AWWSW brainstorm on possible topics for group to take up

We request that contributions to this page be limited to those who agree to the terms of the W3C patent policy http://www.w3.org/Consortium/Patent-Policy-20040205/.

This page gives a list of *potential* topics for discussion by the AWWSW group. Everyone at the first telecon was wary of scope creep, so it was felt that a survey of potential tarpits should be conducted so that we are better able to maintain our guard. Our actual work will be restricted to some subset of these items.

This page was prepared in order to satisfy http://www.w3.org/2007/11/13-awwsw-minutes.html#action02 .

HTTP semantics

Clarify web architecture around what is (or should be) implied by HTTP responses.

- What can you infer from a 200 response?
- What can you infer from a 303 response?

It is desirable to capture the answers to these questions formally, and in particular as RDF statements. To this end we will need an ontology (possible starting point: http://www.w3.org/2006/gen/ont) for expressing such statements.

Clarify whether this clarification activity is descriptive (intended to interpret HTTP 2616 but not build on it) or prescriptive (intended to specify additional constraints on what servers "should" do, a la AWWW).

Resources and representations

Clarify the relationship between a resource and its representations (HTTP responses).

For a resource, what is a satisfactory representation? Can it be anything? If one representation is a photo, another perhaps shouldn't be a cartoon, but a lossy photo might be acceptable.

Are a resource's representations sufficient to figure out what the resource is - do they define the resource? What do 200 responses tell you about the resource, if anything?

Is http://news.google.com/ an information resource? If so, then its representations are representations of what?

Is there a difference between an information resource and its essence? Between its essence and its representations? What is the ontological type of an essence, what is its identity, and what are the operations on it?

How does Content-location: relate to representations?

Giving teeth to "web architecture"

How can you write a program (validator) to determine whether a web site is not following "web architecture"?

Just using HTTP

Clarify the argument that you can do everything with HTTP.

There's a tag issue and finding saying "just use HTTP". So in scope for this group would be explaining and embellishing how to use HTTP. This may help in the struggle to explain why LSID and other [schemes] are unnecessary.

Metadata

There's a need to be able to obtain metadata about a data source (similar to "getMetadata" in the LSID protocol). Maybe write this up and liaise with the HTTP WG.

E.g. How do you know how many representations there are (or will be) for a resource? Should there be a way?

Other issues

The specification of 303 See Other is not necessarily precise enough for the semantic web use case: It would be nice if we could at least expect RDF, and maybe specific kinds of information.

Location independence: What happens when a resource moves and the community wants to do something about it (issue a "third-party redirect")?

Another issue is what to say about time and RDF. This keeps coming up. (Timbl: "architecture doesn't have time; new model of time is out of scope; but HTTP has its own notion of time.")

Comment [skw3]: By the TAG httpRange-14 http://news.google.com *is* an information resource because it responds with a 200 OK (though that took a while to determine because it responds 403 Forbidden to wget).

As regards what are the respesentations of then... they would appear to be representations of the front page of Google News - an online news publication. In a Wittgensteinian sense our repeated experiences of that page would work to confirm that conclusion. The resource is useful in part because of the consistency ... [3]

Comment [skw4]: The question is probably rethorical. Personnally, I have come to prefer a formulation based either on an actual lack of representations (pragmatic) or that a resource is incapable of webarch:representation ... [4]

Comment [skw5]: I think that you are now trying to make too much capital our of a linguist expression. Please suggest how you might frame what you would understand an information resource to be - there seem to me to be things that clearly are ... [5]

Comment [skw6]: It doesn't - it relates to resources. The way I look at it is that it establishes the resource referenced by the content-location as a variant of the resource referenced by the corresponding request. Variation may be by media-type, natu ... [6]

Comment [skw7]: I doubt it. One experiences resources through their representations. In as much as machine processable claims are made about resources and there are sufficient axioms/ontologies available - some claims may lead to contradictions that could ... [7]

Comment [skw8]: That's UrnsAndRegistries-50: henry is in the process of rewriting that coming at the question from a different angle.

Comment [skw9]: Of an instant? - possibly! Over time? - No!

Amongst the headers in the response to " wget -d --header=Accept: http://www.w3.org/Icons/w ... [8]

Comment [skw10]: You *cannot* weld such a media-type dependency into HTTP itself. The meaning of response codes really should be orthonal to media types and vice versa.

What's the web analog for doing citation? (I.e. how to cite articles in published literature in such a way that we can tell when two RDF documents are citing the same article. Problems: common names, stability, third-party metadata.)

Possible work products

Any output of this group is intended to be fed back to the TAG or other groups in order to inform or guide further action.

- HTTP semantics ontology
- List of problems that need to be solved, missing functionality, possible implementations
- List of things needing better exposition
- Set of best practices, to be folded into web architecture as a TAG finding
- FAQ on web architecture and/or HTTP semantics

I think that as far as is possible one should try to either infer "as little as possible" or "only what is absolutely necessary" from http response codes in general. I think that there is probably a useful formalism to be explored in terms of better explaining the HTTP specs... and potentially uncovering 'quirks' therein. eg. Location headers associated with 301 and 302 response provide alternate URI for the original resource (moved permanently and temporarily); whereas 303s do not establish the redirection target as an alternate in the same sense - which suggests that 300 should have weaker seeAlso like semantics that the stronger replacement URI semantics.

