

XML Metadata Standards and Topic Maps

Erik Wilde

16.7.2001

XML Metadata Standards and Topic Maps

1

Outline

- what is XML?
 - a syntax (not a data model!)
- what is the data model behind XML?
 - XML Information Set (basically, trees...)
- what can be described with XML?
 - describing the content syntactically (schemas)
 - describing the content abstractly (metadata)
- XML metadata is outside of XML documents
- ISO Topic Maps
 - a "schema language" for meta data

16.7.2001

XML Metadata Standards and Topic Maps

2

Extensible Markup Language

- standardized by the W3C in February 1998
- a subset (aka *profile*) of SGML (ISO 8879)
- coming from a document world
 - data are documents
- defined in syntax
 - no abstract data model
- problems in many real-world scenarios
 - how to compare XML documents
 - attribute order, white space, namespace prefixes, ...
 - how to search for data within documents
 - query languages operate on abstract data models
 - often data are not documents

Why XML at all?

- because it's simple
 - easily understandable, human-readable
- because of the available tools
 - it's easy to find (free) XML software
- because of improved interoperability
 - all others do it!
 - easy to interface with other XML applications
- because it's versatile
 - the data model behind XML is very versatile

XML Information Set

- several XML applications need a data model
 - style sheets for XML (CSS, XSL)
 - interfaces to programming languages (DOM)
 - XML transformation languages (XSLT)
 - XML fragment identifiers (XPointer)
 - XML query languages (XQuery)
- XML does not have a real data model
 - implicitly defined, but not authoritatively
- XML Information Set (XML Infoset)
 - describes a set of *information items*
 - each XML document is a set of such items

XML Infoset Essentials

- only Namespace-compliant XML allowed!
- so what's in the Infoset?
 - elements
 - attributes
 - Namespace declarations and prefixes
 - comments
 - processing instructions
- and what's not in the Infoset?
 - whitespace within element tags
 - the order of attributes within element tags
 - any information about the DTD

XML Schema Languages

- XML represents structured Information
 - XML Infoset defines the data model (trees)
 - XML 1.0 defines a character-based syntax
- XML 1.0 also defines DTDs
 - element types and their content models
 - attributes and their data types
- every XML application has to support DTDs
 - the only globally accepted schema language
 - almost 20 years old
 - many drawbacks for non-document scenarios

XML Schema

- developed because of user demand
 - B2B scenarios need better data types
 - data modeling needs better structuring
- XML Schema W3C standard since 5/2001
 - implementations available
 - rapid adoption is very likely
- Part I defines structuring mechanisms
 - element types may be derived from each other
- Part II defines a data type vocabulary
 - a set of application-oriented *simple types*

Schemas and Metadata

- XML resources may contain any type of data
 - documents (as originally intended by SGML)
 - order forms (as is common in B2B scenarios)
 - generic things such as RPC requests and responses
 - SOAP and XML RPC are two popular variants
 - or even information about other resources
- XML metadata describes data resources
 - not necessarily XML data (eg, image descriptions)
 - not necessarily attached to the resources
 - making comments on other people's resources
 - metadata is also data (ie, structured information)
 - XML metadata needs schema definitions

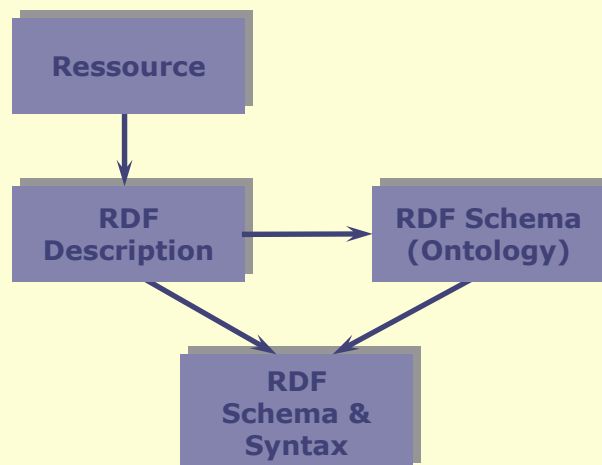
XMLizing the World...

