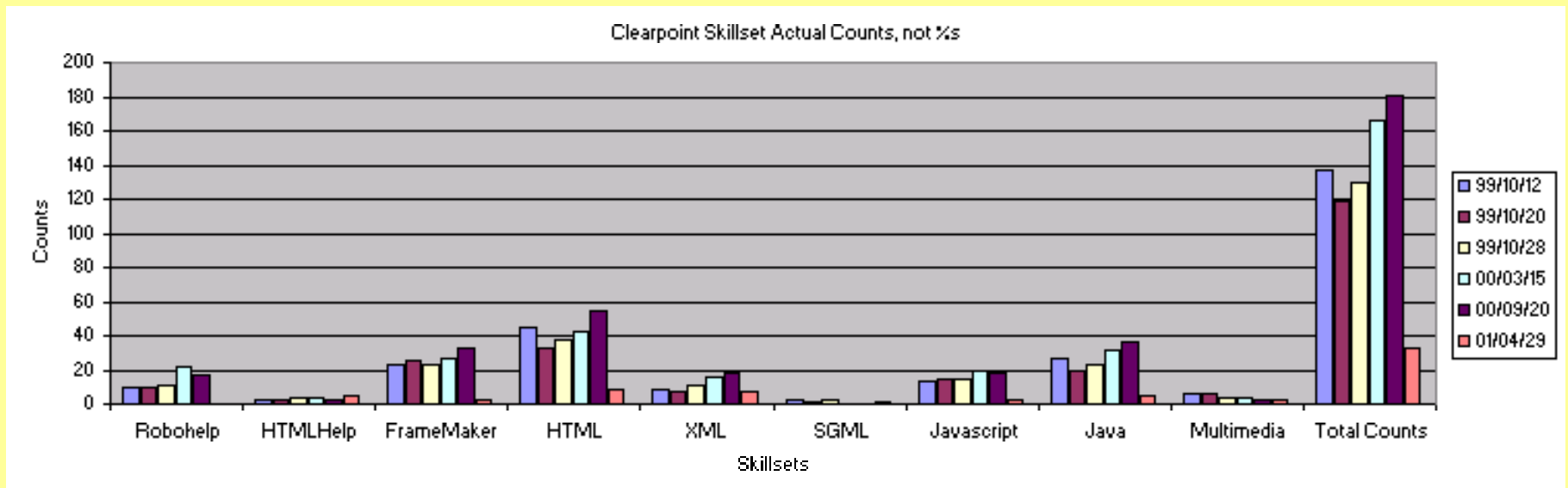
- ◆ Your knowledge of HTML will help.
- ◆ XML is here to stay.
- ◆ The future of publishing is content-oriented.
- ◆ XML publishing is just getting started.
- ◆ It's not too early to learn.
- ◆ There are abundant learning resources.
- ◆ Tools are relatively inexpensive.



- ◆ XML is complex (especially when you include companion standards).
- ◆ There are no easy routes from legacy word processor and HTML content to XML.
- ◆ Tools are somewhat immature.
- ◆ You may have a 'sell job.'

# Demand for XML Skills Grows

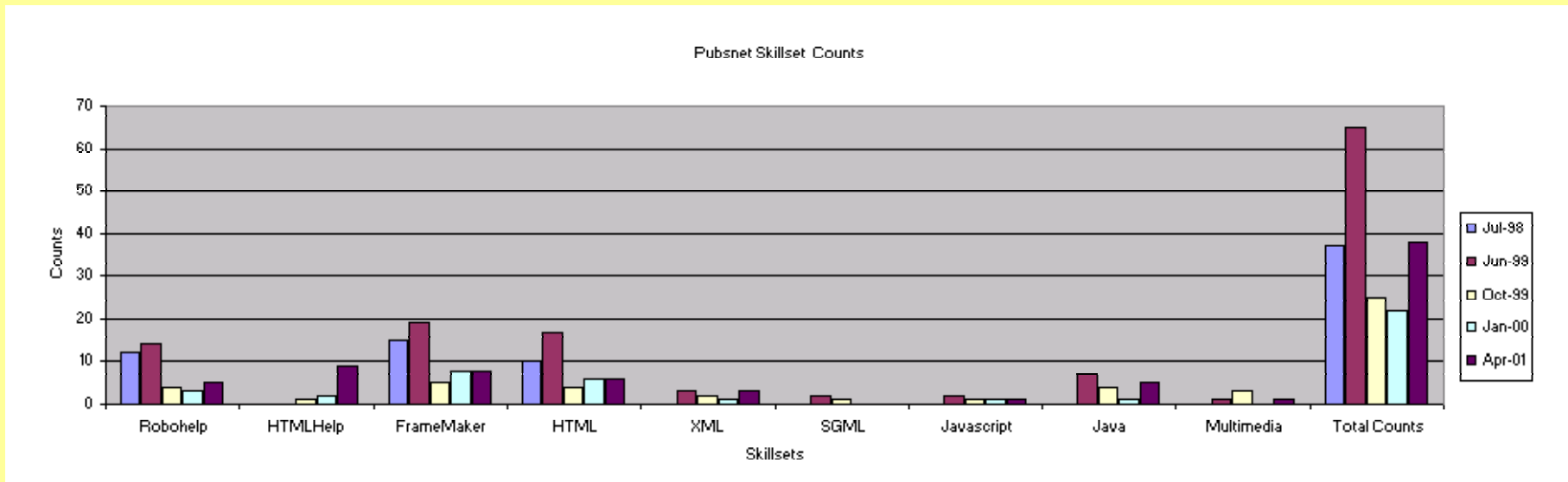
## ClearPoint Consultants Keyword Counts



**!impact:** Perl and XSLT skills beginning to be listed.

# Demand for XML Skills Grows

## Pubsnet Consultants Keyword Counts



**!impact:** Imagine (for example) applying XSL to SVG graphics for interactive job aids.

# Presentation Overview

---

- ◆ **Recap: XML, HTML, and XHTML**
- ◆ **XML Impact from a Developer's Perspective**
- ◆ **XML Impact from an Analyst Perspective**
- ◆ **XML Impact on Your Career**
- ◆ **Learn more**

# Learning More

---

- ◆ Download free software
- ◆ Read XML Books
- ◆ Download and use the XML resources handout
- ◆ Download and use the glossary handout
  
- ◆ Check out <http://world.std.com/~bboeri> for XML status updates, articles, and program handouts
- ◆ Provide us with your business card for a free CAP Ventures whitepaper on single-sourcing strategies