

Is XML in Your Future?



***A Presentation for The 16th Annual North American Serials
Interest Group Conference - May 26, 2001***

Art Rhyno, Leddy Library, University of Windsor

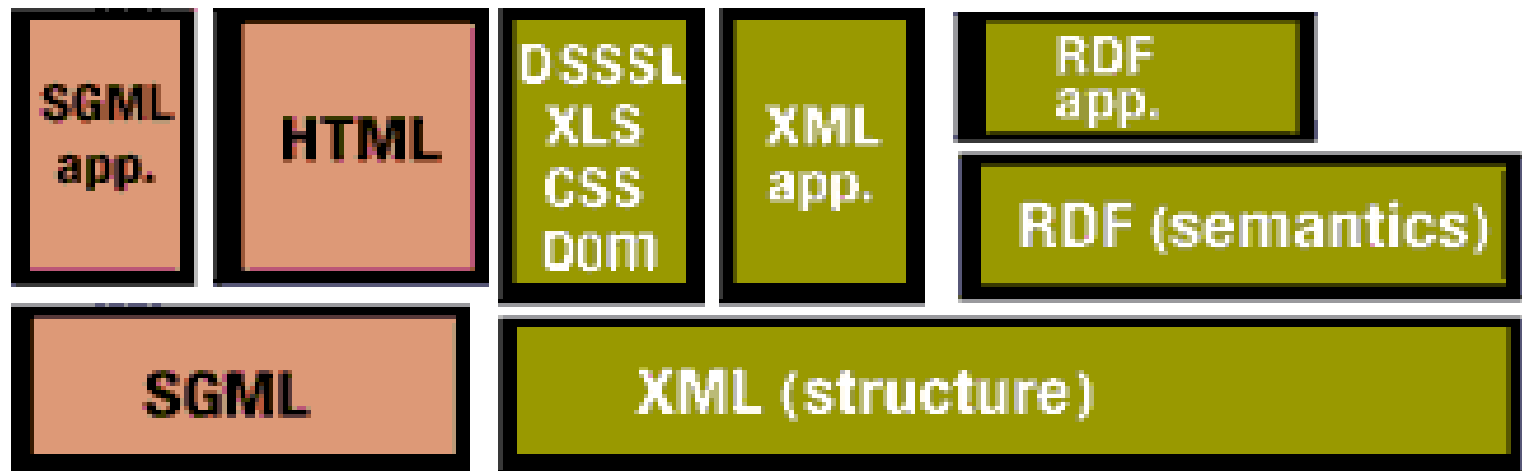
Outline



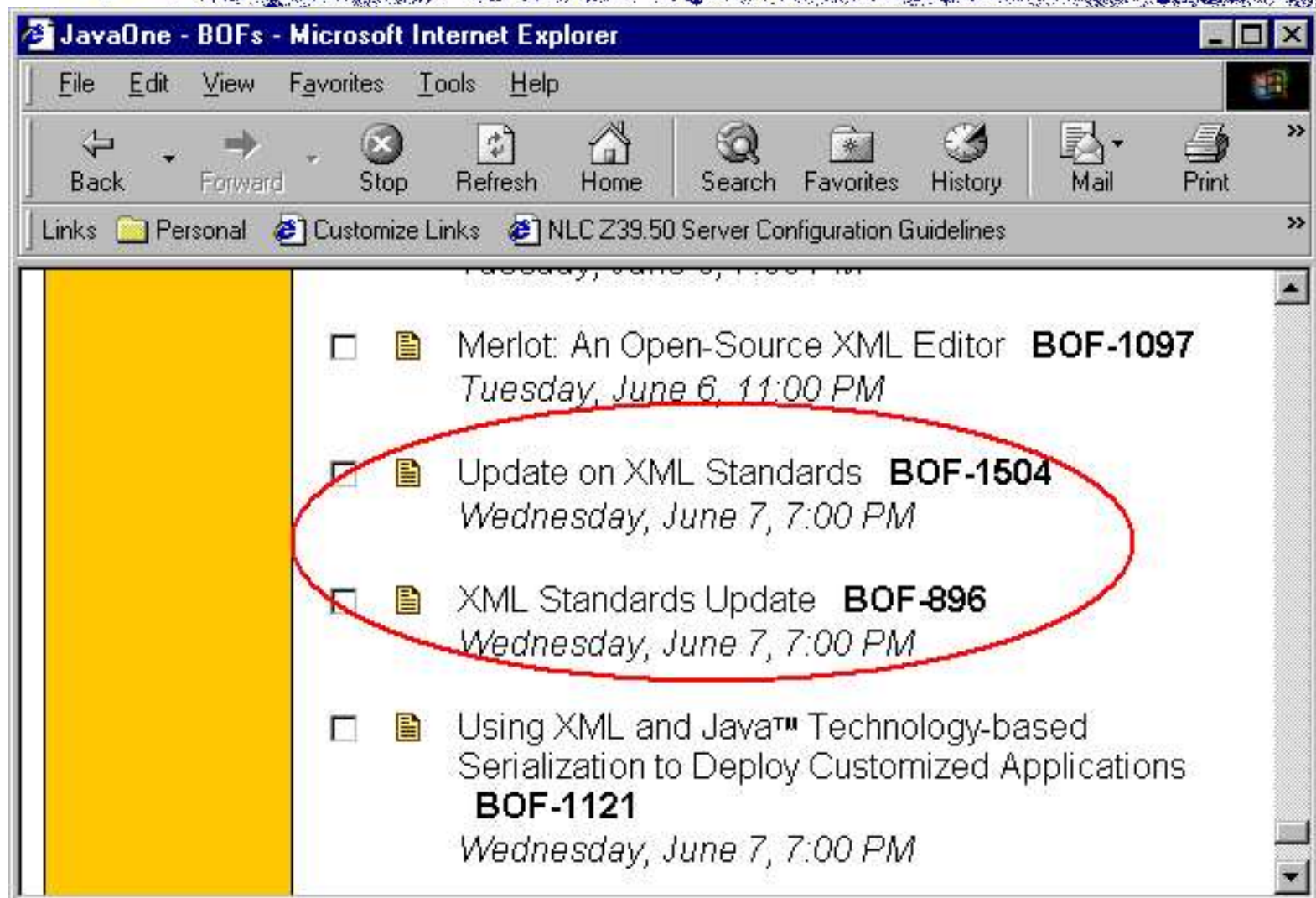
- What is XML & Why Should We Care About It Anyway?
- XML for Content Publication & Management
- XML for Integration of Systems
- Metadata & the Semantic Web
- “XML-enabling” MARC
- The Web Browser as a Global Desktop

What is XML & Why Should We Care About It Anyway – Where Does XML Come From?

- **SGML** (Standard Generalized Markup Language) - an open, non-proprietary language for describing information. SGML provides a language that can be used to define codes suitable for describing a class of documents.
- **XML** (eXtensible Markup Language) is a **subset** of SGML, as compared to HTML, which is an **application** of HTML.



What is XML & Why Should We Care About It Anyway – You know you're popular when...



What is XML – Markup Concepts

- Tags:

<titleproper>Bruce J. S. Macdonald Papers,
1896-1986**</titleproper>**

- Tags can have attributes:

<unittitle label="Title:"
encodinganalog="245\$a">

- Markup should be descriptive and **NOT** concerned with presentation:

<profiledesc> instead of **<i>**, ****, ****

- Document should be well-formed

What is XML & Why Should We Care About It Anyway – A real example



CHAPTER 38

'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also has rest from the polishing process; but Mary, bending again over the roast, said only --

'Have you, miss? Well, for sure!'

A short time after she pursued, 'I seed you go out with the master, but I didn't know you were gone to church to be wed'; and she basted away. John, when I turned to him, was grinning from ear to ear.

