An Introduction To The Semantic Web

Brian Kelly

UK Web Focus

UKOLN

University of Bath

Email

`B.Kelly@ukoln.ac.uk`

URL

`http://www.ukoln.ac.uk/`

UKOLN is supported by:

re:source The Council for Museums Archives and Libraries

Joint Information Systems Committee
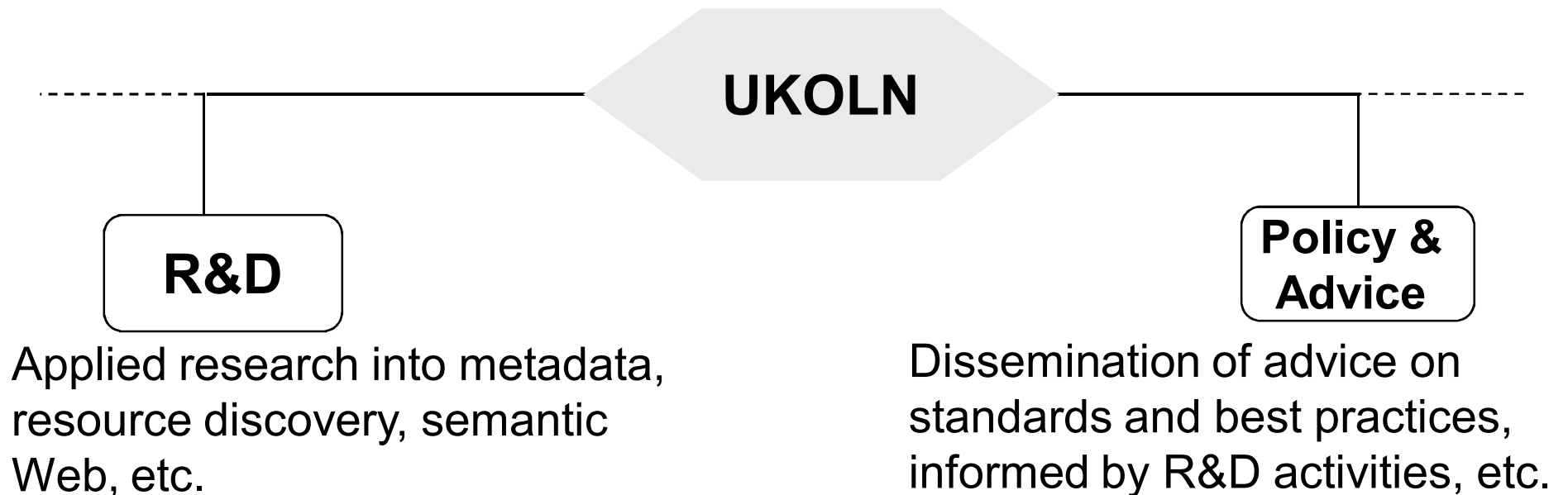
UNIVERSITY OF BATH

UKOLN

Contents

- Introduction
- Development Of A Metadata Architecture For The Web
- Metadata, RDF And The Semantic Web
- Critique
- Where To From Here?
- Conclusions

About UKOLN

UKOLN:

- A national focus of expertise in digital information management

- Provides policy, research & awareness services to UK library, information & cultural heritage communities

- Based at University of Bath

UKOLN

R&D

Policy & Advice

Applied research into metadata, resource discovery, semantic Web, etc.

Dissemination of advice on standards and best practices, informed by R&D activities, etc.

Remember When ...

The Web:

- Launched in early 1990s
- Exponential growth in mid-1990s
- Search engines took off as tool for finding resources
- Self-publishing increased amount of low quality hits from search engines
- Spammers (e.g. porn companies) attempt to fool search engines

Difficulties in using the Web to find relevant resources

Librarians To The Rescue

In the mid 1990s:

- Librarians, computer scientists, etc. got together to try to find a solution to resource discovery problem

- First meeting organised by OCLC in Dublin, Ohio led to the **Dublin Core** set of attributes for resource discovery

- *"This metadata – it's just cataloguing isn't it?"*

- Needs to be cross-sectoral, scalable & extensible

Title: How Web Was Born
Author: Robert Cailliau
ISBN: …
Classification:

DC.Title: Mona Lisa
DC.Creator:
    Leonardo da Vinci
…

Representing Dublin Core

Dublin Core (DC):

- Consists of 15 core attributes for resource discovery

- Documented at <http://dublincore.org/documents/dces/>

Name: Title
Identifier: Title
Definition: A name given to the resource.
Comment: … a name by which the resource is formally known

- Neutral on how DC should be represented

- HTML found to be inadequate for representing complexities of structured use of DC

Meanwhile

Also in the mid 1990s:

- Development of PICS for resource labelling (e.g. porn) in response to threat of US CDA legislation:
  - "*This page contains nudity*" described in a **machine-understandable** way
- Work on digital signatures:
  - "*This bill was signed by Bill Clinton and is legally binding*"
- Work on privacy:
  - "*This Web site will store your personal information only to ensure goods by be delivered / will give you a 10% discount if you let us send you info on further offers* "
- etc.