Do do a comprehensive - general purpose job for http across a full set of response codes, request and response headers and media-types is... well, huge - and concensus with the IETF on http interaction semantics, which would be important to such a venture, would also take considerable work. I'd prefer, I think, to focus on formalising a limited set of patterns that make, as far as is possible, legitimate use of existing facilities. I think it then possible for some community - say the LinkDataCommunity or through AWWSW and/or TAG establish a community practice and a set of community accepted semantics/inferences associated with the use of those patterns. In the longer run, some inferences may be of more general utility and with careful documentation and presentation, one might encourage the IETF community to adopt them as well, and possibly get engaged in filling out more of the general inferences that could be made.

I used to be of the opinion that say a JPEG encode image of a person was an acceptable webarch:Rrepresentation of that person. However, I am no longer of that view. I have come to regard, say a person, as something that defies representation in the sense of having a webarch:Representation - I am not information, "...I am a bag of mostly water." So, I'd regard the JPEG encoded bitstream of conveying an image of me as a webarch:Representation of a resource which depicts me. Likewise for most if not all physical objects. The grey area form me come with say RDF properties, conceptual things - certainly abstract conceptual things (see also Pat's discussion of Unicorns with John Cowan some way back), and namespaces.

So I see images of things as being separate from the things they depict and in many (almost said most - but realised that would leave a target :-)) cases the depicted thing from my pov defies webarch:representation - in which case, simply don't deploy or claim to have deployed webarch:repesentations of such things - provide descriptions/depictions and indirect to those either with '#s' or 303's or possibly some other pragmatic that leads you to a decription/depiction of the thing on interest that you are prepared to trust.

I find Pat's infamous PatHayes page interesting on a couple of fronts. Firstly, at least in narrative form, it contains a number of invariants that by an large taken together distinguish the individual Pat Hayes from all others (people let alone Pat Hayes'). Secondly, that the objects of many of those invariant statements could infact only be established by similar treatment of the corresponding referring names (IIRC, because I haven't looked recently and I'm being lazy) the description at least makes reference to parents and birthplace.

[Aside: I'd really like something like owl:DistinguishingProperty such that for a individuals of given object class taken together all 'it's' required distinguishing properties (min cardinality >1 restrictions) establish a 'complex key' that discriminates the indiviudal from all others.]

By the TAG httpRange-14 http://news.google.com *is* an information resource because it responds with a 200 OK (though that took a while to determine because it responds 403 Forbidden to wget).

As regards what are the respesentations of then... they would appear to be representations of the front page of Google News - an online news publication. In a Wittgensteinian sense our repeated experiences of that page would work to confirm that conclusion. The resource is useful in part because of the consistency of the conclusion we reach through repeated visits. Our inutition builds and we are not surprised by subsequent visits to the resource. Of course, nothing tells us authoratively what the resource actually is. Certainly the browser has no idea. Google may tell us if we ask them - but we don't know how to do that.

A semantic web resource could include and element of self-description for machines, This page could have, but doesn't AFAICT, contain an element of self-description for humans.

The question is probably rethorical. Personnally, I have come to prefer a formulation based either on an actual lack of representations (pragmatic) or that a resource is incapable of webarch:representation (factual/philosopical). webarch:Representations are really of the *current* state of a resource and I think our (the TAGs) defn viz message conveyable essential characteristics fails to take in temporal change - though on could regards the set of available representations over all time organised say by media type and time (extending into the future and the past) as a structure that could be convey in part by a message and incrementally grown in further messages - but that is to wriggle on the point.

I am also intrigued by the PatHayes declaration page - because it actually conveys as set of invariant (and possibly essential) characteristics of Pat - I don't think I've heard Pat pose the question of whether he is an information resource by claiming that "all his essential characteristics can be conveyed in a message." but he seems pretty close to having established that such is in fact possible :-)

I think that you are now trying to make too much capital our of a linguist expression. Please suggest how you might frame what you would understand an information resource to be - there seem to me to be things that clearly are; things that clearly are not; and some that at least at present are grey. Cover the first two and see where the others land.

It doesn't - it relates to resources. The way I look at it is that it establishes the resource referenced by the content-location as a variant of the resource referenced by the corresponding request. Variation may be by media-type, natural language... (can't think of other dimensions right now but there are probably a few more common ones).

I doubt it. One experiences resources through their representations. In as much as machine processable claims are made about resources and there are sufficient axioms/ontologies available - some claims may lead to contradictions that could be caught - eg. a 200 response with a representation that also claims the URI identifies/denotes a person with axioms that state people and information resources are distinct.
At a human level, it is possible to spot inconsistencies between ones experience of a resource and what is claimed (sometimes authoratively) about a resource eg. http://www.markbaker.ca/ (Mark, possbily for the purposes of argument, has claimed that the URI identifies/denotes himself, the person, Mark Baker rather than (one of) his homepage(s).

Of an instant? - possibly! Over time? - No!

Amongst the headers in the response to " wget -d --header=Accept: http://www.w3.org/Icons/w3c_home" is:
        Alternates: {"w3c_home.png" 0.7 {type image/png} {length 1936}}, {"w3c_home.gif" 0.5 {type image/gif} {length 1865}}

Would need check the specs to see if this is a generic way to find out and whether the Alternatives: header is supposed to be a complete account for a given instant.