'I telled Mary how it would be,' he said: 'I knew what Mr Edward' (John was an old servant, and had know his master when he was the cadet of the house, therefore he often gave him his Christian name) -- 'I knew what Mr Edward would do; and I was certain he would not wait long either: and he's done right,

What is XML & Why Should We Care About It Anyway – A real example



```
<pb n='474'>
```

```
<div1 type=chapter n='38'>
```

```
<p><q>Mary, I have been married to Mr Rochester this morning.</q>
```

The housekeeper and her husband were of that decent decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also has rest from the polishing process; but Mary, bending again over the roast, said only ‐

```
<p><q>Have you, miss? Well, for sure</q>
```

```
<p>A short time after she pursued, <q>I seed you go out with the master, but I didn't know you were gone to church to be wed</q>; and she basted away. John, when I turned to him, was grinning from ear to ear. <q>I telled Mary how it would be,</q> he said: <q>I knew what Mr Edward</q> (John was an old servant, and had know his
```

What is XML & Why Should We Care About It Anyway – Another example



```
20000305:102'DTM+158:20000305:102'DTM+159:20000722:102'NAD+SU+9876543
NY'NAD+MI+88835::92'GIS+37'NAD+ST+72681::92'LIN+++93235494:IN'PIA+1+0
04'RFF+ON:XXX00004'QTY+79:6660:EA'DTM+51:19991225:102'DTM+52:20000304
91225:102'DTM+11:20000302:102'SCC+1++W:16'QTY+1:960:EA'DTM+158:200003
20000313:102'SCC+4++W:16'QTY+1:900:EA'DTM+158:20000320:102'QTY+1:900:
:1080:EA'DTM+158:20000403:102'QTY+1:1080:EA'DTM+158:20000410:102'QTY+
'QTY+1:630:EA'DTM+158:20000424:102'QTY+1:990:EA'DTM+158:20000501:102'
:102'QTY+1:810:EA'DTM+158:20000515:102'QTY+1:810:EA'DTM+158:20000522:
0529:102'QTY+1:810:EA'DTM+158:20000605:102'QTY+1:630:EA'DTM+158:20000
20000619:102'QTY+1:810:EA'DTM+158:20000626:102'QTY+1:810:EA'DTM+158:2
158:20000710:102'QTY+1:766:EA'DTM+158:20000717:102'SCC+2'QTY+3:12610:
:20000416:102'SCC+3'QTY+3:17485:EA'DTM+51:19991225:102'DTM+52:2000052
0:EA'UNT+73+770001'UNZ+1+77'UNB+UNOA:2+BFT:ZZ+CAI:ZZ+000305:2338+78++
```



What is XML & Why Should We Care About It Anyway – Another example

```
<!-- InvoiceParties contains names and address of parties and -->
<!-- their functions -->
<InvoiceParties>
  <Buyer>
    <NameAddress>
      <Name1>Ralph's Automotive Parts</Name1>
      <Address1>10 Main St.</Address1>
      <City>Boulder Creek</City>
      <StateOrProvince>California</StateOrProvince>
      <PostalCode>96005</PostalCode>
      <Country>US</Country>
    </NameAddress>
  </Buyer>
  <Supplier>
    <NameAddress>
      <Name1>ABC Wholesale</Name1>
      <Address1>1222 Industrial Park way</Address1>
      <City>South San Francisco</City>
      <StateOrProvince>California</StateOrProvince>
      <PostalCode>96045</PostalCode>
      <Country>US</Country>
    </NameAddress>
  </Supplier>
</InvoiceParties>
```



What is XML & Why Should We Care About It Anyway – It Likes to *Behave* (with apologies to Austin Powers)



- With XML, you can define a list of all tags that can be used as well as the rules that describe how they can be used.
- The definitions are stored in a DTD (Document Type Definition) or XML *Schema* and utilize a specific vocabulary.
- DTDs and Schemas are key to the effective processing of XML. They allow a document's structure to be verified.

Example – Seinfeld Dialogue



<!xml version="1.0"?>

<dialogue>

<jerry>What, you rented <quote>Home Alone</quote>?</jerry>

<george>Yeah.</george>

<jerry>I thought you saw that already...</jerry>

<george>No, I saw <quote>Home Alone II</quote>.</george>

<jerry>Oh, right... But you hated it!</jerry>

<george>Well I was lost, I never saw the first one. By the way, do you mind if I watch it here?</george>


<jerry>What for?</jerry>

<george>Because if I watch it at my apartment I feel like I'm not doing anything. If I watch it here, I'm out of the house; I'm doing something.</george>

<laughter/>

</dialogue>

DTD – Seinfeld Dialogue



```
<!DOCTYPE dialogue [  
<!ELEMENT dialogue  
(jerry+,george,elaine,kramer,laughter?)>  
<!ELEMENT jerry (#PCDATA | quote | em)*>  
<!ELEMENT george (#PCDATA | quote | em)*>  
<!ELEMENT elaine (#PCDATA | quote | em)*>  
<!ELEMENT kramer (#PCDATA | quote | em)*>  
<!ELEMENT quote (#PCDATA)*>  
<!ELEMENT em (#PCDATA)*>  
<!ELEMENT laughter EMPTY]>
```

XML Schema – Seinfeld Dialogue



```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:complexType name="JerryDialogueType">
    <xsd:element name="jerry" minOccurs="1"
maxOccurs="unbounded" type="SitcomCharacter">
    <xsd:element name="george" minOccurs="0"
maxOccurs="unbounded" type="SitcomCharacter">
    <xsd:element name="elaine" minOccurs="1"
maxOccurs="unbounded" type="SitcomCharacter">
    <xsd:element name="kramer" minOccurs="1"
maxOccurs="unbounded" type="SitcomCharacter">
    <xsd:element ref="quote" minOccurs="0"/>
    <xsd:element ref="em" minOccurs="0"/> <snip>
```

What is XML & Why Should We Care About It Anyway – It has *Style*

- XML gives a lot more control back to the user for displaying elements via ***Stylesheets***.
- Stylesheets describe how tags should be displayed (font,size, colour, etc.) XML documents can use Cascading Stylesheets (CSS) or Extensible Stylesheet Language (XSL).
- The same stylesheet can be shared by different documents. The same document can be viewed with different stylesheets. One can also be **transformed** into another using XSLT.


```
kramer { color: yellow;  
        font-size: 20 pt}
```

Or, in other words, the reasons to care about XML are...



- With XML you can capture and publish information about your data and how it is structured.
- You can exchange data with others based on an agreed upon definition.
- There are a ton of technologies that use it, whether you like it or not...

Just to name a few...



XUL	RDF	CDF	CML
MCF	MML	OSD	ONIX
OpenTag	WIDL	XML-Data	XSL
XSLT	XSQL	XQL	EXTRA
BizTalk	RSS	Xlink	Xpointer
PSML	SML	TeXML	MRML
Xforms	XHTML	Xinclude	Xpath
SVG	ebXML	SMIL	XML-RPC
SOAP	XMI	DESSERT	XBRL
XIOP	XAP	ParlML	VoiceXML
RAX	MASP	SDLIP	XMLMARC
XML-MP	NewsML	adXML	DSD...