Realisation that:

- This is all metadata
- Need to provide a common metadata framework

XML

Around same time XML (Extensible Markup Language):

- Developed within W3C

- Provided a lightweight alternative to SGML

- Allows for extensibility

- To be used for development of all new formats within W3C

- Accompanied by much related work:

☐	XML Schemas	cf DTDs
☐	XML Namespaces	my `<title>` means "Mr." and not title of CD
☐	XLink and XPointer	Better hyperlinking
☐	XSLT	Transform XML resources
☐	…	

XML Is Not Enough!

XML:

- Should be used
- Is extensible (DC qualifiers)

```
<DC.Creator>Brian Kelly</DC.Creator>
<DC.Creator.email>B.Kelly@ukoln.ac.uk</DC.Creator.email>
```

But:

- XML describes the syntax
- Does not provide semantics (what does DC.Creator *mean*?)
- The meaning may be agreed & understood within DC applications – but this does not allow for *extensibility*
- Similar applications may be described using different XML DTDs: e.g. is `<Creator>` the same as `<le-Créator>` or `<Доклады>`

Scenario – Buying A Car

User

> You live in London and want to buy a car locally. You can afford up to £500. The car must be red.

Honest EuroJoe's Used Car Web Site

> Joe uses:

```
<car>
      <location>Brixton</location>
      <price>€400</price>
      <colour>maroon</colour>
      <description>Old banger</description>
      <model>Ford Escort</model>
</car>
```

Result

- Car not found – even though structured information is provided.
- A human would know that this was a valid match, because it understands the meanings and relationships.
- The Semantic Web aims to solve this problem.

We Need Extensibility!

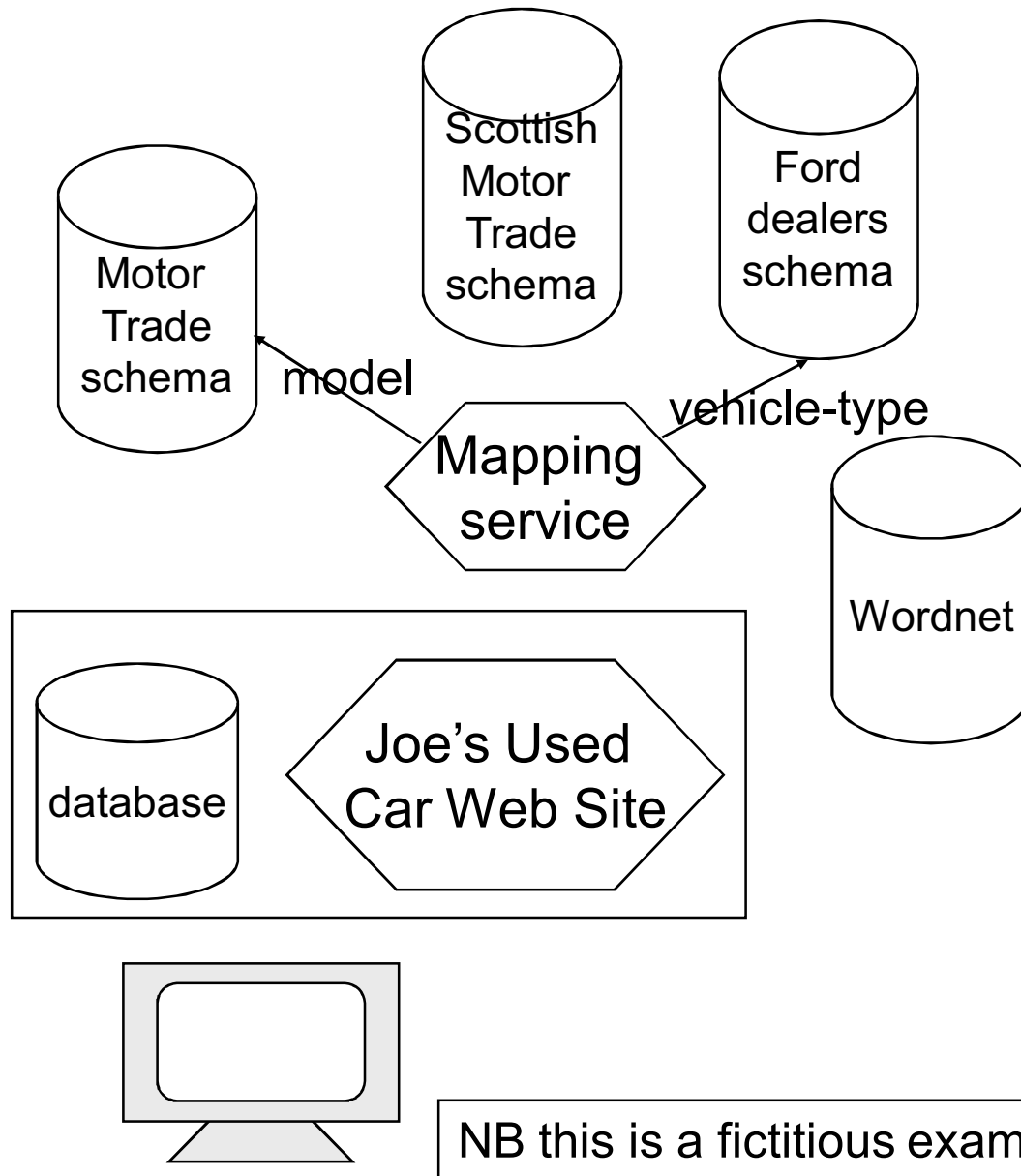We can see a progression from Web sites which are:

- Understandable by humans
- Understandable by software "in the know"

```
<h1>Joe's Used Cars</h1>
<h2>Ford Escort</h2>
<p>This maroon car costs €400
```

```
<company>Joe's Used Cars</company>
<model>Ford Escort</model>
<colour>maroon</colour>
<price>€400</price>
```

We need a mechanism which allows equivalent resources to be identified, without programming this knowledge into software

Buying Car On The Semantic Web

Scottish Motor Trade schema

Ford dealers schema

Motor Trade schema

model

vehicle-type

Mapping service

Wordnet

database

Joe's Used Car Web Site

NB this is a fictitious example

Joe is part of the motor trade association, which has defined its own schema for selling cars.

The Scots use a different schema, as do the car manufacturers (which mainly sell new cars).

A mapping service provides a mapping between these machine-understandable schemata.

Wordnet maps relationships between words (e.g. red and maroon)

The Semantic Web

A Vision Of Possibilities

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

-- Tim Berners-Lee, James Hendler and Ora Lassila, The Semantic Web, Scientific American, May 2001

Scenario – Buying A Car (2)

We've seen how this query can be answered:

Find me a red car in London for < £500.
How about this maroon Escort in Brixton for €400?

The Semantic Web will be extensible enabling interactions with other services which may use different XML DTDs:
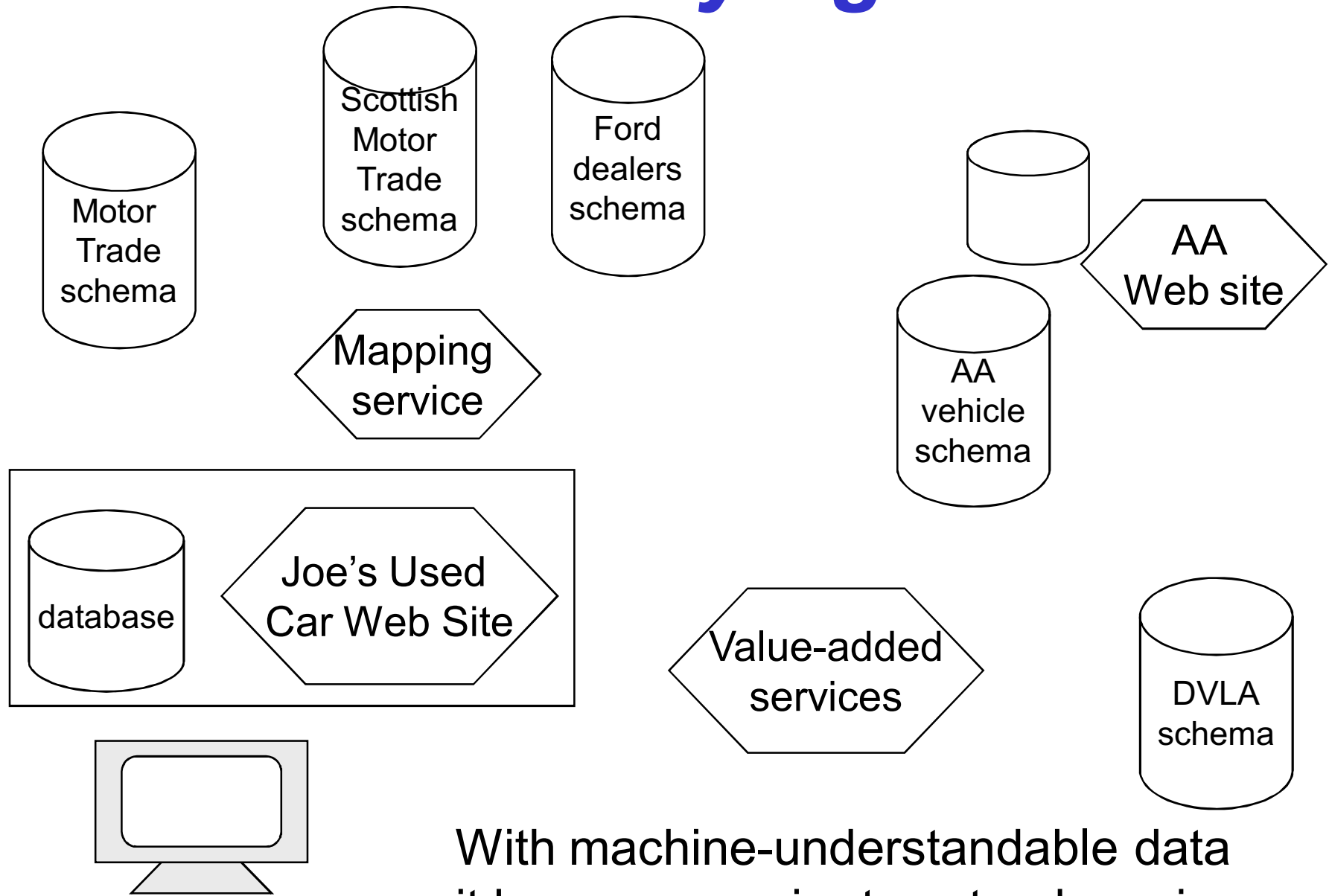
Give me the AA's report on this type of car.
OK here it is
Check the DVLA details for the reg. no.
OK – the car is registered correctly