- should everything be XML?
 - structured data would be an appropriate target
 - but what about GIF, JPEG, MPEG, ... ?
- everything should be described using XML
 - descriptions of resources are metadata
 - metadata is structured data
 - metadata should be in XML
- so there must be an XML metadata standard
 - TimBL's favorite: *Resource Description Framework*
 - coming from ISO standardization: *Topic Maps*

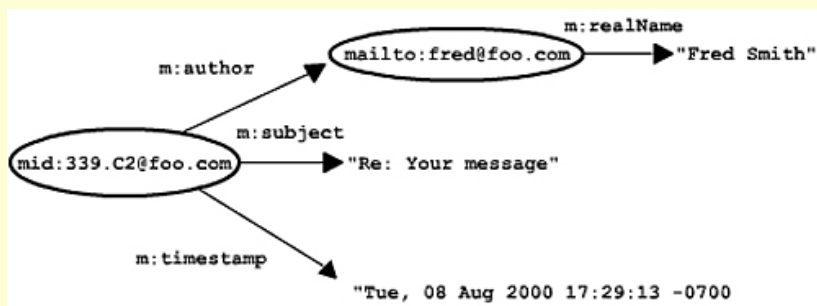
Resource Description Framework

- RDF starts with a data model
 - and defines an XML syntax for representation
- everything in RDF can be represented by a graph with nodes and arcs
 - each node is a resource
 - each arc represents a property
 - properties and resources are named with URIs
- describes the whole Web and beyond
 - anything which can be named with a URI
 - which is almost anything (phone, tv-channels, ...)
- RDF graphs describe logical assertions

RDF Metadata



RDF-based Email Description



16.7.2001

XML Metadata Standards and Topic Maps

13

But ... what is it good for?

- ask questions about the email
 - who sent me mail on a particular topic?
 - get me all the mail from Fred Smith
 - who where the people who I mailed with on Friday?
- join the email graphs with other ones
 - address books
 - home pages
 - browser history
 - organizational affiliations

16.7.2001

XML Metadata Standards and Topic Maps

14

Topic Maps

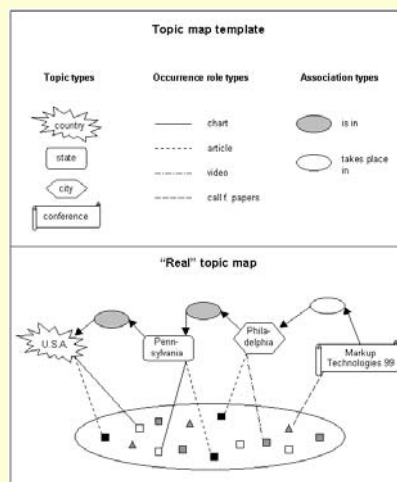
- Topics are "things of interest"
 - loosely defined, widely usable
 - each Topic has name(s) and/or occurrence(s)
 - Topics have "topic types" (which are Topics...)
- Associations are used to connect Topics
 - they have an "association type" (which is a Topic...)
 - Topics references in Associations have an "association role type" (which are Topics...)
- Topic Occurrences point to resources
 - anything addressable by a name (URI)
 - described by an "occurrence role type" (a Topic...)