XML for Content Publishing & Management



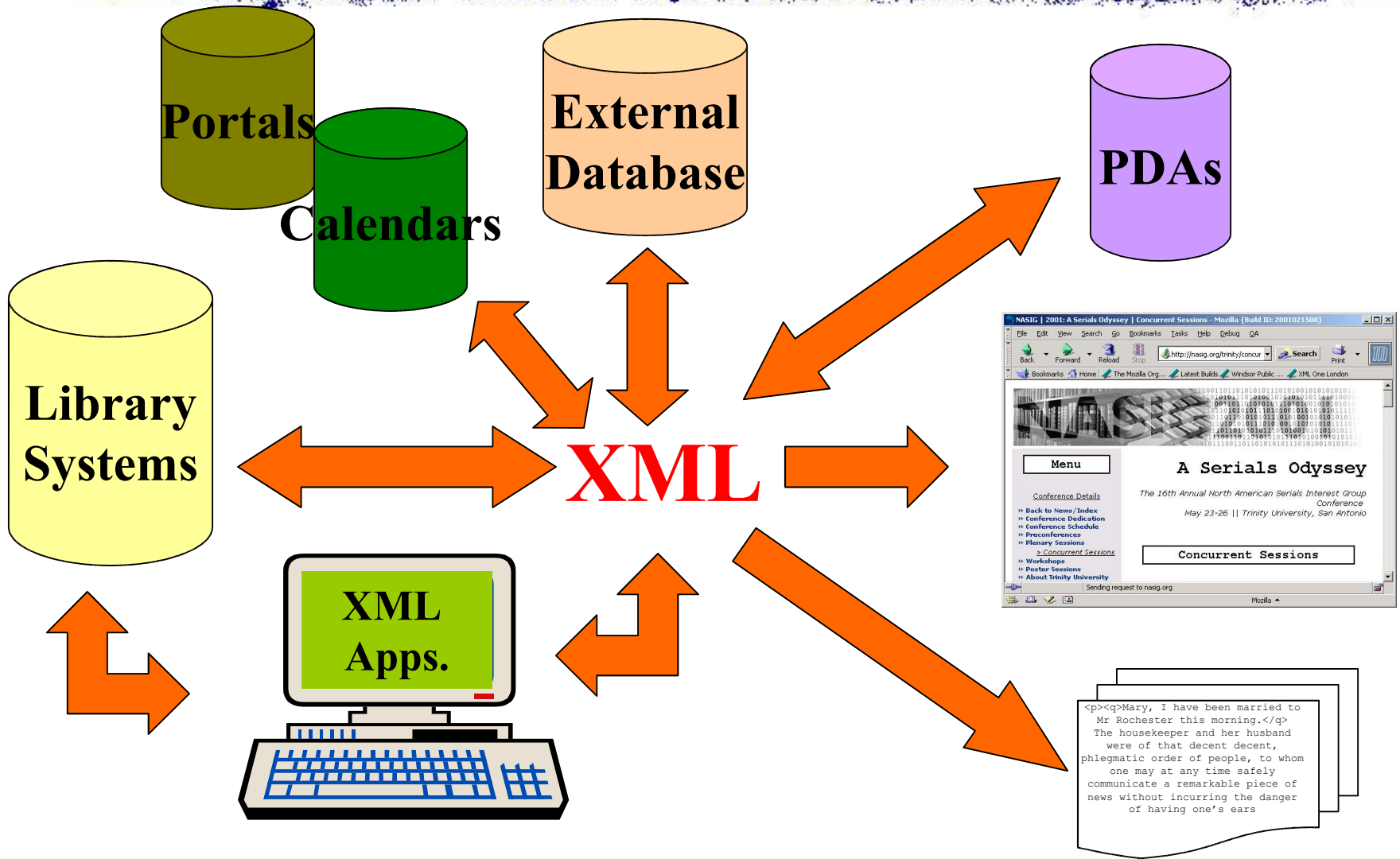
- XML “web-empowers” existing SGML standards.
- Examples include the **Text Encoding Initiative** (TEI) and the **Encoded Archival Description** (EAD).
- HTML itself can be marked up with XML using **XHTML**

XML for Content Publishing & Management



- **Cocoon** publishing framework from **Apache** can serve XML documents to non-XML sources.
- **Zope** is a credible XML repository. Includes a powerful search engine.
- ILS vendors jumping in – Endeavor's **Encompass**.
- Lots of **XML Editors**, support in **Word**, **Wordperfect**. **Oracle**, **SQL server** and lots of others

XML for Integration of Systems



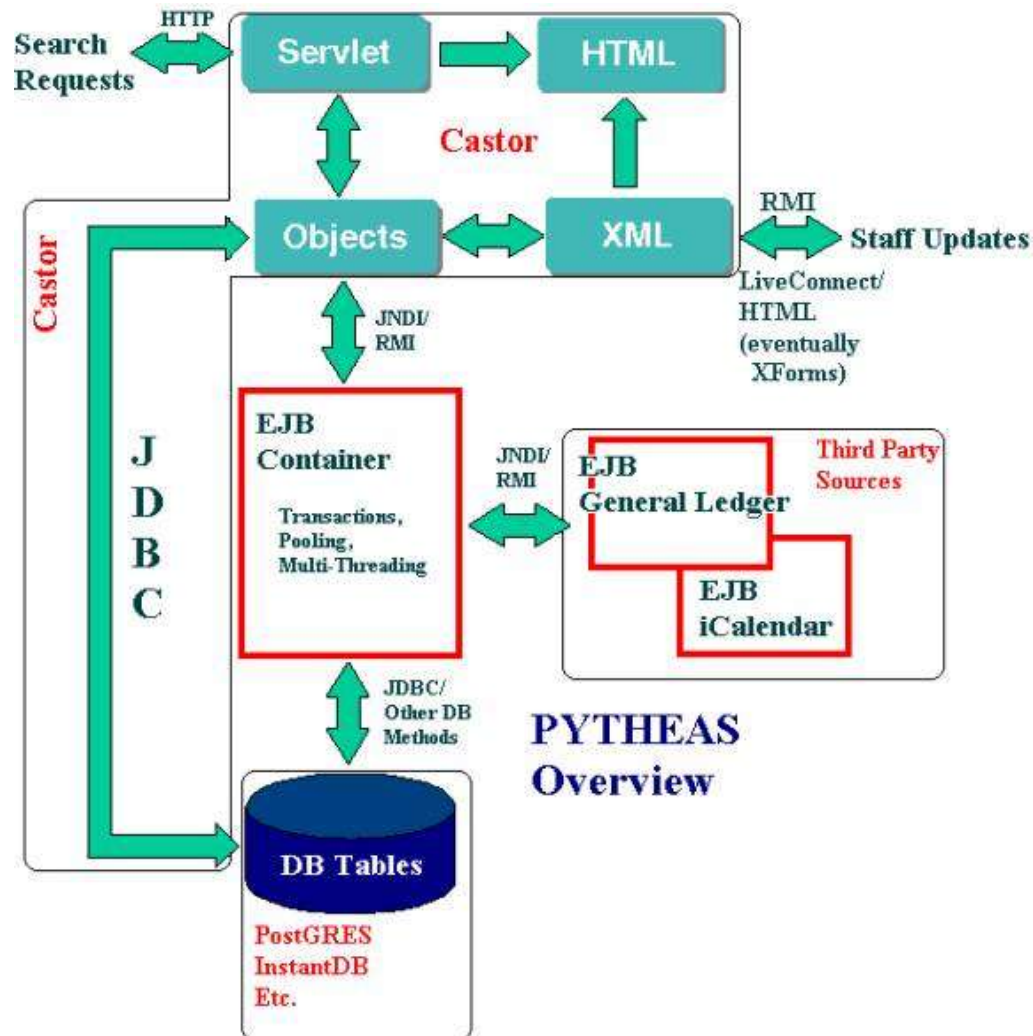
XML for Integration of Systems

– a few to watch out for



- **Rich Site Summary (RSS)** – A lightweight vocabulary to provide a “what’s new” category.
- **Open Archives Initiative (OAI)** – uses XML for carrying information about e-prints.
- **Portal Markup Language (PML)** – allows portals to share information
- **Open eBook (OEB) Specification** – XML based syntax for defining e-book file format and structure

