Web Architecture Part 3

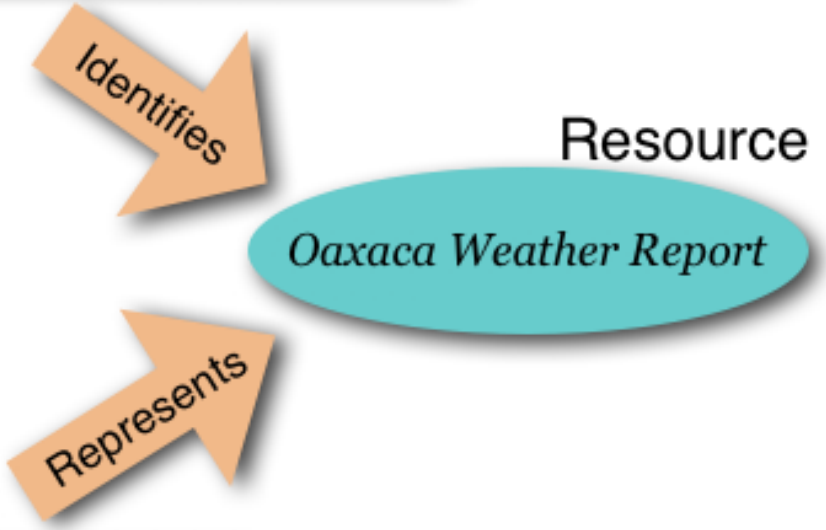
http://www.w3.org/TR/2004/REC-webarch-20041215/

Web Architecture – so far

- Collection of details of how technology works
  - At a rather shallow level
- Now:
  - Underlying principles
- Objectives:
  - Reflection
  - Re-assessment
  - Connecting the dots

URI

`http://weather.example.com/oaxaca`



Representation

Metadata:
`Content-type: application/xhtml+xml`

Data:

```
<!DOCTYPE html PUBLIC "...
    "http://www.w3.org/...
<html xmlns="http://www...
<head>
<title>5 Day Forecaste for
Oaxaca</title>
...
</html>
```

IDENTIFICATION

Principle: Global Identifiers

Global naming leads to global network effects.

- the value of an identifier increases the more it is used consistently
  - Analogous interesting example:
    no (!) proliferation of ontologies

- Every object addressable
  [Engelbart 1990, http://www.dougengelbart.org/pubs/augment-132082.html#2a1]

Good practice: Identify with URIs

To benefit from and increase the value of the World Wide Web, agents should provide URIs as identifiers for resources.

- A resource should have an associated URI if another party might reasonably want to
  - create a hypertext link to it,
  - make or refute assertions about it,
  - retrieve or cache a representation of it,
  - include
    - all or
    - part of it by reference into another representation,
  - annotate it, or
  - perform other operations on it.

Much more than hypertext linking!

- Software developers should expect that sharing URIs across applications will be useful, even if that utility is not initially evident.

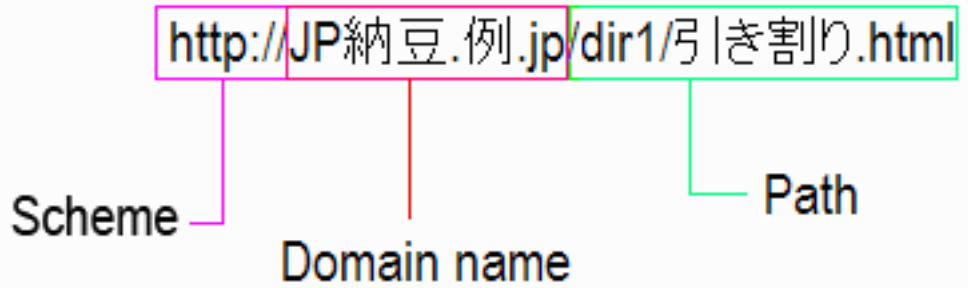
URI Syntax

- Full specification: http://www.ietf.org/rfc/rfc3986.txt
- Examples:
  - ftp://ftp.is.co.za/rfc/rfc1808.txt
  - http://www.ietf.org/rfc/rfc2396.txt
  - ldap://[2001:db8::7]/c=GB?objectClass?one
  - mailto:John.Doe@example.com
  - news:comp.infosystems.www.servers.unix
  - tel:+1-816-555-1212
  - telnet://192.0.2.16:80/
  - urn:oasis:names:specification:docbook:dtd:xml:4.1.2

URI, URL, URN, IRI

- Uniform Resource Identifier
  - Uniform Resource Locator
  - Uniform Resource Name

- IRI
  - http://ヒキワリ.ナットウ.ニホン



Cool URIs don't change

- What makes a cool URI?
- A cool URI is one which does not change.
- What sorts of URI change?
- URIs don't change: people change them.