16.7.2001

XML Metadata Standards and Topic Maps

15

A Simple Topic Map



16.7.2001

XML Metadata Standards and Topic Maps

16

Topic Map View (II)

ABCDEFGHIJKLMNPQRSTUVWXYZ	ABCDEFGHIJKLMNPQRSTUVWXYZ
<p>XLink (Xlink Data) XML Linking Language</p> <p>XLink defines how to insert links in XML documents. It specifies a framework making it possible for XML applications to recognize XML elements as having link semantics. In addition to the simple, two-ended, unidirectional links which are well-known from HTML, XLink allows more general links, which must not be embedded in the document, can have any number of ends, and can be multidirectional.</p> <ul style="list-style-type: none"> XLink is based on <ul style="list-style-type: none"> XML Namespaces XLink has been standardized by <ul style="list-style-type: none"> W3C XLink is authoritatively described at <ul style="list-style-type: none"> W3C XML Linking (external link) XLink is described at <ul style="list-style-type: none"> The XML Core Pages - XML Linking and Addressing Languages (XPath, XPointer, XLink) (external link) <p><small>document(s) with a view technology previously unknown</small></p>	<p>Welcome to the online version of W3C's XML glossary. The abbreviation index (shown below) provides a concise list of all abbreviations, while the individual letter indices (shown as clickable via the letter bar) list the full names as well. In addition to this online version, a PDF version is available. Please note that the PDF version contains less topics than the online version, and also less structural information. Eventually, the PDF version contains all topics which have descriptive text, and with this text in hyperlinked form, but without the structural information for those topics.</p> <p>If you have any comments, please send me an email. Thanks.</p> <p>AAC ABNF AC-3 ACAP Actix4 AENK ADSE AEN AF5 AFB ABF-C ABFH ABH ARAvita Anova ARS Apache HTTP Server Apache Software Foundation API apply ARP ARPA ARPNAT ASA ASCE ASF AOL1 ASP ATA ATAPI ATH ATN Forum AT&T AVE AVE1 EUB BSC Bascis Base32 Base64 Base64 Authentication BC BEB EUP EUP1 EUP2 EUP3 EUP4 EUP5 EUP6 EUP7 EUP8 EUP9 EUP10 EUP11 EUP12 EUP13 EUP14 EUP15 EUP16 EUP17 EUP18 EUP19 EUP20 EUP21 EUP22 EUP23 EUP24 EUP25 EUP26 EUP27 EUP28 EUP29 EUP30 EUP31 EUP32 EUP33 EUP34 EUP35 EUP36 EUP37 EUP38 EUP39 EUP40 EUP41 EUP42 EUP43 EUP44 EUP45 EUP46 EUP47 EUP48 EUP49 EUP50 EUP51 EUP52 EUP53 EUP54 EUP55 EUP56 EUP57 EUP58 EUP59 EUP60 EUP61 EUP62 EUP63 EUP64 EUP65 EUP66 EUP67 EUP68 EUP69 EUP70 EUP71 EUP72 EUP73 EUP74 EUP75 EUP76 EUP77 EUP78 EUP79 EUP80 EUP81 EUP82 EUP83 EUP84 EUP85 EUP86 EUP87 EUP88 EUP89 EUP90 EUP91 EUP92 EUP93 EUP94 EUP95 EUP96 EUP97 EUP98 EUP99 EUP100 EUP101 EUP102 EUP103 EUP104 EUP105 EUP106 EUP107 EUP108 EUP109 EUP110 EUP111 EUP112 EUP113 EUP114 EUP115 EUP116 EUP117 EUP118 EUP119 EUP120 EUP121 EUP122 EUP123 EUP124 EUP125 EUP126 EUP127 EUP128 EUP129 EUP130 EUP131 EUP132 EUP133 EUP134 EUP135 EUP136 EUP137 EUP138 EUP139 EUP140 EUP141 EUP142 EUP143 