XML for Integration - PYTHEAS



Metadata & the Semantic Web



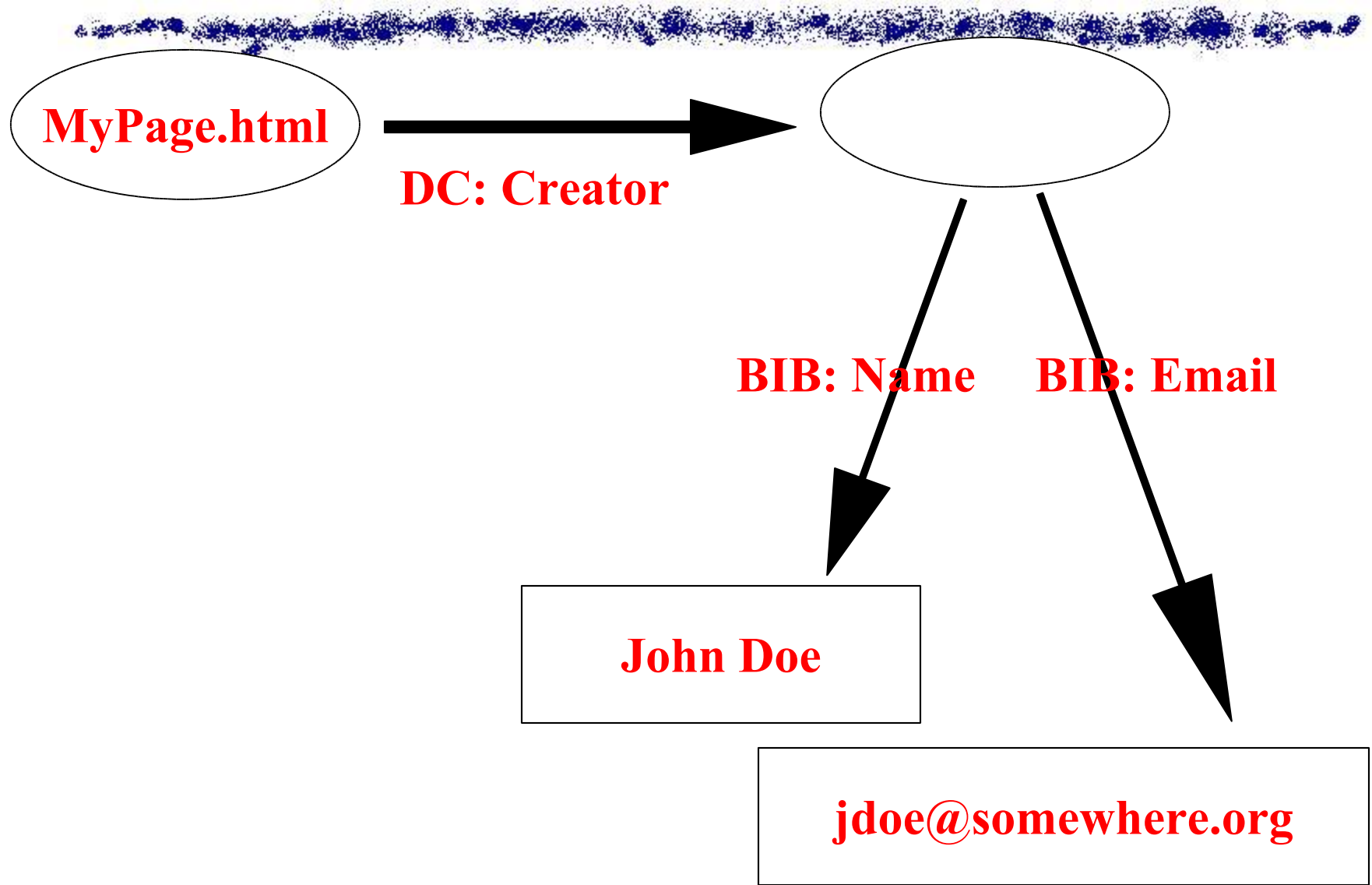
- Metadata is typically defined as “data about data”.
- Useful to remember why it is a big deal on the web
 - Saves bandwidth
 - Allows more sophisticated searching
 - Can define access restrictions
 - Integrates disparate resources

Metadata & the Semantic Web



- The W3C Metadata activities seek to provide “a common framework to express assertions about information on the Web”.
- This work includes **PICS** (Platform for Internet Content Selection), **DSig** (Digital Signatures), **P3P** (Platform for Privacy Preferences), and **CC/PP** (Composite Capabilities/Preferences Profiles).
- But the primary result of this initiative is the **Resource Description Framework (RDF)**.

RDF – Up Close with Labeled Graphs



RDF in XML



<RDF:RDF>

<RDF:Description RDF:Href= "MyPage.html">

<DC:Creator>

<RDF:Description>

<BIB:Name>John Doe</BIB:Name>

<BIB:Email>

jdoe@somewhere.org

</BIB:Email>


</RDF:Description>

</DC:Creator>

</RDF:Description>

</RDF:RDF>

RDF – a tough sell?



“...the RDF spec is particularly obtuse, and every time I have to write something on RDF my heart sinks, because I know it will take me a good 1-3 days research before I am sure I have got it right!” - Frank Boumphrey, author of *XML Applications* and *Beginning XHTML*

“...the XML syntax for RDF has too many annoying variations, granted, but the main problem is that the underlying RDF data model is much, much more complicated than the spec suggests. “ - David Megginson, main creator of SAX

“If Patton were alive, he would slap it!” - comment from one of the RDF lists

A Word About Namespaces



- Namespaces are a simple way to distinguish names used in XML documents. They give programmers a helping hand, enabling them to process the tags and attributes they care about and ignore those that don't matter to them.
- Namespaces are what allows a resource to be built or described using more than one vocabulary

A Namespace Example (courtesy of Tim Bray)



```
<h:html xmlns:xdc="http://www.xml.com/books"
        xmlns:h="http://www.w3.org/HTML/1998/html4">
  <h:head><h:title>Book Review</h:title></h:head>
  <h:body>
    <xdc:bookreview>
      <xdc:title h:style="font-family: sans-serif;">
        XML: A Primer</xdc:title>
      <h:table>
        <h:tr align="center">
          <h:td>Author</h:td><h:td>Price</h:td>
          <h:td>Pages</h:td><h:td>Date</h:td></h:tr>
        <h:tr align="left">
          <h:td><xdc:author>Simon St. Laurent</xdc:author></h:td>
          <h:td><xdc:price>31.98</xdc:price></h:td>
          <h:td><xdc:pages>352</xdc:pages></h:td>
          <h:td><xdc:date>1998/01</xdc:date></h:td>
        </h:tr>
      </h:table>
    </xdc:bookreview>
  </h:body>
</h:html>
```

Ontologies & Knowledge Sharing



- An ontology is a specification of a conceptualization
- For knowledge sharing, an ontology is a specification used for making ontological commitments
- An ontological commitment is an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent with respect to the theory specified by an ontology.

