Standardization for the Semantic Web

Thomas Severiens

Severiens@Mathematik.Uni-Osnabrueck.de

Semantic Web – the vision

- Scenario: Someone needs to see a doctor.
  - Tim Berners-Lee, Scientific American 2001
  - Handheld devices with “agents” negotiate:
    - Doctor’s agent knows prescribed treatment
    - List of doctors, rated at least “very good”, within 20 miles, accepted by insurance plan.
    - Devices match calendars to fix appointment.
  - Semantic Web challenge: provide language that expresses both data and rules for reasoning about that data

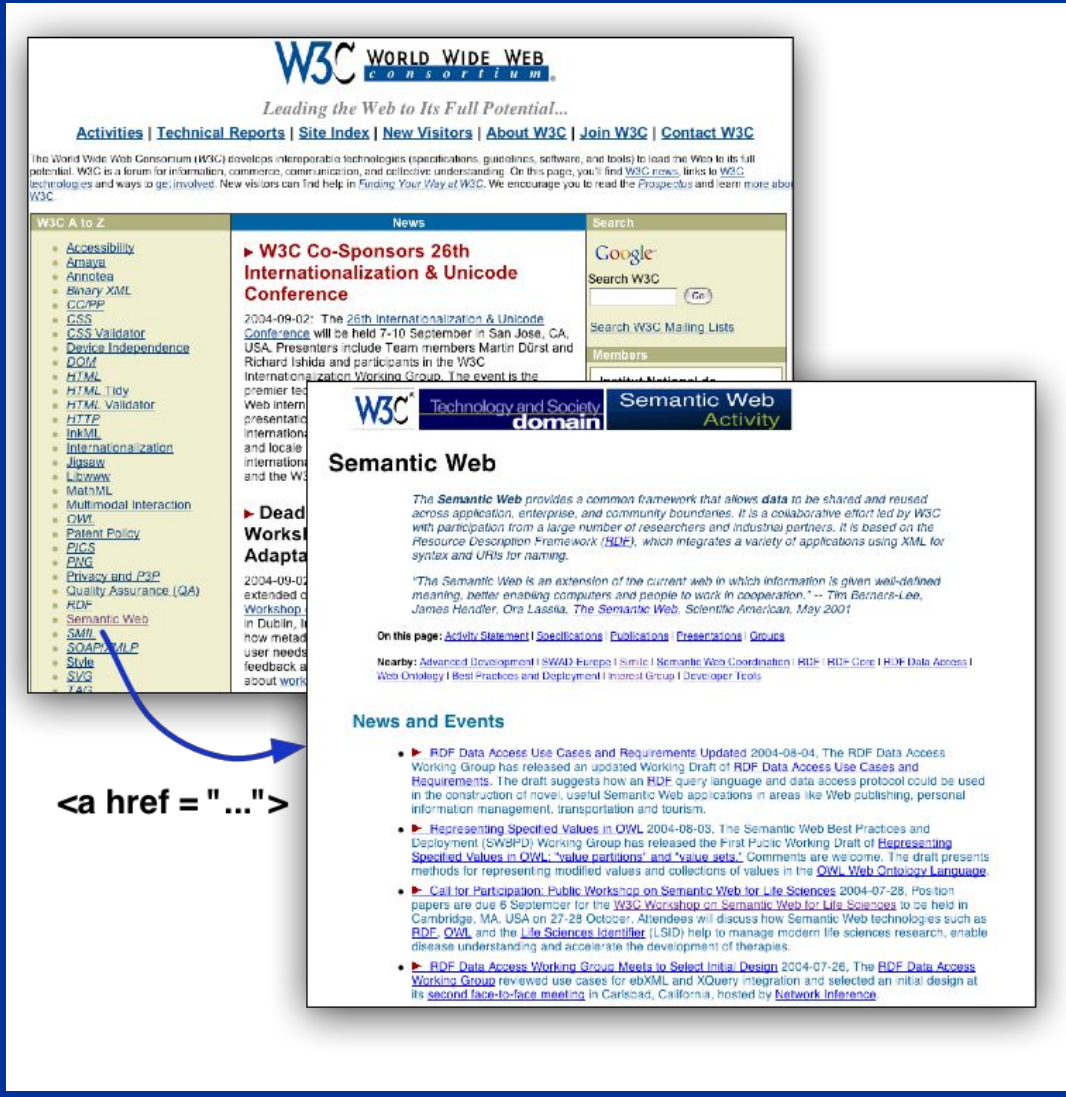
From a presentation of Thomas Baker

Semantic Web – the vision

- The Web of Data
  - Integration of multiple sources of data and services to draw new conclusions.
  - Describe and manage objects (items, collection, processes etc.) so they can be reused at verious scale.
  - Build structure behind the content, to allow preservation and retrieval by making the implicit structures explicit.