EUP144 EUP145 EUP146 EUP147 EUP148 EUP149 EUP150 EUP151 EUP152 EUP153 EUP154 EUP155 EUP156 EUP157 EUP158 EUP159 EUP160 EUP161 EUP162 EUP163 EUP164 EUP165 EUP166 EUP167 EUP168 EUP169 EUP170 EUP171 EUP172 EUP173 EUP174 EUP175 EUP176 EUP177 EUP178 EUP179 EUP180 EUP181 EUP182 EUP183 EUP184 EUP185 EUP186 EUP187 EUP188 EUP189 EUP190 EUP191 EUP192 EUP193 EUP194 EUP195 EUP196 EUP197 EUP198 EUP199 EUP200 EUP201 EUP202 EUP203 EUP204 EUP205 EUP206 EUP207 EUP208 EUP209 EUP210 EUP211 EUP212 EUP213 EUP214 EUP215 EUP216 EUP217 EUP218 EUP219 EUP220 EUP221 EUP222 EUP223 EUP224 EUP225 EUP226 EUP227 EUP228 EUP229 EUP230 EUP231 EUP232 EUP233 EUP234 EUP235 EUP236 EUP237 EUP238 EUP239 EUP240 EUP241 EUP242 EUP243 EUP244 EUP245 EUP246 EUP247 EUP248 EUP249 EUP250 EUP251 EUP252 EUP253 EUP254 EUP255 EUP256 EUP257 EUP258 EUP259 EUP260 EUP261 EUP262 EUP263 EUP264 EUP265 EUP266 EUP267 EUP268 EUP269 EUP270 EUP271 EUP272 EUP273 EUP274 EUP275 EUP276 EUP277 EUP278 EUP279 EUP280 EUP281 EUP282 EUP283 EUP284 EUP285 EUP286 EUP287 EUP288 EUP289 EUP290 EUP291 EUP292 EUP293 EUP294 EUP295 EUP296 EUP297 EUP298 EUP299 EUP300 EUP301 EUP302 EUP303 EUP304 EUP305 EUP306 EUP307 EUP308 EUP309 EUP310 EUP311 EUP312 EUP313 EUP314 EUP315 EUP316 EUP317 EUP318 EUP319 EUP320 EUP321 EUP322 EUP323 EUP324 EUP325 EUP326 EUP327 EUP328 EUP329 EUP330 EUP331 EUP332 EUP333 EUP334 EUP335 EUP336 EUP337 EUP338 EUP339 EUP340 EUP341 EUP342 EUP343 EUP344 EUP345 EUP346 EUP347 EUP348 EUP349 EUP350 EUP351 EUP352 EUP353 EUP354 EUP355 EUP356 EUP357 EUP358 EUP359 EUP360 EUP361 EUP362 EUP363 EUP364 EUP365 EUP366 EUP367 EUP368 EUP369 EUP370 EUP371 EUP372 EUP373 EUP374 EUP375 EUP376 EUP377 EUP378 EUP379 EUP380 EUP381 EUP382 EUP383 EUP384 EUP385 EUP386 EUP387 EUP388 EUP389 EUP390 EUP391 EUP392 EUP393 EUP394 EUP395 EUP396 EUP397 EUP398 EUP399 EUP400 EUP401 EUP402 EUP403 EUP404 EUP405 EUP406 EUP407 EUP408 EUP409 EUP410 EUP411 EUP412 EUP413 EUP414 EUP415 EUP416 EUP417 EUP418 EUP419 EUP420 EUP421 EUP422 EUP423 EUP424 EUP425 EUP426 EUP427 EUP428 EUP429 EUP430 EUP431 EUP432 EUP433 EUP434 EUP435 EUP436 EUP437 EUP438 EUP439 EUP440 EUP441 EUP442 EUP443 EUP444 EUP445 EUP446 EUP447 EUP448 EUP449 EUP450 EUP451 EUP452 EUP453 EUP454 EUP455 EUP456 EUP457 EUP458 EUP459 EUP460 EUP461 EUP462 EUP463 EUP464 EUP465 EUP466 EUP467 EUP468 EUP469 EUP470 EUP471 EUP472 EUP473 EUP474 EUP475 EUP476 EUP477 EUP478 EUP479 EUP480 EUP481 EUP482 EUP483 EUP484 EUP485 EUP486 EUP487 EUP488 EUP489 EUP490 EUP491 EUP492 EUP493 EUP494 EUP495 EUP496 EUP497 EUP498 EUP499 EUP500 EUP501 EUP502 EUP503 EUP504 EUP505 EUP506 EUP507 EUP508 EUP509 EUP510 EUP511 EUP512 EUP513 EUP514 EUP515 EUP516 EUP517 EUP518 EUP519 EUP520 EUP521 EUP522 EUP523 EUP524 EUP525 EUP526 EUP527 EUP528 EUP529 EUP530 EUP531 EUP532 EUP533 EUP534 EUP535 EUP536 EUP537 EUP538 EUP539 EUP540 EUP541 EUP542 EUP543 EUP544 EUP545 EUP546 EUP547 EUP548 EUP549 EUP550 EUP551 EUP552 EUP553 EUP554 EUP555 EUP556 EUP557 EUP558 EUP559 EUP560 EUP561 EUP562 EUP563 EUP564 EUP565 EUP566 EUP567 EUP568 EUP569 EUP570 EUP571 EUP572 