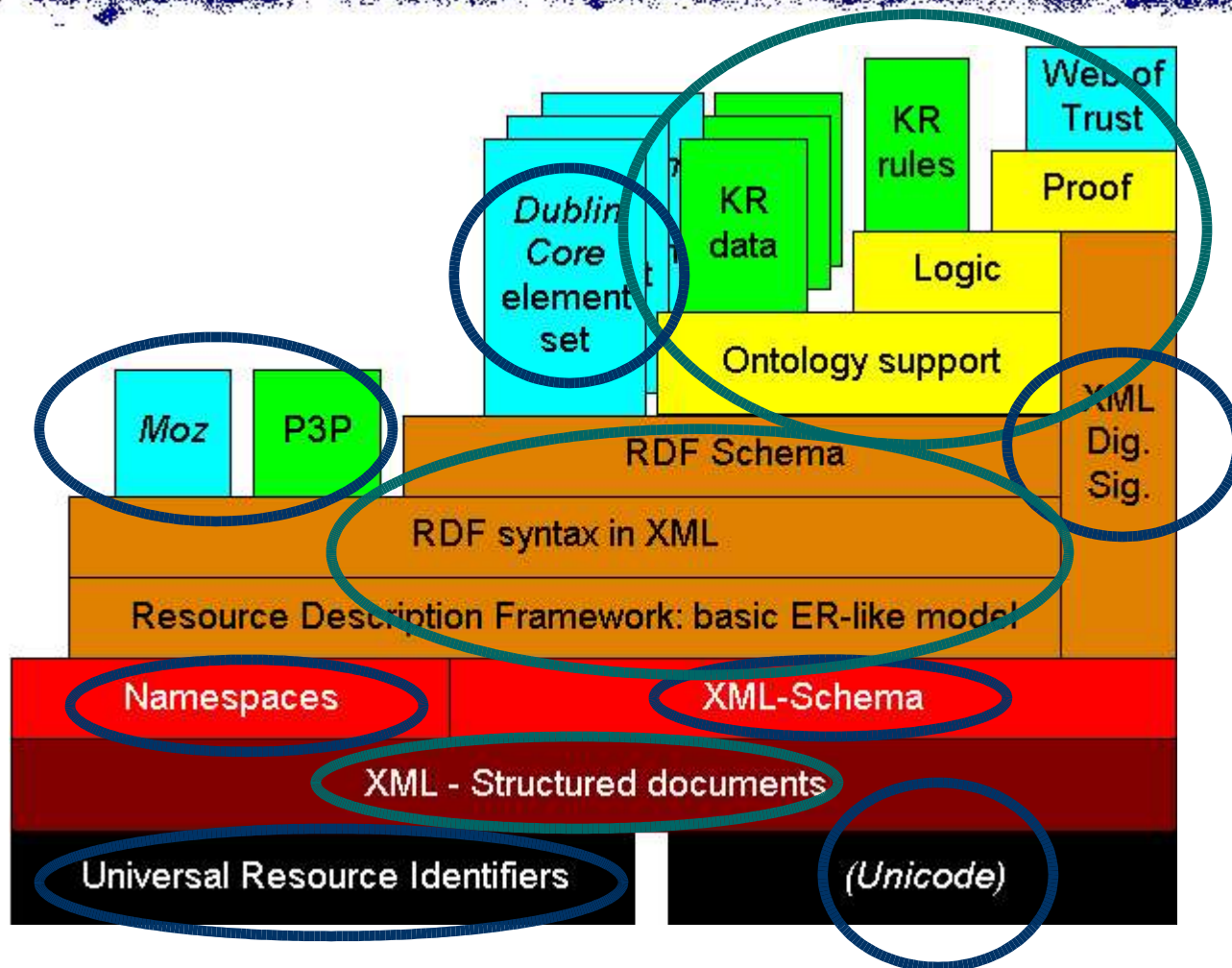
The Semantic Web



“The Semantic Web is a Web that includes documents or portions of documents, describing explicit relationships between things and containing semantic information intended for automated processing by machines.” – W3C Semantic Web Agreement Group

“If anyone thinks that typing “Aunt Hilda” into a Semantic Web search engine will produce an encyclopedias worth of information about their Aunt Hilda, then they are very much mistaken”

The Semantic Web – a view from the Web's creator



The Semantic Web – a few thoughts



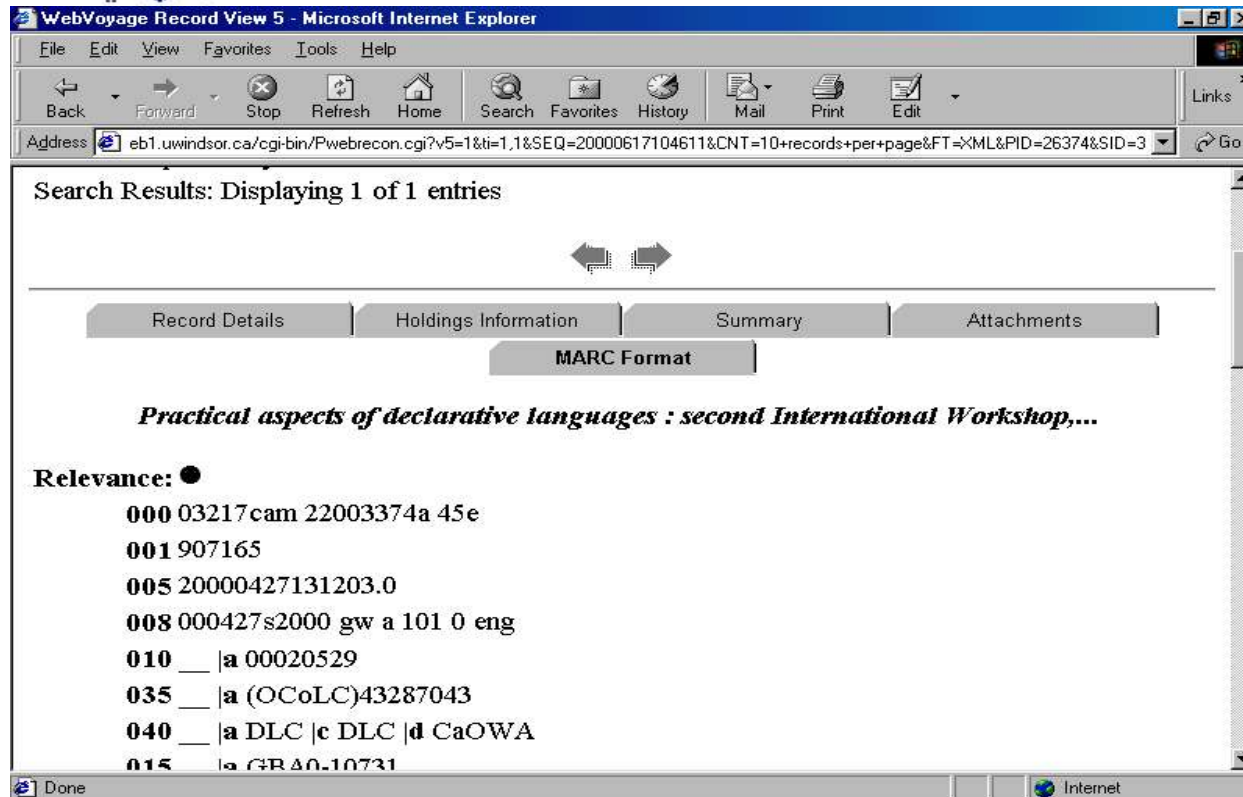
- Information Retrieval is always harder than anyone expects.
- Author-assigned metadata has a dubious track record.
- It is always better to dumb down data on the way out rather than on the way in.

“XML-enabling” MARC



- XMLMARC is an experimental effort from the Lane Medical Library at Stanford to create a flexible retrieval and display mechanism for bibliographic, authority, and other ‘library’ information using XML.
- LC’s SGML for MARC - 500 p.
- Lane’s XML for 80% of MARC (“Reduced complexity--without content loss”)
 - Works - 7 p.
 - Authorities - 4 p.

