Interoperability - Technical Standards and New Developments

Michael Wilson
W3C Office in the UK
CLRC Rutherford Appleton Laboratory

Talk Structure

- Data Usage Process
- RDF
- Dublin Core
- Conclusion

Data Interoperability - The Standard Solution

- Metadata Format
- Controlled Vocabulary
- Common Access Protocol
- Uniform User Interface
- But … too many Metadata formats & vocabularies.
- access through WAIS, Z39.50, HTTP etc..
- No common UI for metadata search, data browsing, statistics & visualisation packages

- Which technology to choose ?

Usage Process

Action	**build houses, roads. establish schools, hospitals**
Decision	**Is there a problem? Which solution is effective? Is expenditure on this solution more efficient than solving another problem?**
Knowledge	**What data/information is required to make a decision ? Analysis, diagnosis etc.. ; Synthesis, planning etc.. What are the analogies to this case ?**
Information	**Statistical analysis; Visualisation Information retrieval of previous knowledge, decisions etc..**
Data	**Store & Query Data**

Distributed Access

- Wide Area Distributed Access is Required
- Therefore Internet & WWW - W3C standards
- W3C Activities
  - separate data from presentation - not HTML V3.2
  - increase the semantic access to information
  - maximise range of presentation options -
    resolution, size, nationalisation, bandwidth
    PC, TV, mobile phone, car IS, fridge etc..
  - Layers of Languages, modules & profiles
    - slim clients containing only required modules

Human Usable & Machine Interoperable

Action	Robots, process control Finance trading system	
Decision	Expert Systems	
Knowledge	Ontologies, Metadata Rule Bases, KMS	Resource Description Framework (RDF)
Information	IR Systems; Stats & Visualisation Tools	XHTML, SMIL, SVG, MathML, ChemML
Data	DBMS	eXtensible Markup Language (XML)

RDF - Example

- a 'resource', http://doc, has a 'property', author, describing some aspect of it. The value of the 'property' is Joe Smith.
- Joe Smith is the author of http://doc.
- Beyond controlled vocabulary RDF can be used to define the semantics of an ontology

Resource Description Framework

- RDF is the W3C recommendation for metadata to describe resources available over the WWW
- It is like a mid-1980's Knowledge Representation Semantic Network Language - with reification
- It is best thought of as a structured graph model with nodes and links
- The Nodes represent RDF Resources while the links (arcs) represent RDF Properties describing the attributes and relationships of the resources.
- Properties and Resources are identified as URI's drawing upon multiple namespaces and vocabularies

RDF & XML

- RDF Model is independent of XML
  - It is a higher level model over XML: XML is Syntax, RDF is an Object Model
- RDF data may, or may not, be stored in XML
- All processing can be done at a higher level in RDF before conversion to XML if necessary
- XML conversion may be necessary since most web systems understand XML
- RDF evolves the Warwick Framework for metadata vocabularies, where a single model and syntax are used.

Epistemologically Backwards

- Attributes are first class entities
  - objects are only second class objects
- NOT A document is an object with a creator, title, publisher, date, language etc..
- BUT The attributes creator, title, publisher, date, language etc.. combine in the object document
- This appears backwards to those used to object centred design & programming
- But it allows anything to be said about existing resources, by anybody

Viewing RDF

- 3 ways to look at RDF
  - Diagramatic Representation
  - XML Serialisation Syntax