EUP573 EUP574 EUP575 EUP576 EUP577 EUP578 EUP579 EUP580 EUP581 EUP582 EUP583 EUP584 EUP585 EUP586 EUP587 EUP588 EUP589 EUP590 EUP591 EUP592 EUP593 EUP594 EUP595 EUP596 EUP597 EUP598 EUP599 EUP600 EUP601 EUP602 EUP603 EUP604 EUP605 EUP606 EUP607 EUP608 EUP609 EUP610 EUP611 EUP612 EUP613 EUP614 EUP615 EUP616 EUP617 EUP618 EUP619 EUP620 EUP621 EUP622 EUP623 EUP624 EUP625 EUP626 EUP627 EUP628 EUP629 EUP630 EUP631 EUP632 EUP633 EUP634 EUP635 EUP636 EUP637 EUP638 EUP639 EUP640 EUP641 EUP642 EUP643 EUP644 EUP645 EUP646 EUP647 EUP648 EUP649 EUP650 EUP651 EUP652 EUP653 EUP654 EUP655 EUP656 EUP657 EUP658 EUP659 EUP660 EUP661 EUP662 EUP663 EUP664 EUP665 EUP666 EUP667 EUP668 EUP669 EUP670 EUP671 EUP672 EUP673 EUP674 EUP675 EUP676 EUP677 EUP678 EUP679 EUP680 EUP681 EUP682 EUP683 EUP684 EUP685 EUP686 EUP687 EUP688 EUP689 EUP690 EUP691 EUP692 EUP693 EUP694 EUP695 EUP696 EUP697 EUP698 EUP699 EUP700 EUP701 EUP702 EUP703 EUP704 EUP705 EUP706 EUP707 EUP708 EUP709 EUP710 EUP711 EUP712 EUP713 EUP714 EUP715 EUP716 EUP717 EUP718 EUP719 EUP720 EUP721 EUP722 EUP723 EUP724 EUP725 EUP726 EUP727 EUP728 EUP729 EUP730 EUP731 EUP732 EUP733 EUP734 EUP735 EUP736 EUP737 EUP738 EUP739 EUP740 EUP741 EUP742 EUP743 EUP744 EUP745 EUP746 EUP747 EUP748 EUP749 EUP750 EUP751 EUP752 EUP753 EUP754 EUP755 EUP756 EUP757 EUP758 EUP759 EUP760 EUP761 EUP762 EUP763 EUP764 EUP765 EUP766 EUP767 EUP768 EUP769 EUP770 EUP771 EUP772 EUP773 EUP774 EUP775 EUP776 EUP777 EUP778 EUP779 EUP780 EUP781 EUP782 EUP783 EUP784 EUP785 EUP786 EUP787 EUP788 EUP789 EUP790 EUP791 EUP792 EUP793 EUP794 EUP795 EUP796 EUP797 EUP798 EUP799 EUP800 EUP801 EUP802 EUP803 EUP804 EUP805 EUP806 EUP807 EUP808 EUP809 EUP810 EUP811 EUP812 EUP813 EUP814 EUP815 EUP816 EUP817 EUP818 EUP819 EUP820 EUP821 EUP822 EUP823 EUP824 EUP825 EUP826 EUP827 EUP828 EUP829 EUP830 EUP831 EUP832 EUP833 EUP834 EUP835 EUP836 EUP837 EUP838 EUP839 EUP840 EUP841 EUP842 EUP843 EUP844 EUP845 EUP846 EUP847 EUP848 EUP849 EUP850 EUP851 EUP852 EUP853 EUP854 EUP855 EUP856 EUP857 EUP858 EUP859 EUP860 EUP861 EUP862 EUP863 EUP864 EUP865 EUP866 EUP867 EUP868 EUP869 EUP870 EUP871 EUP872 EUP873 EUP874 EUP875 EUP876 EUP877 EUP878 EUP879 EUP880 EUP881 EUP882 EUP883 EUP884 EUP885 EUP886 EUP887 EUP888 EUP889 EUP890 EUP891 EUP892 EUP893 EUP894 EUP895 EUP896 EUP897 EUP898 EUP899 EUP900 EUP901 EUP902 EUP903 EUP904 EUP905 EUP906 EUP907 EUP908 EUP909 EUP910 EUP911 EUP912 EUP913 EUP914 EUP915 EUP916 EUP917 EUP918 EUP919 EUP920 EUP921 EUP922 EUP923 EUP924 EUP925 EUP926 EUP927 EUP928 EUP929 EUP930 EUP931 EUP932 EUP933 EUP934 EUP935 EUP936 EUP937 EUP938 EUP939 EUP940 EUP941 EUP942 EUP943 EUP944 EUP945 EUP946 EUP947 EUP948 EUP949 EUP950 EUP951 EUP952 EUP953 EUP954 EUP955 EUP956 EUP957 EUP958 EUP959 EUP960 EUP961 EUP962 EUP963 EUP964 EUP965 EUP966 EUP967 EUP968 EUP969 EUP970 EUP971 EUP972 EUP973 EUP974 EUP975 EUP976 EUP977 EUP978 EUP979 EUP980 EUP981 EUP982 EUP983 EUP984 EUP985 EUP986 EUP987 EUP988 EUP989 EUP990 EUP991 EUP992 EUP993 EUP994 EUP995 EUP996 EUP997 EUP998 EUP999 EUP1000</p>