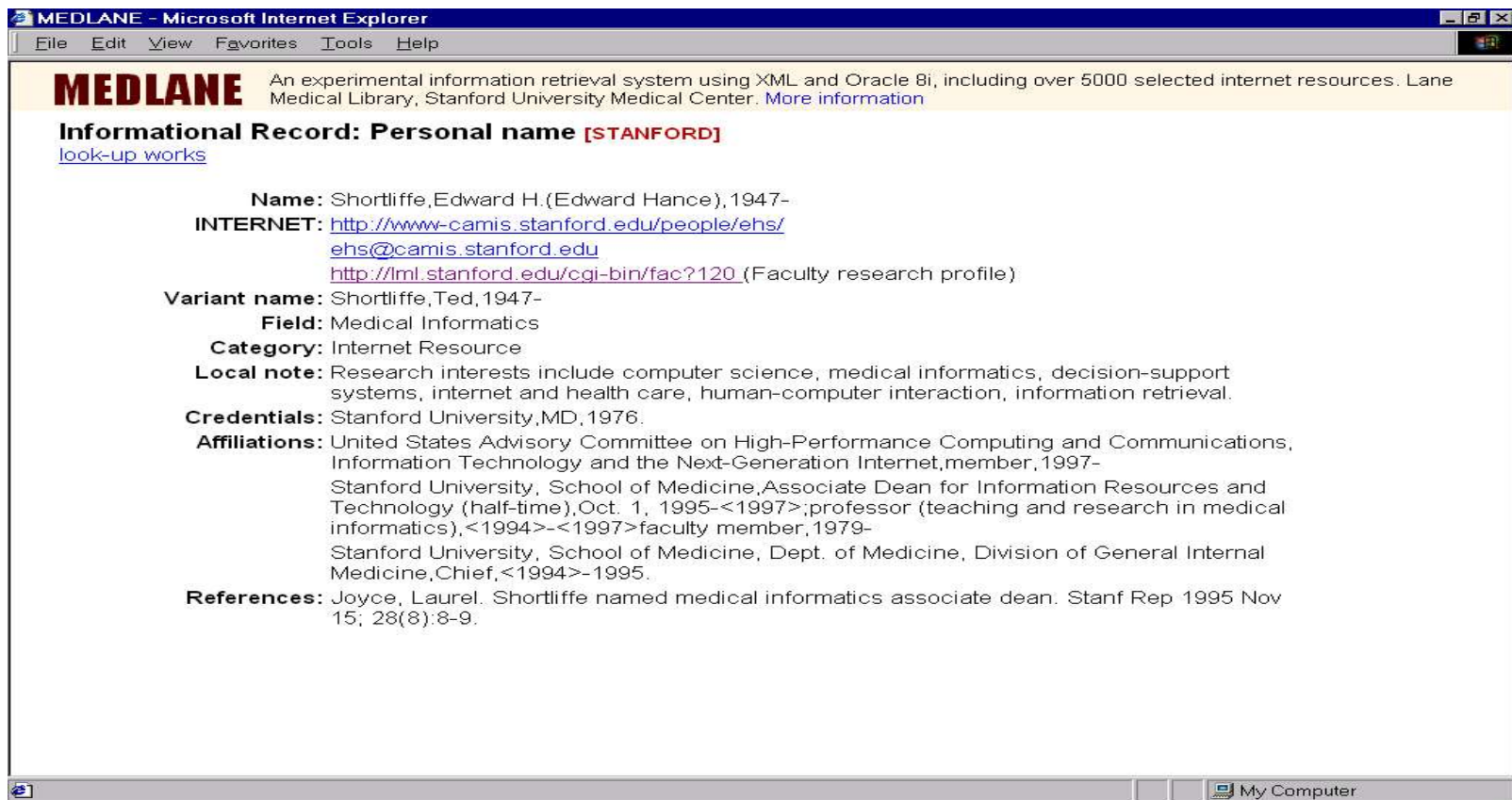
“XML-enabling” MARC



“Who is this MARC guy and why does he get his own format?” - comments sent from Web site

“XML-enabling” MARC

XML + XSL = Flexibility



MEDLANE An experimental information retrieval system using XML and Oracle 8i, including over 5000 selected internet resources. Lane Medical Library, Stanford University Medical Center. [More information](#)

Informational Record: Personal name [STANFORD]
[look-up works](#)

Name: Shortliffe, Edward H. (Edward Hance), 1947-

INTERNET: <http://www-camis.stanford.edu/people/ehs/>
ehs@camis.stanford.edu
<http://lml.stanford.edu/cgi-bin/fac?120> (Faculty research profile)

Variant name: Shortliffe, Ted, 1947-

Field: Medical Informatics

Category: Internet Resource

Local note: Research interests include computer science, medical informatics, decision-support systems, internet and health care, human-computer interaction, information retrieval.

Credentials: Stanford University, MD, 1976.

Affiliations: United States Advisory Committee on High-Performance Computing and Communications, Information Technology and the Next-Generation Internet, member, 1997-
Stanford University, School of Medicine, Associate Dean for Information Resources and Technology (half-time), Oct. 1, 1995-<1997>; professor (teaching and research in medical informatics), <1994>-<1997> faculty member, 1979-
Stanford University, School of Medicine, Dept. of Medicine, Division of General Internal Medicine, Chief, <1994>-1995.



References: Joyce, Laurel. Shortliffe named medical informatics associate dean. Stanf Rep 1995 Nov 15; 28(8):8-9.

“XML-enabling” MARC – Next Steps?



- Lessons to be learned from EDI – SIMPL EDI.
- XML Schema/RDF Schema combination holds great potential for bringing MARC to XML for “front-end” and “back-end” functions.
- Metadata Term Thesauri holds possibilities for Authority Control – MetaNet and Harmony Project.

“XML-enabling” MARC – Next Steps?

Results of Search for metadata term: *event*

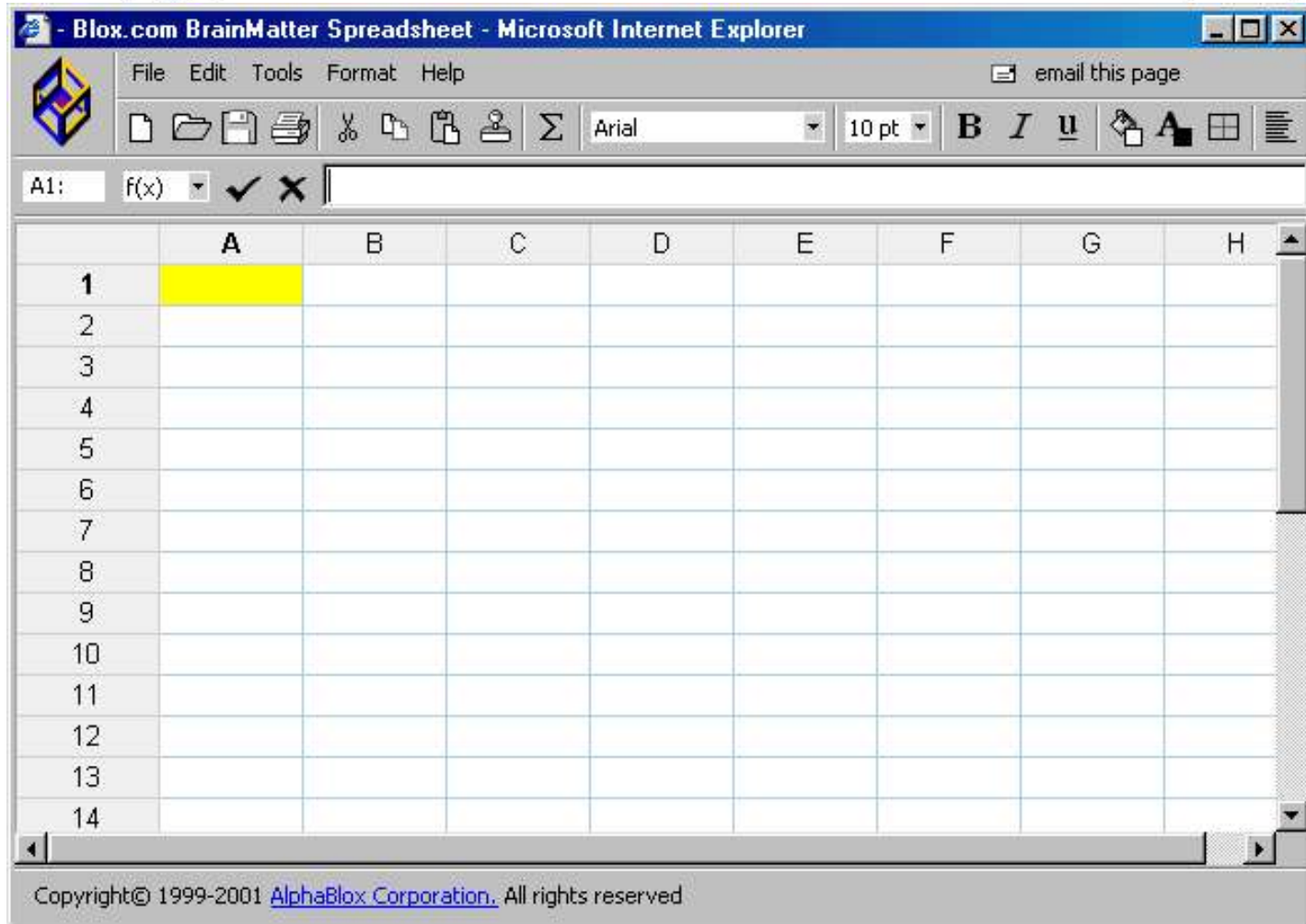
Core Term	Synonyms/Equivalent Terms	Hyponyms/Narrower Terms	Hypo-hyponyms/Narrowest Terms:
<i>event</i>	happening, occurrence, activity, act, action, phenomenon, occasion, episode, development, incident, circumstance, situation, proceeding, eventuality,	origin, creation, conception, birth, composition	-
		performance, expression	-
		modification, translation, edition, conversion, interpretation, adaptation, transformation, transmutation, metamorphosis, alteration, extraction, compilation, arrangement, remixing, copying, replication, printing, recording	-
		publication, publish, broadcast, transmission, airing, dissemination, disseminate, distribution, issued,	-