Model For Buying A Car

Motor Trade schema

Scottish Motor Trade schema

Ford dealers schema

AA Web site

Mapping service

AA vehicle schema

database

Joe's Used Car Web Site

Value-added services

DVLA schema

With machine-understandable data it becomes easier to extend services

RDF

RDF:

- Resource Description Framework
- An XML application
- "Not just tags" – RDF makes use of a formal model
- Basis for "The Semantic Web" (SW)

RDF Data Model

Resource —PropertyType→ Value

Property

05-Mar-02

↑ on

page.html —written by→ Brian
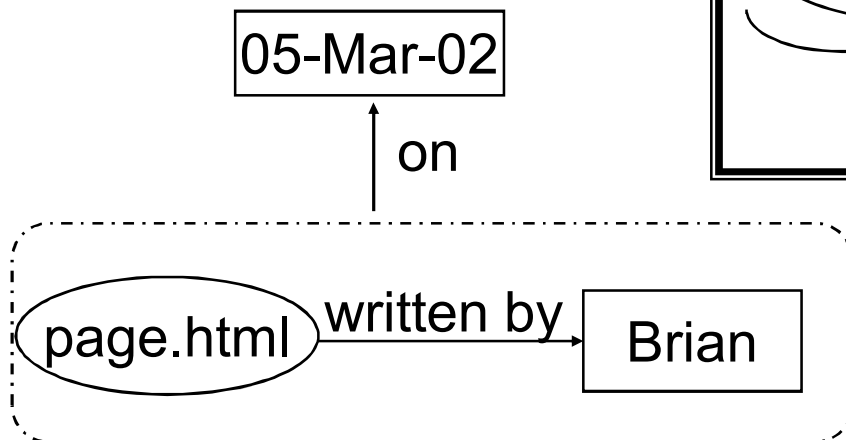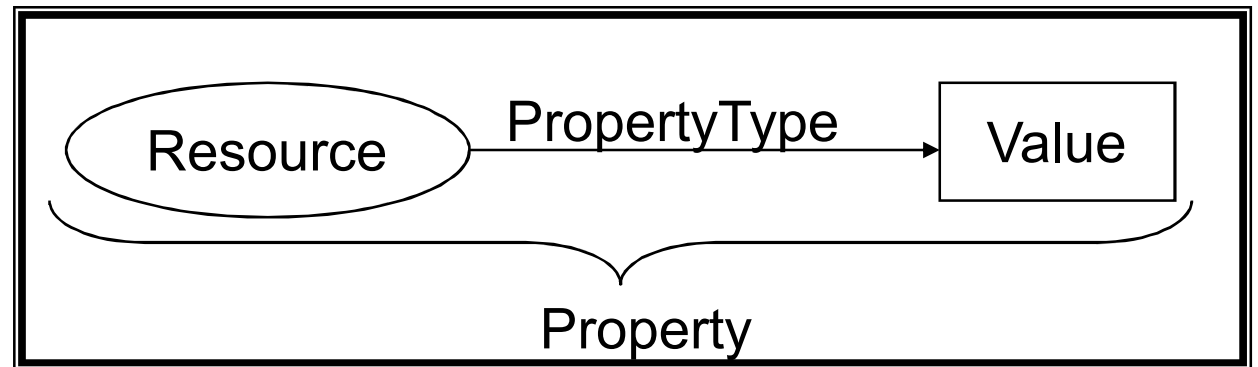
Resource   has property value
Page.html   written-by   Brian