16.7.2001

XML Metadata Standards and Topic Maps

19

Topic Map View (III)

ABCDEFGHIJKLMNPQRSTUVWXYZ	W3C ARCHITECTURE CENTER
<p>XLink (Xlink Data) XML Linking Language</p> <p>XLink defines how to insert links in XML documents. It specifies a framework making it possible for XML applications to recognize XML elements as having link semantics. In addition to the simple, two-ended, unidirectional links which are well-known from HTML, XLink allows more general links, which must not be embedded in the document, can have any number of ends, and can be multidirectional.</p> <ul style="list-style-type: none"> XLink is based on <ul style="list-style-type: none"> XML Namespaces XLink has been standardized by <ul style="list-style-type: none"> W3C XLink is authoritatively described at <ul style="list-style-type: none"> W3C XML Linking (external link) XLink is described at <ul style="list-style-type: none"> The XML Core Pages - XML Linking and Addressing Languages (XPath, XPointer, XLink) (external link) <p><small>document(s) with a view technology previously unknown</small></p>	<p>W3C XML Pointer, XML Base and XML Linking</p> <p>The XML Activity Statement explains the W3C's work on this topic in more detail.</p> <p>The XML Linking WG is looking for implementer feedback. If you have implemented or reused one of those specifications, please take the time to answer the related questionnaire, thanks.</p> <p>Check the implementation Chart.</p> <p>By unanimous decision from the XML Linking Working Group on 14th December 2000, the Issue List status has been changed from W3C Member only access to publicly available. The Issue List is now available!</p> <p>XML Linking and XML Base Reached Recommendation status on June 27th 2001.</p> <p>XML Linking Language (XLink)</p> <p>Description</p> <p>This specification defines the XML Linking Language (XLink), which allows elements to be inserted into XML documents in order to create and describe links between resources. It uses XML syntax to create structures that can describe the simple unidirectional hyperlinks of today's HTML, as well as more sophisticated links.</p> <p>W3C Status</p> <ul style="list-style-type: none"> Current version: Recommendation 1.0, June 27th 2001 Proposed Recommendation, December 20 2000 The XML CR DPC Expiration of Last Call comments Last Call 21 February 2000, last call ended 20 March Feb '99: W3C Note: XML XLink Requirements: Version 1.0 Jul '99: W3C Note: XML Linking Language (XLink) Design Principles Apr '97: initial XML Linking Working Draft XML Linking: An introduction by Steven J. DeRose working group: XML Linking Feedback: www.xml-linking-comments <p>Documentation and associated resources</p> <ul style="list-style-type: none"> discussion: also-der started Dec '98 archive: also-der_comp.html working group statements from the XML Linking WG 29 September 2000 Presentation at XML '99 Dec 1999 (somewhat outdated) XML Linking: An Executive Summary, by Eve Meyer Other Design documents: www.xml-linking.org