The Web as a Global Desktop

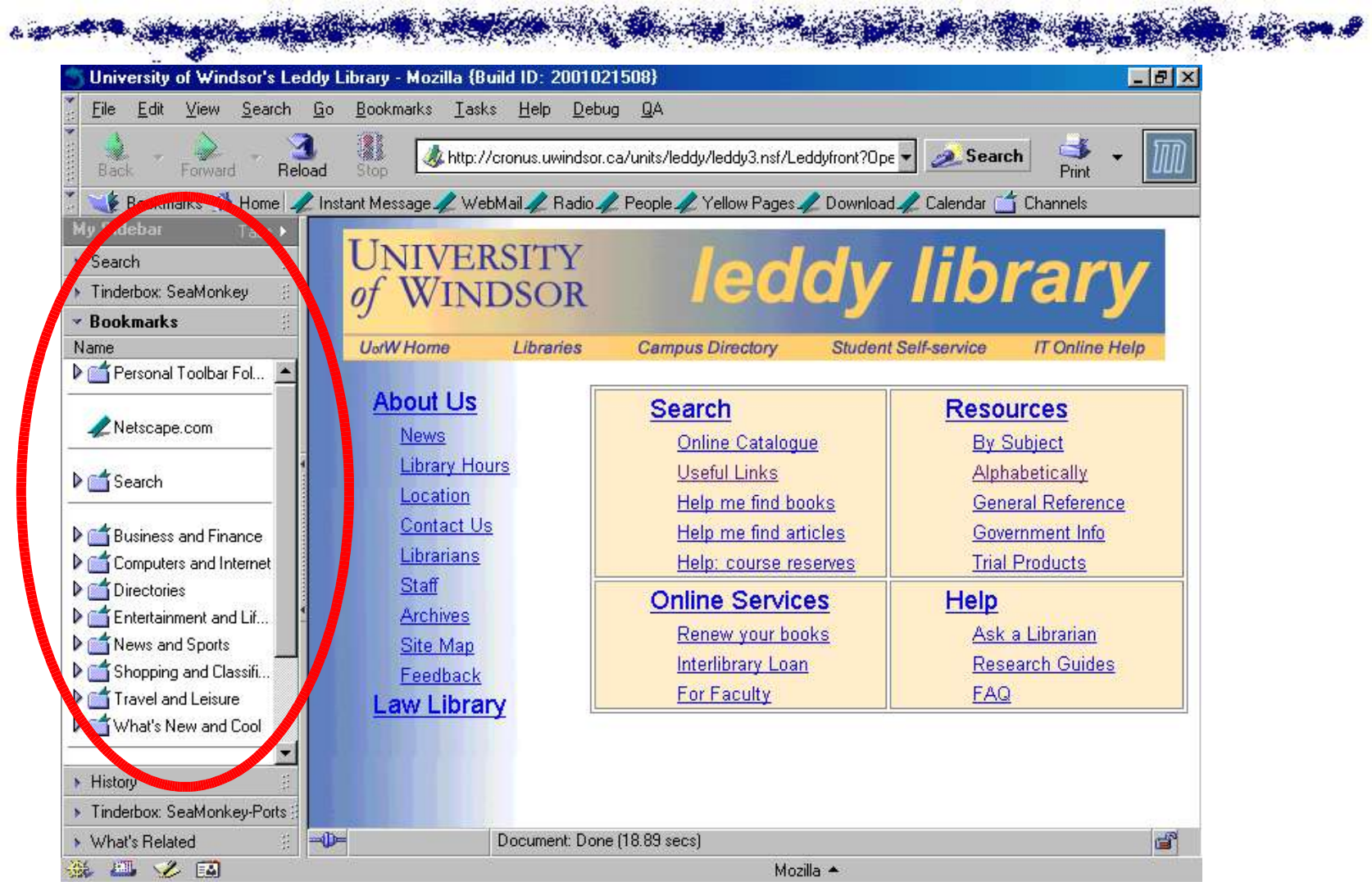


- Distinctions between web browser and desktop starting to blur.
- Mozilla has strong toolkit for richer browser applications. Microsoft's .Net creates mobile desktop.
- W3C is working to enhance the capabilities of web forms and other web interactions.
- Can see a glimpse of this future at Blox.com