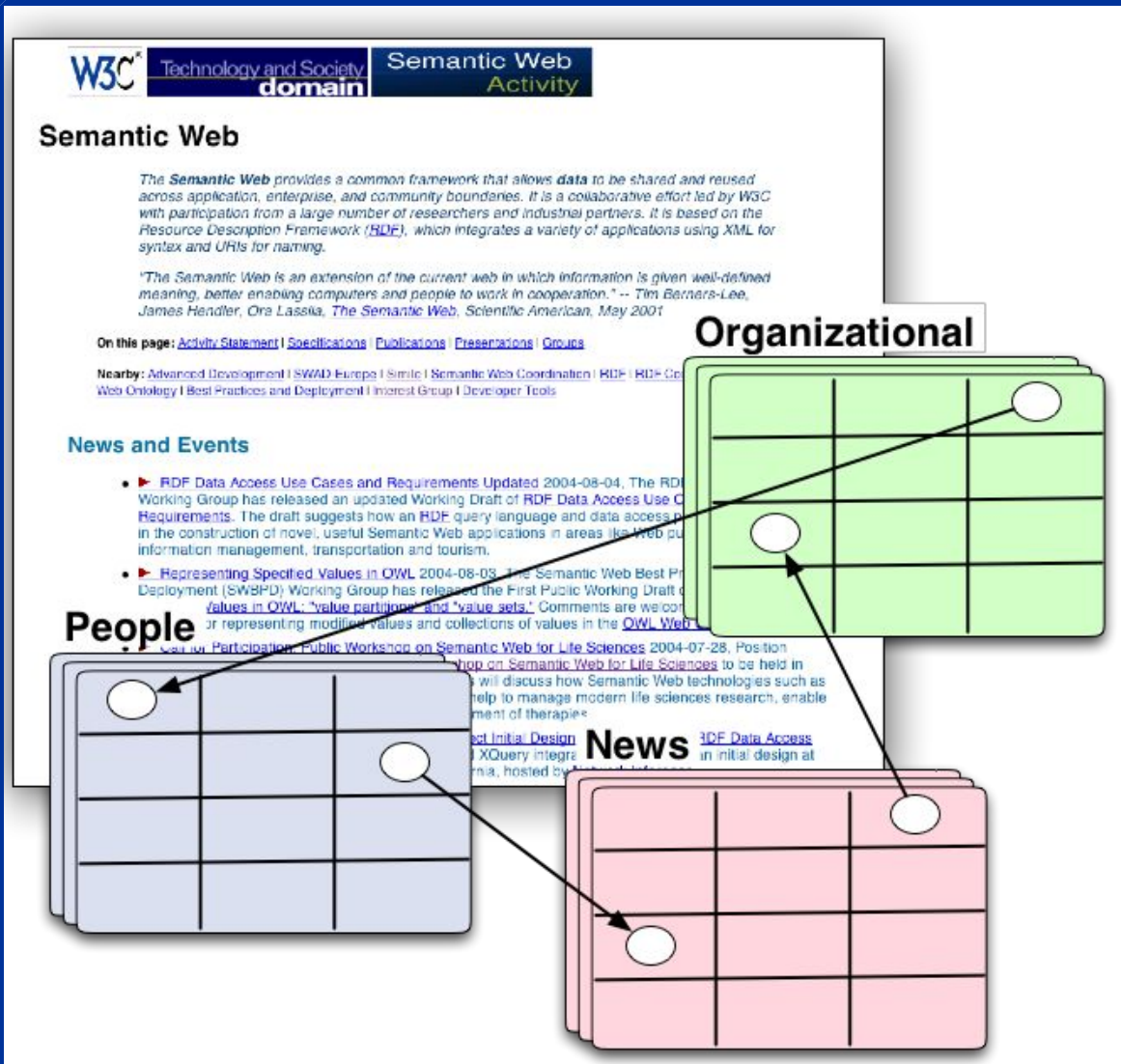
Semantic Web – the vision

– The Web of Data
  - HTML and URL
  - Text-level



Semantic Web – the vision

– The Web of Data
  - RDF, OWL, URI
  - Below Text-level
  - Data-level



Semantic Web – Languages

- RDF – Resource Description Framework
  - See http://www.w3.org/TR/rdf-primer/
- OWL Web Ontology Language
  - See http://www.w3.org/TR/owl-features/
- Core specifications are W3C Recommendations as of February 2004:
  - See http://www.w3.org/2001/sw/