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://doc">
      <author> Joe Smith </author>
  </rdf:Description>
</rdf:RDF>
```

  - RDF Statements - triples

{“http://doc”,author,x}

{x,author, “Joe Smith”}

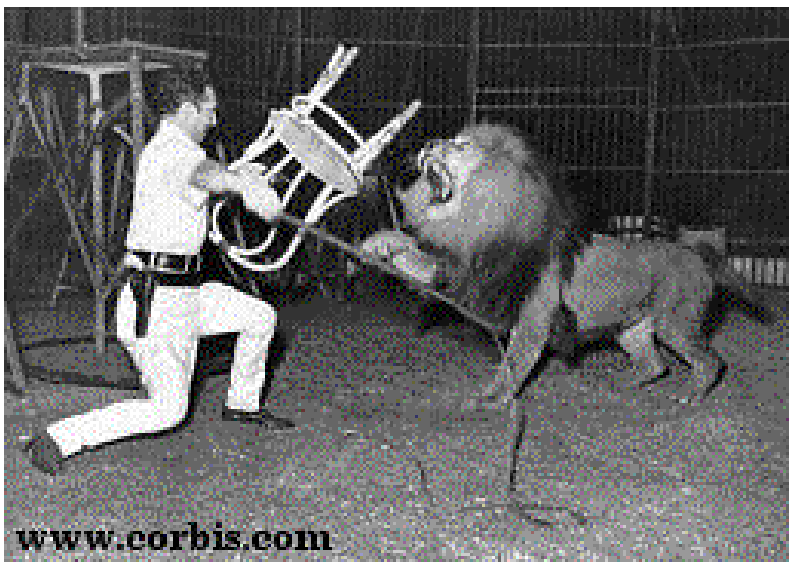
RDF Schemas

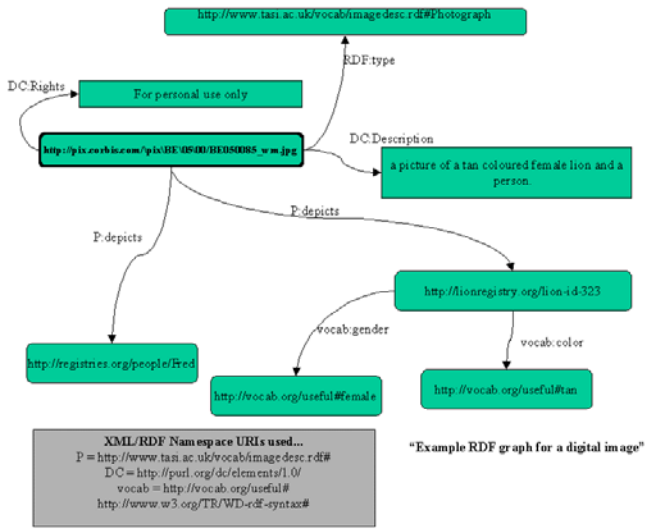
- RDF Schema provides and extensible object model and type system for RDF
- Simpler to implement than full predicate calculus languages such as CycL or KIF.
- It defines constraints on the property types and their values
- e.g. - this property can only by applied to Minivans
  - <rdfs:domain rdf:resource="Minivan"/>
- e.g. - values for this property must be numbers

<rdfs:range rdf:resource="http://www.w3.org/datatypes#Number"/>

Self Describing Images & RDF

- GIF image & RDF text file combine into a single PNG image using giftopnm and pnmtogif tools





- http://www.tasi.ac.uk/building/note_rdfmeta.html

RDF Tools

- W3C SiRPAC - RDF viewer, syntax checker & triple producer
  http://www.w3.org/RDF/Implementations/SiRPAC
- DSTC (Australia) Reggie RDF Metadata editor
  http://metadata.net/dstc/
- Automatic web page metadata generator in DDC
  http://www.scit.wlv.ac.uk/~ex1253/rdf_paper/
- IBM's RDF4XML - creating, manipulating, storing, querying & transformng RDF
  http://www.alphaworks.ibm.com/formula/rdf4xml
- Netscape Mozilla - in Communicator 4.5

XML Namespaces & standard terminologies

- content providing communities can declare their own definitions for the description of resources of importance to them
- a single description may comprise elements drawn from any number of other accessible recording practices
- an XML Namespace provides context for any resource description element
- E.G. - the Dublin Core namespace for digital libraries, the WHO namespace for medical terminology etc…
- Similarlay a label in one language (e.g. French) may be linked to the authoritative definition of the concept elsewhere (e.g. UKDA)
- An RDF definition will declare the namespaces used at the beginning - for example to include the RDF & Dublin Core namespaces

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.0/">
```

Dublin Core & RDF

- 15 elements core to metadata definitions for resource discovery - not retrieval or request
- Agreed at NCSA March 95, trialed widely
- To promote global interoperability, element descriptions may be associated with a controlled vocabulary for the respective element values
- Tool support e.g. - http://www.ukoln.ac.uk/metadata/dcdot/
- Translations available in various languages
- Defined in RDF to produce RDF metadata
- http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/

Interoperable Heritage Metadata

- A resource description can be built up of vocabulary elements from different metadata formats -
- Dublin Core - Identify Resource
- VRA Core - visual documents
- Object ID - track stolen items
- CIDOC Data Model - list of DB fields
- FDA/ADAG - architectural drawings
- MESL - site licensing info
- CDWA - Full heritage taxonomy
- USMARC - generic publication details

http://www.getty.edu/gri/

Conclusion

- Too many metadata technologies to choose between
- Different Subject areas have the same problem - medicine, heritage, science, libraries
- W3C standards have a good track record
- XML appears to be adopted
- Cross domain interoperability requires use of common metadata and ontology
- RDF has attracted a lot of interest - expressive
- Dublin Core is picking up users
- No formal method of subsidiarity to standardise metadata in different domains