16.7.2001

XML Metadata Standards and Topic Maps

20

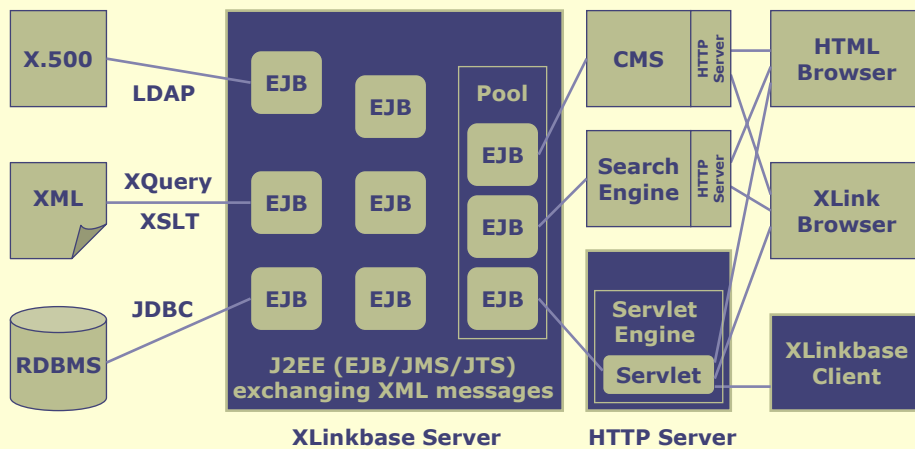
Comparison

RDF	Topic Maps
standardized by W3C	standardized by ISO
explicit data model	defined by syntax
properties have data types	associations aren't constrained
inherently distributed	centralized
separates schema from instance (resource description)	everything (almost...) is a topic, there are no types

What is missing?

- for Topic Maps only
 - a clean way to separate schemas and instances
 - a constraint language for topic associations
 - a way to distribute Topic Maps
- for RDF only
 - a unified data model with XML Schema
- for both approaches
 - tools for creating and managing metadata
 - a query language for actually using metadata
 - support from a wide range of vendors & users
 - an approach for achieving vocabulary consensus
 - smart ways to handle distribution

XLinkbase System Architecture



16.7.2001

XML Metadata Standards and Topic Maps

23

XLinkbase Status

- where the implementation is going
 - currently concentrating on EJB environment
 - hard to keep up with commercial engines
 - case study with simpler model & implementation
 - case study for generating DHTML links
- where the concept is going
 - proof of concept with the case study
 - 1 or 2 DAs dealing with Topic Map distribution
 - looking into data model improvements
 - constraint language for associations
 - schema/instance separation or separability

16.7.2001

XML Metadata Standards and Topic Maps

24