Semantic Web – Fulltext-Semantics

- OMDoc www.openmath.org/omdoc generate semantic MathML from LaTeX-like sources
- CML www.xml-cml.org describe Objects in Chemistry giving their bindings and reaction parameters
- E = m c² ??? Need for Physics Semantic Markup!
- Some nice examples from www.vs-c.de ...

Institutional Repositories

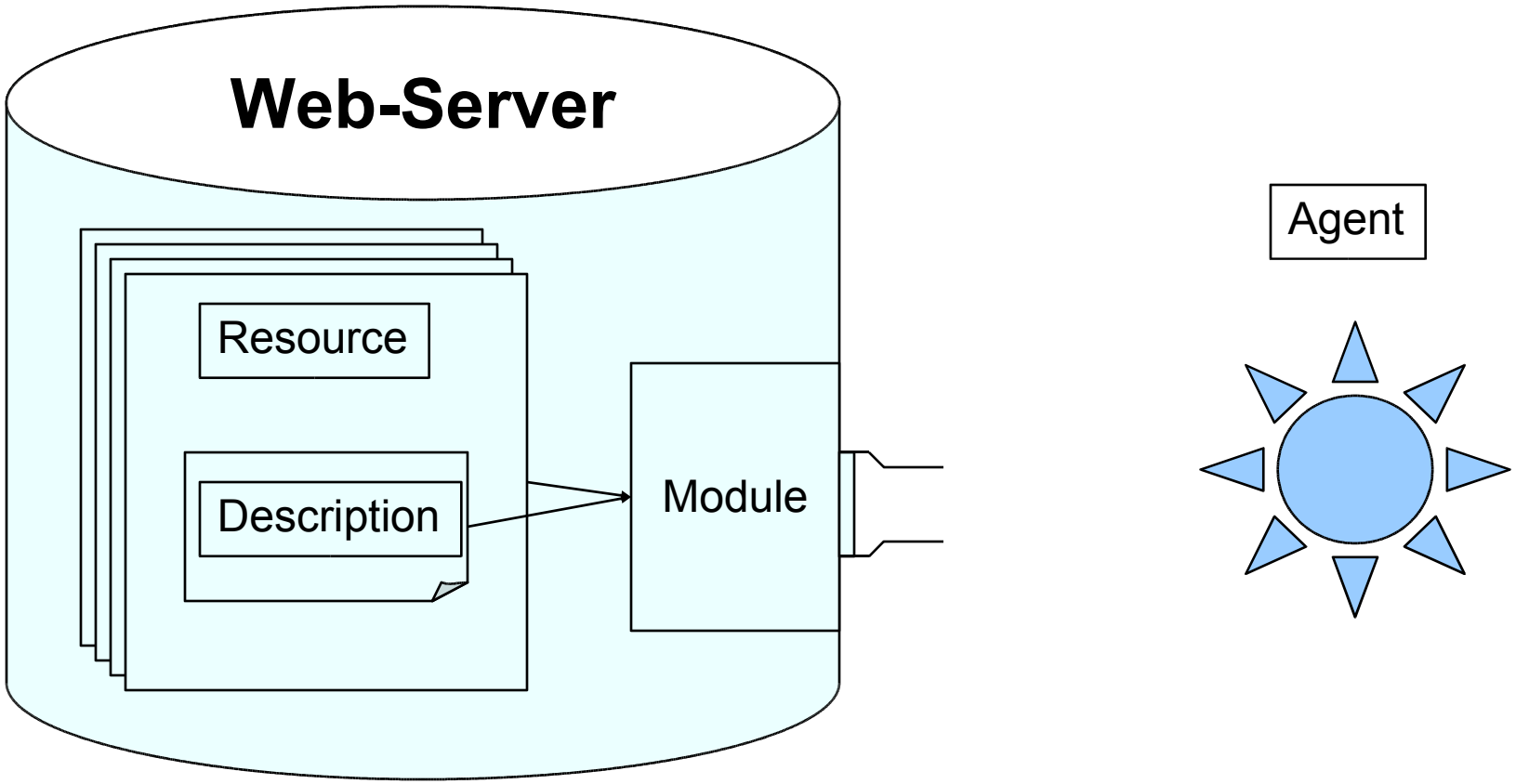
- Repositories contain Metadata, describing objects.
- User wants to retrieve the Metadata, without asking for the fulltext.