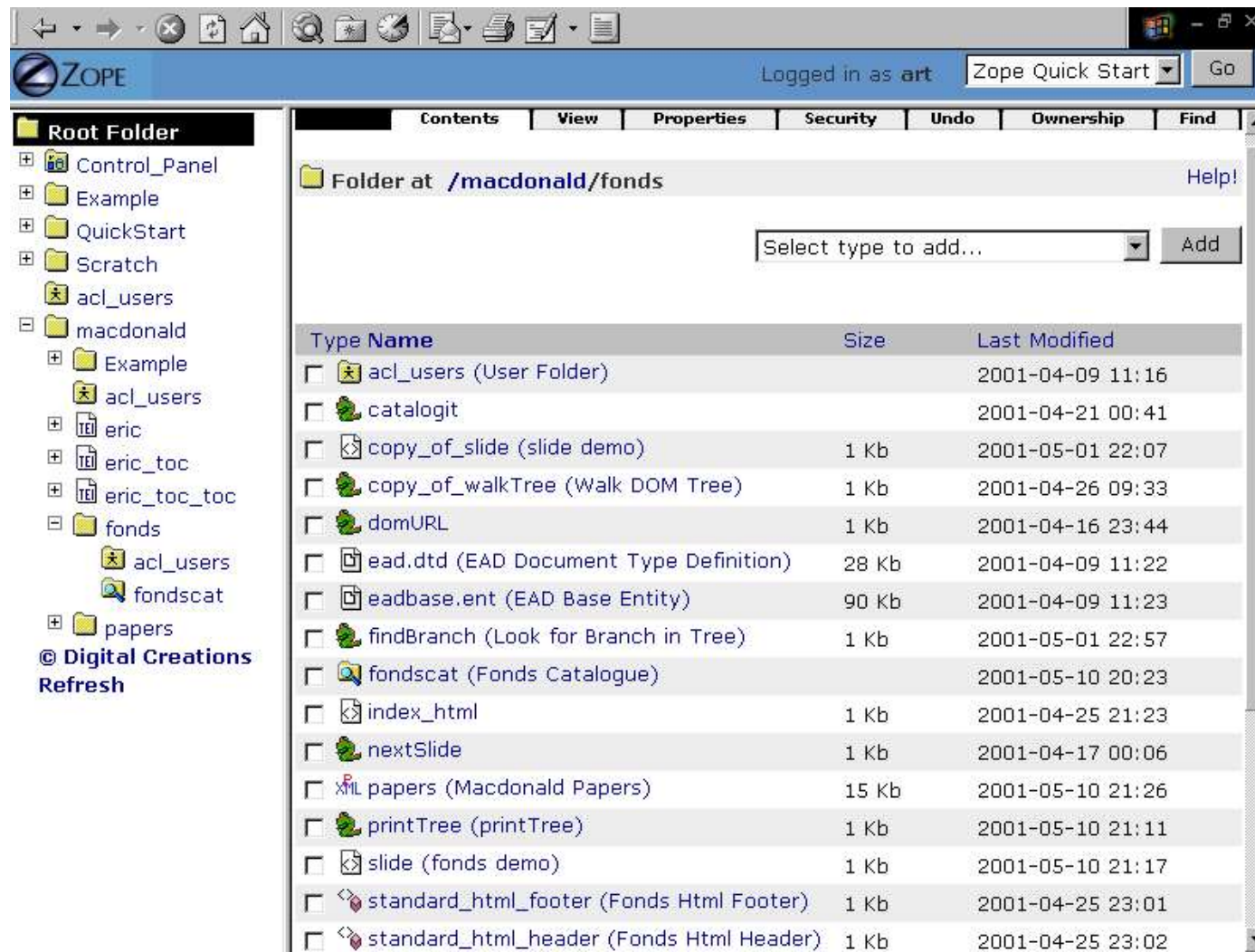
The Web as a Global Desktop



The Web as a Global Desktop



The Web as a Global Desktop



The screenshot displays the Zope web interface, a web-based file management system. The interface includes a navigation pane on the left, a main content area, and a top toolbar. The navigation pane shows a tree structure starting from a 'Root Folder', with the path '/macdonald/fonds' selected. The main content area shows the contents of this folder, including a list of files and folders with their sizes and last modified dates. A 'Help!' link is visible in the top right of the main area. The top toolbar contains various icons for navigation and file management, along with a login status 'Logged in as art' and a 'Zope Quick Start' dropdown menu.

Root Folder

- Control_Panel
- Example
- QuickStart
- Scratch
- acl_users
- macdonald
 - Example
 - acl_users
 - eric
 - eric_toc
 - eric_toc_toc
 - fonds
 - acl_users
 - fondscat
 - papers

© Digital Creations Refresh

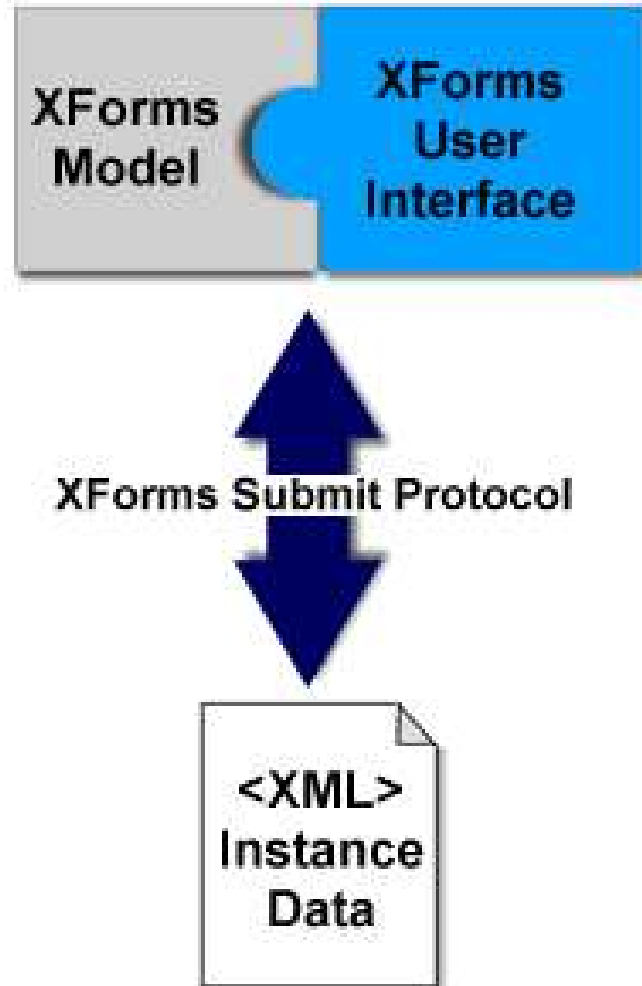
Contents View Properties Security Undo Ownership Find

Folder at /macdonald/fonds Help!

Select type to add... Add

Type	Name	Size	Last Modified
<input type="checkbox"/>	acl_users (User Folder)		2001-04-09 11:16
<input type="checkbox"/>	catalogit		2001-04-21 00:41
<input type="checkbox"/>	copy_of_slide (slide demo)	1 Kb	2001-05-01 22:07
<input type="checkbox"/>	copy_of_walkTree (Walk DOM Tree)	1 Kb	2001-04-26 09:33
<input type="checkbox"/>	domURL	1 Kb	2001-04-16 23:44
<input type="checkbox"/>	ead.dtd (EAD Document Type Definition)	28 Kb	2001-04-09 11:22
<input type="checkbox"/>	eadbase.ent (EAD Base Entity)	90 Kb	2001-04-09 11:23
<input type="checkbox"/>	findBranch (Look for Branch in Tree)	1 Kb	2001-05-01 22:57
<input type="checkbox"/>	fondscat (Fonds Catalogue)		2001-05-10 20:23
<input type="checkbox"/>	index_html	1 Kb	2001-04-25 21:23
<input type="checkbox"/>	nextSlide	1 Kb	2001-04-17 00:06
<input type="checkbox"/>	papers (Macdonald Papers)	15 Kb	2001-05-10 21:26
<input type="checkbox"/>	printTree (printTree)	1 Kb	2001-05-10 21:11
<input type="checkbox"/>	slide (fonds demo)	1 Kb	2001-05-10 21:17
<input type="checkbox"/>	standard_html_footer (Fonds Html Footer)	1 Kb	2001-04-25 23:01
<input type="checkbox"/>	standard_html_header (Fonds Html Header)	1 Kb	2001-04-25 23:02

The Web as a Global Desktop



The Web as a Global Desktop



■ A place where

- Library information can be passed seamlessly into many other types of applications.
- There is complete control over web interfaces and web interactions.
- Metadata is shared with other communities.
- Library applications and tools use mainstream technologies rather than focused on a “niche” market.

Some XML Projects You Can Try Now



1. Create small XHTML, TEI or EAD documents with an XML-authoring program.
2. Convert some MARC Records to XML with XMLMARC software.
3. Use MARC/GILS/DC crosswalks to convert library records to DC and then represent in RDF.
4. Use Zope to store and index some XML documents.
5. Retrieve some RDF resources from the Open Directory Project.

Is XML in Your Future?



Some URLs



<http://www.w3.org/XML/>

<http://xml.coverpages.org/sgml-xml.html>

<http://www.xml.com>

<http://xmlmarc.stanford.edu>

<http://www.zope.org>

<http://lcweb.loc.gov/marc/dccross.html>

<http://xml.apache.org>

<http://www.ilrt.bris.ac.uk/discovery/harmony>

<http://dmoz.org>

<http://www.blox.com/>