Known as *triples* or *tuples*

Ontologies

XML DTDs:
- Document Type Definition
- Define structure: **Car** application contains a **price** (integer), **description** and **colour**

XML Schemas:
- Allows richer definitions
- Define structure: **Car** application contains a **price** (+ve integer between 1 and 20,000), **description** and **colour** (taken from fixed vocabulary)

Ontologies:
- Define relationships: relationship between, say, a postcode, a town, a suburb, etc
- Builds on AI techniques

UKOLN

Importance Of URIs

The Semantic Web will build on the distributed nature of the Web:

- No central naming authority

Schema definitions:

- Not implied in applications (cf. Web browsers and HTML DTD)
- Accessible in a machine-understandable format using a URI

SW Scenario

"At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules."

(The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web.)

Scientific American: The Semantic Web: May 2001
<http://www.sciam.com/2001/0501issue/0501berners-lee.html>

Public Sector Scenario

Highways
Road works which will force changes to bus routes

DTLR
Transxchange schema for public transport – "exchange of bus timetable information ... Provide a national passenger transport information system".

Public Utilities
Road works which will force changes to bus routes

Schools
Information about bus times

Community Information
Information about bus times

How do we ensure that all services provide up-to-date information?
- Agreement across different sectors?
  - Difficult (different histories, cultures, timing, ownership, etc.)
- Describe schema so that software can automatically process information

Semantic Web

In the Semantic Web we will need:

- Machines talking to machines – semantics need to be unambiguously declared

- Joined-up data – enabling complex tasks based on information from various sources

- Wide scope – from, say, home to government to commerce

- Trust – both in data and who is saying it

This is not going to be easily achieved

What's Needed

Semantics

Shared schemas: conventions about declaring meaning

Agreed ontologies (both terms and 'rules' as to how terms relate)

Agreed data model (RDF)

Infrastructure

Schema Registries to share schemas

Common syntax (XML)

The Web for connectivity: URI, HTTP...

UKOLN

Reality Check (1)

Reservations have been expressed about the SW:

It's Too Complex

- The Web took off because page creation was simple
- The RDF model is felt to be complex
- The RDF representation in XML looks complex