Cool URIs don't change

- Idea: URIs are an abstract space
  - Different from file or database space

- Do‘s:
  - Describe your documents with stable properties
    /pubs/2012/DellschaftHypertextPaper.html
    will never have to change

- Don‘ts:
  - Describing your documents according to
    - the software you use now: /cgi-bin/paper?nr=1234
    - the owner of the document: /fb4/AGStaab/Publications
    - ownership, access, archive level security level, ...

The trouble with identifiers

- Experience: Knowledge acquisition for the knowledge acquisition community
  - V. Richard Benjamins, Dieter Fensel, Stefan Decker, Asunción Gómez-Pérez: (KA)2: building ontologies for the Internet: a mid-term report. Int. J. Hum.-Comput. Stud. 51(3): 687-712 (1999)
  - Steffen Staab, Jürgen Angele, Stefan Decker, Michael Erdmann, Andreas Hotho, Alexander Maedche, Hans-Peter Schnurr, Rudi Studer, York Sure: Semantic community Web portals. Computer Networks 33(1-6): 473-491 (2000)
  - Few hundred facts, yet
    - 3 identifiers for Dieter Fensel!

URI aliases

- Web architecture allows the association of more than one URI with a resource.
- URIs that identify the same resource are called URI aliases.

- Good practice: Avoiding URI aliases
  - A URI owner SHOULD NOT associate arbitrarily different URIs with the same resource.

- Good practice: Consistent URI usage
  - An agent that receives a URI SHOULD refer to the associated resource using the same URI, character-by-character.

- If aliases are used, then
  1. select one of them as
  2. Server-side redirect the other by the web server (e.g. 30x error code)

Aliasing vs Dereferencing URIs

- Dereferencing
  - http://www.koblenz.de/weather/current/
  - http://www.koblenz.de/weather/2012-11-02

- May give identical results, but there are no aliases, because on Nov 3 they will have different content

- The meaning of the two URIs is different!
  - Even on November 2, 2012

URI collision

- URI for the webpage of Steffen Staab
  - is not the same as
- URI for Steffen Staab

- They have different
  - Creation dates
  - Creation methods
  - Interfaces
  - ....

- KA2 Experience (and beyond): often done wrongly

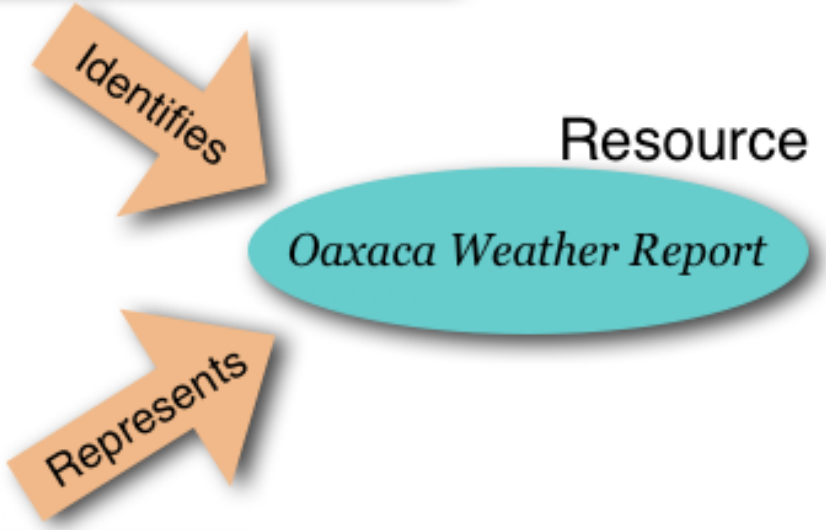
Semantics of identity

- "Web Ontology Language (OWL)" [OWL10] defines RDF properties
  - sameAs to assert that two URIs identify the same resource
  - inverseFunctionalProperty to imply it

INTERACTION/PROTOCOLS

URI

`http://weather.example.com/oaxaca`



Representation

Metadata:
`Content-type: application/xhtml+xml`

Data:

```
<!DOCTYPE html PUBLIC "...
    "http://www.w3.org/...
<html xmlns="http://www...
<head>
<title>5 Day Forecaste for
Oaxaca</title>
...
</html>
```

Accessing resources

- Using a URI to Access a Resource
  - URI typically prefixed by a protocol scheme, e.g.
    - http://west.uni-koblenz.de
    - With default protocol methods, e.g. get, post,...