- Content vs. Description

- Full text search vs. Metadata retrieval

=> 

Requested extension of Web-Server



HTTP

- “HTTP (for HyperText Transfer Protocol) is the primary method used to convey information on the World Wide Web. The original purpose was to provide a way to publish and receive HTML pages.” www.wikipedia.org

HTTP methods

- **GET:** request for a resource
- **POST:** request for a resource (with additional data)
- **PUT:** stores entity under the supplied Request-URI
- **DELETE:** requests that the origin server deletes the resource identified by the request-URI.
- **HEAD:** same as GET, without expecting a resource.
- **OPTIONS:** retrieves information about communication options.
- Other **TRACE**, **CONNECT**.

HTTP methods

- **GET, POST, PUT, DELETE, HEAD, OPTIONS, TRACE, CONNECT**
- Do not distinguish between resource and its description
- Extension of the protocol is needed

HTTP protocol extension

- Proposal
  - Extension of http protocol

MGET	Return a description of the resource denoted by the request URI
MPUT	Add the statements to the description of a resource denoted by the request URI
MDELETE	Remove the statements about the resource denoted by the request URI

MPUT and MDELETE usually only accessible for the web server administrator.

Current development

- NOKIA offers a commercial version for URIQA.
  - Interesting especially for mobile services
- Cashmere-int develops an open source version of URIQA, called RDS (Resource Description Server). Available as perl module.
- Cashmere-int wants to bring MGET MPUT MDELETE into HTTP standard.

CASHMERE-int: Goals

Ziel des Vorhabens ist die aktive Beteiligung an ausgewählten Standardentwicklungen im Kontext Semantic Web und die kompetente Transmission in den Hochschulbereich der Bundesrepublik.

Damit wird eine schnellere und vertiefte Umsetzung von Standards in neue Dienste gefördert.

Ziel der Aktivitäten zur Transmission ist auch die Motivation weiterer Wissenschaftler ... zur aktiven Teilnahme an internationaler Standardisierung ... fortgeschrittener Webtechnologien.

R. Schwänzl, Projektantrag

CASHMERE-int: Focus

- Semantic Web
  - Methoden und Werkzeuge zur maschinellen und intellektuellen Aufbereitung von Information und deren Weiterverarbeitung in search and retrieval.
  - Beschränkung auf ein relativ schmales Themengebiet, exemplarisch für die gesamte Breite offener Standardisierung.
- Preservation Metadata
  - Als eine vertikale Implementation der angestrebten Integration, beispielhaft an diesem Teilgebiet des Semantic Web

CASHMERE-int: Method

- Durch die Entwicklung und Erforschung neuer Dienste und Verfahren fokussiert Cashmere-int auf die
- Beteiligung an internationalen Standardisierungs-ansätzen, die auf innovativen Methoden und Werkzeugen
- zur maschinellen und intellektuellen Aufbereitung von Informationen und deren Weiterverbreitung in search and retrieval abheben.
- Dabei soll darauf hingewirkt werden, dass leistungsfähige Standards entstehen, also solche, die eine reiche Produktentwicklung entfalten können.

Standardization Activities

- Preservation Metadata
  - PREMIS
  - OAIS
  - METS
  - Format Registry
- Semantic Web
  - Qualified DC in RDF/XML
  - CARA Parser
  - XSLT and Application Profiles
  - RDF Data-Access / XQuery
  - Development of Tools (⇨ DC-Tools)
- WebServices
  - SOAP
  - URIQA
  - RSS

Member of W3C via IWI e.V.
Member of Dublin-Core Advisory Board

For further information and links to all the sources, just visit the CASHMERE-int homepage: www.iwi-iuk.org/cashmere