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:foaf="http://xmlns.com/0.1/foaf/" >
   <rdf:Description rdf:about="">
      <dc:creator rdf:parseType="Resource">
         <foaf:name>Sean B. Palmer</foaf:name>
      </dc:creator>
      <dc:title>The Semantic Web</dc:title>
   </rdf:Description>
</rdf:RDF>
```

This says that the article has the title "*The Semantic Web*",
and was written by someone whose name is "*Sean B. Palmer*"

Reality Check (2)

Industry Isn't Interested

- The Semantic Web won't take off unless the IT sector develops tools

Its Too Researchy

- The Semantic Web is an idea for the AI research community and not for mainstream use

W3C Using A Flawed Knowledge Representation Model

- At the WWW 10 conference "the ontologists met the Web geeks" and told them they'd got their knowledge representation model wrong

Consensus Not Yet Reached On Architectural Approach

- There is still debate on RDF, patent issues, etc.

Other Issues - Trust

In our used car example we have:

>   Car ▯ has-colour ▯ maroon

But can we trust the person who made the statement (they may be colour-blind)

>   Car ▯ has-previous-owners ▯ 1
>   Owner ▯ has-sex ▯ lady
>   Owner ▯ has-occupation ▯ vicar
>   Car ▯ has-mileage ▯ 10,000

Can we trust these statements?

For the Semantic Web to be scalable we will need a *Web of trust*

>   Car-company ▯ has-status ▯ CA-approved

See E-GIF document on "Trust Services Framework" at
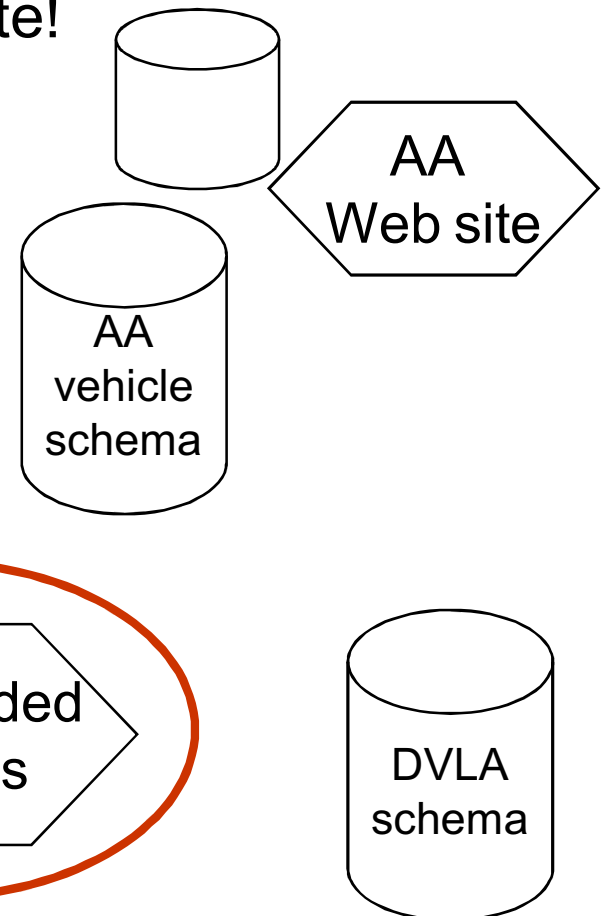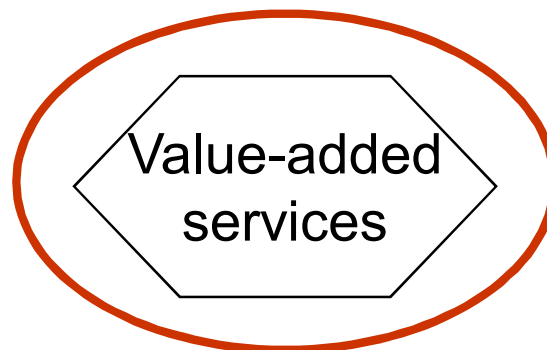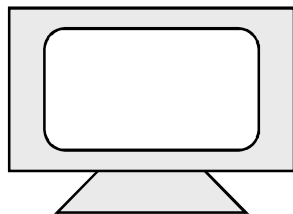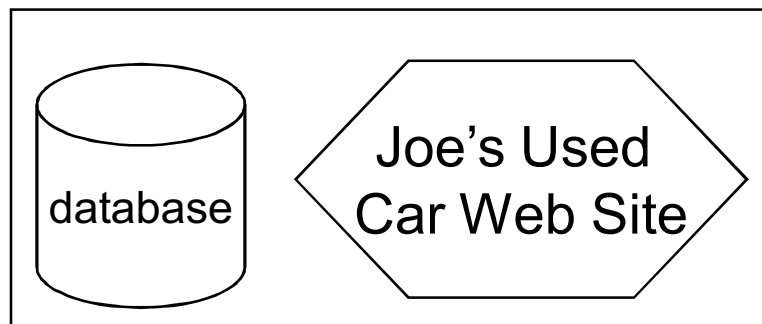<http://www.govtalk.gov.uk/rfc/rfc_document.asp?docnum=469>

Other Issues - Business Model

Machine-understandable schema definitions should be derived automatically,

But who will develop value-added schemas, mapping services, etc. – e.g. Joe's `<model>` is the same as Ford's `<vehicle-type>`?

This could drive users away from Joe's Web site!

Will the public sector be better placed to implement such services ?

Responses

Complexity

- RDF in XML looks complex. So what? Postscript looks complex. Software will generate RDF.
- RDF modelling (arcs, nodes, tuples) is complex. So is database modelling (SQL, third normal forms, etc.) but that doesn't stop relational databases being a multi-billion dollar market.

Industry View

- Is changing? Interest from companies such as IBM.
- Links with "Web services" and "The Grid"- which look to be a major growth areas.
- Will recognise:
  - the cost implications of not doing this
  - the dangers of multiple, non-interoperable Semantic Webs (Microsoft camp and Sun camp)

RDF Developments

We have seen that:

- RDF looks complex
- There are still some uncertain areas

Let's now look at:

- A simple RDF application
- Browser support
- Project work
- Related work which may:
  - require the Semantic Web
 - be used to build the Semantic Web

A Lightweight RDF Application

RSS (RDF Site Summary):

- Example of a lightweight RDF application
- A format for news syndication
- Worth looking at for:
    - News syndication
    - Gaining experience of an RDF application
- **Note** beware of versions – RSS 0.9 is not RDF, but RSS 1.0 is
- See:

    <http://blogspace.com/rss/>

   <http://www.oreillynet.com/rss/>

    <http://www.webreference.com/authoring/languages/
      xml/rss/intro/>

Browser Support

The **Mozilla** open source browser is using RDF to integrate and aggregate Internet resources.

http://www.mozilla.org/rdf/doc/
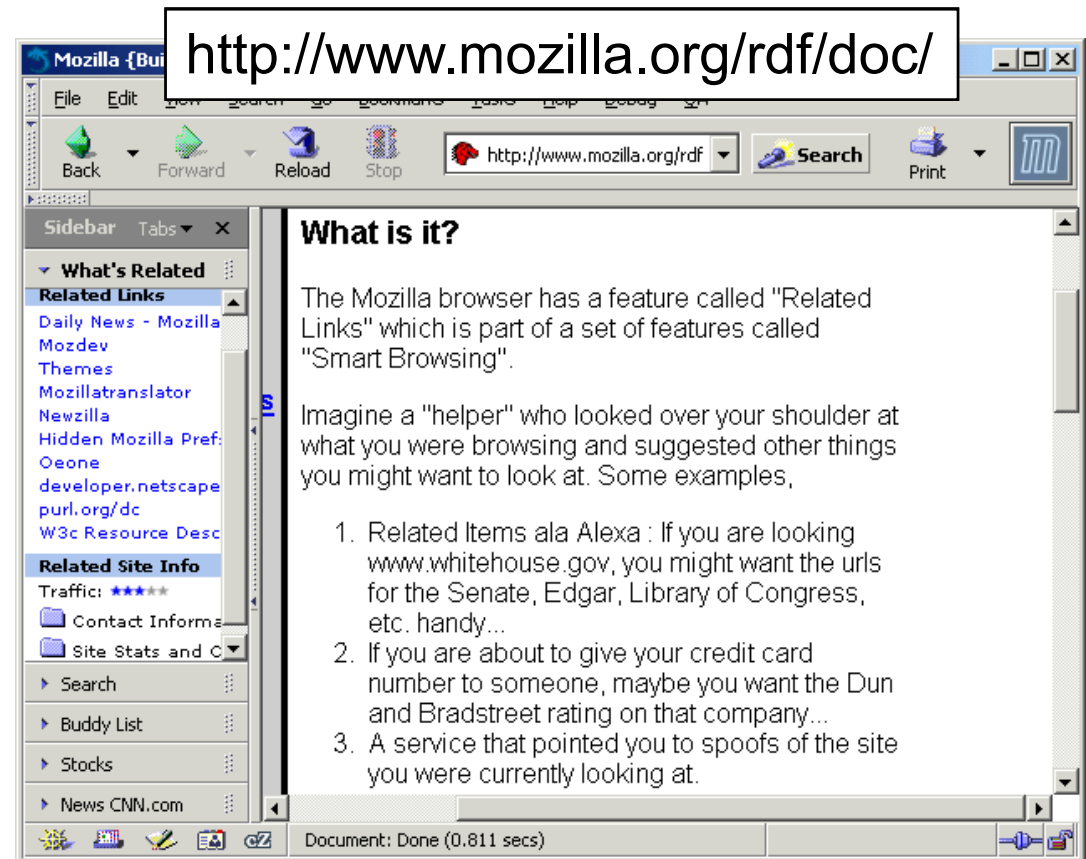
SCHEMAS

UKOLN involved in the EU-funded SCHEMAS project:

- Provides a forum for metadata schema designers
- Will inform schema implementers about the status and proper use of emerging metadata standards
- Provides a registry of metadata schema
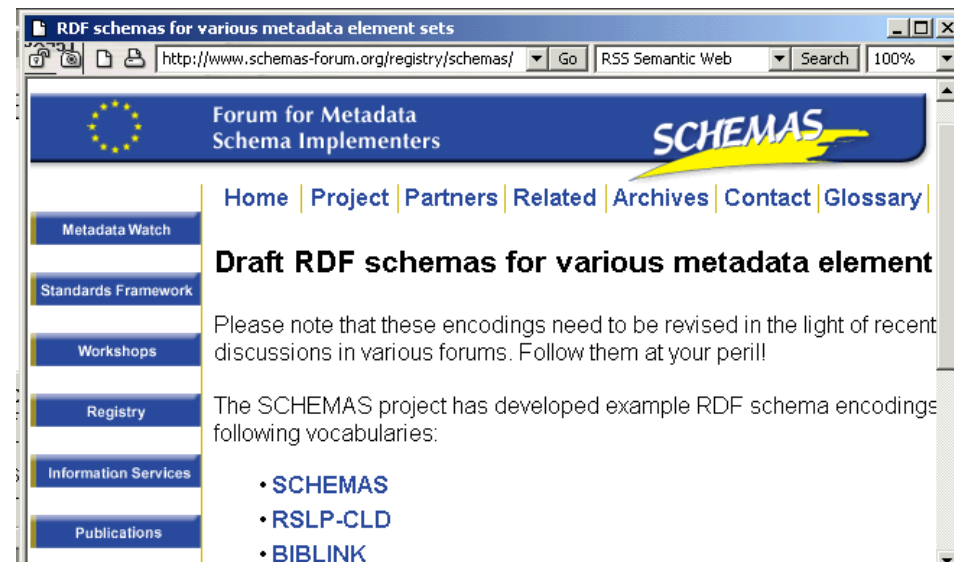
See <http://www.schemas-forum.org/>

Related Work – Web Services

Web Services:

- Foundations for automated use of the Web
- Three standards normally mentioned:
  - **SOAP** (Simple Open Access Protocol)
    Enables rich messages to be send across HTTP
    (compare with http query string:
    http://foo.com/get?userid=Fred&profile=bar)
  - **WSDL** (Web Services Description Language)
    A specification for describing the operational
    information of a Web service, such as interface
    See <http://xml.coverpages.org/wsdl.html>
  - **UDDI** (Universal Description, Discovery and
    Integration)
    A framework for describing services, discovering and
    integrating services
- Provide building blocks for Semantic Web

Related Work – The GRID

The GRID:

- Access to distributed resources (computation, storage, databases, etc.)
- Much interest in areas such as Physics, Chemistry, etc.
- Likely to utilise Web Services (find this molecule, find services which can apply this technique on it, …)
- Likely to need a Semantic Web in order to allow software to understand descriptions of database, etc.
- See `<http://www.gridforum.org/>`

Related Work – Ontology

The WWW 10 conference (Hong Kong, May 2001) provided a forum for the Web and Knowledge Representation community. We now have:

- Close links between two communities
- Building on **DAML** (DARPA Markup Language) work in US and **OIL** (Ontology Inference Layer) funded by EU
- **DAML+OIL**: a semantic markup language for Web resources which builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modelling primitives. See:

    <http://www.w3.org/TR/daml+oil-walkthru/>
    <http://www.w3.org/TR/daml+oil-reference>