- Access to representation depends on
  - Whether URI owner makes available any representations at all;
  - Whether the agent making the request has access privileges
  - If the URI owner has provided more than one representation (in different formats such as HTML, PNG, or RDF; in different languages; or transformed dynamically according to the hardware or software capabilities of the recipient), the resulting representation may depend on negotiation between the user agent and server.
  - The time of the request; the world changes over time, so representations change

Representation types & Internet media types

- *A representation is data that encodes information about resource state*.

- Good practice: Reuse representation formats
  - New protocols created for the Web SHOULD transmit representations as octet streams typed by Internet media types.

Separation of Concerns / Orthogonality

```
<body>
...
<a href="http://www.example.com/images/
nadia#hat">Nadia's hat</a>
...
</body>
```

1. Body understood by HTML processor (does not know about SVG)
2. Link extracted, clicked by user
3. Image served as image/svg+xml
4. SVG processor renders image (does not know about HTML)
5. SVG processor interprets „#hat“ appropriately

Content Negotiation

- User agent (browser) informs the server what media types it understands with ratings of how well it understands them

Accept HTTP header

```
Accept-Language: de; q=1.0, en; q=0.5
Accept: text/html; q=1.0, text/*; q=0.8, image/gif;
q=0.6, image/jpeg; q=0.6, image/*; q=0.5, */*; q=0.1
```

Availability and consistency of URI dereferencing

In practice, URIs are bookmarked, forwarded, stored; dereferencing them should lead to predictable performance

- Good practice: Available representation
  - A URI owner SHOULD provide representations of the resource it identifies
- Good practice: Consistent representation
  - A URI owner SHOULD provide representations of the identified resource consistently and predictably.

Linking and access control

- Don‘t assume that a page is unaccessible because it has not been told about to anyone
  - E.g. browsers store where you last have been
  - Server logs and search engines extract such links
    - Search engines thus find parts of the hidden web

- Rather use
  - Access control
  - Login + password

FORMATS

When to use an XML-based format

Design constraints that suggest the use of XML include:

- Requirement for a hierarchical structure.
- Need for a wide range of tools on a variety of platforms.
- Need for data that can outlive current applications
- Ability to support internationalization in a self-describing way that makes confusion over coding options unlikely.
- Early detection of encoding errors with no requirement to "work around" such errors.
- A high proportion of human-readable textual content.
- Potential composition of the data format with other XML-encoded formats.
- Desire for data easily parsed by both humans and machines.
- Desire for vocabularies that can be invented in a distributed manner and combined flexibly.

Binary data formats

ARCHITECTURAL PRINCIPLES

WeST

Orthogonal abstractions benefit from orthogonal specifications.

Application of the principle: Identification, interaction, and representation are orthogonal concepts, their technologies may evolve independently.

Examples:

- Resources are identified with URIs (no need for knowing about ist representation or availability).
- Generic URI syntax often does not require knowing specifics of URI schemes
- Representation of a resource may be changed without disrupting references to the resource (eg, by content negotiation).
- Even orthogonal specifications may change independently (e.g. when adopting SVG)

Steffen Staab
staab@uni-koblenz.de

Principle: Extensibility

- Nesting of languages
  - E.g. Web ontology language: OWL 2
  - Every legal OWL Lite ontology is a legal OWL DL ontology.
  - Every legal OWL DL ontology is a legal OWL Full ontology.
  - Every valid OWL Lite conclusion is a valid OWL DL conclusion.
  - Every valid OWL DL conclusion is a valid OWL Full conclusion.

- Try to avoid compatibility problem by ignoring unknown language constructs of the more expressive language
- Still be able to „do something“
  - E.g. show text only if you cannot display graphics

Principle: Error recovery

Agents that recover from error by making a choice without the user's consent are not acting on the user's behalf.

- Consent may be pre-configured

- Error correction
  - Repair the error to make it unhappen
  - E.g. request packet again if it fails to arrive in time

- Error recovery
  - Continue working by living with the error

Web is a software environment where the user is expected to be aware of some errors!

Protocol-based Interoperability

- interfaces are defined in terms of protocols,
  - syntax,
  - semantics, and
  - sequencing constraints
  - (time outs)
- of the messages interchanged.

- Contrast: APIs

- Protocol-based design
  - Agents are developed faster than the protocols
- Make protocols visible
  - Allow for understanding and engineering