Where To From Here? (1)

What should we be doing now?

- Ensure that your information is stored in a neutral, structured way:
    - Reuse of resources (e.g. digital TV, PDA, etc.)
    - Remember that HTML is an output format (not for storage)
- Use management tools (e.g. CMS, RDBMS)

This will enable you to manage existing services more effectively:

- Manage information
- Exploit new devices
- Provide new functionality (personalisation, etc.)

Where To From Here? (2)

Best Practice For Today's Web

- Ensure that your resources are well-structured and managed

Moving To The Semantic Web

You will need to:

- Ensure that you have definitions of the structure of your resources

- Agree on these definitions at an appropriate level (international agreement is good from a technical perspective, but agreements may be difficult!)

- Provide these descriptions in a machine-understandable format

Note work on E-GIF XML Schema Registry – see <www.govtalk.gov.uk/documents/schema%20register.doc>

But Should We Do It?

Should local government make use of the Semantic Web?

How do you see the future of your Web services if you don't?

- Inter-departmental wars which aim to provide a standard descriptions of data
- HTML scraping – application specific technique which needs to be rewritten when the interface changes
- Failure to implement the joined-up Government vision
- Difficulties in interoperating with services outside the public sector

Conclusions

To conclude:

- The first version of the Web lacked a metadata framework which was needed to describe resources
- W3C developed RDF to provide this framework
- As well as providing an framework for metadata applications, RDF allows software to reach beyond individual Web sites
- The Semantic Web will be based on registries of machine-understandable definition
- There is significant investment in related areas such as the GRID and Web Services
- The Semantic Web will be difficult to achieve
- It will be expensive to provide rich interoperable services without a Semantic Web

Find Out More (1)

Semantic Web, W3C
<http://www.w3c.org/2001/sw/>

Semantic Web Road map, Tim Berners-Lee
<http://www.w3c.org/DesignIssues/Semantic.html>

The Semantic Web, Scientific American
<http://www.sciam.com/2001/0501issue/0501berners-lee.html>

The Semantic Web Community Portal,
<http://www.semanticweb.org/>

The Semantic Web: A Primer
<http://www.xml.com/pub/a/2000/11/01/semanticweb/>

All found using Google to search for "*semantic Web*"

Find Out More (2)

An introduction to RDF

<http://www-106.ibm.com/developerworks/xml/library/w-rdf/>

The Semantic Web Community Portal

<http://www.semanticweb.org/>

The Semantic Web: An Introduction

<http://infomesh.net/2001/swintro/>