

## Tortured Terminology and Problematic Prescriptions

by Mike Bergman - Tuesday, March 27, 2012

<http://www.mkbergman.com/1002/tortured-terminology-and-problematic-prescriptions/>



### Casting My Vote on Revising `httpRange-14`

The [httpRange-14](#) issue and its predecessor "identity crisis" debate have been active for more than a decade on the Web [1]. It has been around so long that most acknowledge "fatigue" and it has acquired that rarified status as a [permathread](#). Many want to throw up their hands when they hear of it again and some feel -- because of its duration and lack of resolution -- that there never will be closure on the question. Yet everyone continues to argue and then everyone wonders why actual consumption of linked data remains so problematic.

Jonathan Rees is to be thanked for refusing to let this sleeping dog lie. This issue is not going to go away so long as its basis and existing prescriptions are, in essence, incoherent. As a member of the W3C's TAG ([Technical Architecture Group](#)), Rees has worked diligently to re-surface and re-frame the discussion. While I don't agree with some of the specifics and especially with the constrained approach proposed for resolving this question [2], the sleeping dog has indeed been poked and is awake. For that we can thank Jonathan. Maybe now we can get it right and move on.

I don't agree with how this issue has been re-framed and I don't agree that responses to it must be constrained to the prescriptive approach specified in the TAG's [call for comments](#). Yet, that being said, as someone who has been vocal for years about the poor semantics of the semantic Web community, I feel I have an obligation to comment on this official call.

Thus, I am casting my vote behind David Booth's alternative proposal [3], with one major caveat. I first explain the caveat and then my reasons for supporting Booth's proposal. I have chosen not to submit a separate alternative in order to not add further to the noise, as Bernard Vatant (and, I'm sure, many, many others) has chosen [4].

### Bury the Notion of 'Information Resource' Once and for All

I first commented on the absurdity of the 'information resource' terminology about five years ago [5]. Going back to Claude Shannon [6] we have come to understand information as entropy (or, more precisely, as differences in energy state). One need not get that theoretical to see that this terminology is confusing. "Information resource" is a term that defies understanding (meaning) or precision. It is also a distinction that leads to a natural counter-distinction, the "non-information resource", which is also an imprecise absurdity.

What the confusing term is meant to encompass is *web-accessible content* ("documents"), as opposed to *descriptions* of (or statements about) things. This distinction then triggers a different understanding of a

URI (locator *v* identifier alone) and different treatments of how to process and interpret that URI. But the term is so vague and easily misinterpreted that all of the guidance behind the machinery to be followed gets muddled, too. Even in the current chapter of the debate, key interlocutors confuse and disagree as to whether a book is an "information resource" or not. If we can't basically separate the black balls from the white balls, how are we to know what to do with them?

If there must be a distinction, it should be based on the idea of the actual *content* of a thing -- or perhaps more precisely *web-accessible content* or *web-retrievable content* -- as opposed to the *description* of a thing. If there is a need to name this class of content things (a position that David Booth prefers, *pers. comm.*), then let's use one of these more relevant terms and drop "information resource" (and its associated IR and NIR acronyms) entirely.

The motivation behind the "information resource" terminology also appears to be a desire that somehow a URI alone can convey the name of what a thing is or what it means. I recently tried to blow this notion to smithereens by using Peirce's discussion of signs [1]. We should understand that naming and meaning may only be provided by the owner of a URI through additional explication, and then through what is understood by the recipient; the string of the URI itself conveys very little (or no) meaning in any semantic sense.

We should ban the notion of "information resource" forever. If the first exposure a potential new publisher or consumer of linked data encounters is "information resource", we have immediately lost the game. Unresolvable abstractions lead to incomprehension and confusion.

The approach taken by the TAG in requesting new comments on httpRange-14 only compounds this problem. First, the guidance is to not allow any questioning of the "information resource" terminology within the prescribed comment framework [7]. Then, in the suggested framework for response, still further terminology such as "probe URIs", "URI documentation carrier" or "nominal URI documentation carrier for a URI" is introduced. Aaaaarggghh! This only furthers the labored and artificial terminology common to this particular standards effort.

While Booth's proposal does not call for an outright rejection of the "information resource" terminology (my one major qualification in supporting it), I like it because it purposefully sidesteps the question of the need to define "information resource" (see his Section 2.7). Booth's proposal is also explicit in its rejection of implied meaning in URIs and through embrace of the idea of a protocol. Remember, all that is being put forward in any of these proposals is a mechanism for distinguishing between retrievable content obtainable at a given URL and a description of something found at a URI. By racheting down the implied intent, Booth's proposal is more consistent with the purpose of the guidance and is not guilty of overreach.

## **Keep It Simple**

One of the real strengths of Booth's proposal is its rejection of the prescriptive method proposed by the TAG for suggesting an alternative to httpRange-14 [7]. The parsimonious objective should be to be simple, be clear, and be somewhat relaxed in terms of mechanisms and prescriptions. I believe use patterns -- negotiated via adoption between publishers and consumers -- will tell us over time what the "right" solutions may be.

Amongst the proposals put forward so far, David Booth's is the most "neutral" with respect to imposed meanings or mechanisms, and is the simplest. Though I quibble in some respects, I offer qualified support for his alternative because it:

- Sidesteps the "information resource" definition (though weaker than I would want; see above)
- Addresses only the specific HTTP and HTTPS cases
- Avoids the constrained response format suggested by the TAG
- Explicitly rejects assigning innate meanings to URIs
- Poses the solution as a *protocol* (an understanding between publisher and consumer) rather than defining or establishing a *meaning* via naming
- Provides multiple "cow paths" by which resource definitions can be conveyed, which gives publishers and consumers choice and offers the best chance for more well-trodden paths to emerge
- Does not call for an outright repeal of the httpRange-14 rule, but retains it as one of multiple options for URI owners to describe resources
- Permits the use of an HTTP 200 response with RDF content as a means of conveying a URI definition
- Retains the use of the hash URI as an option
- Provides alternatives to those who can not easily (or at all) use the 303 see also redirect mechanism, and
- Simplifies the language and the presentation.

I would wholeheartedly support this approach were two things to be added: 1) the complete abandonment of all "information resource" terminology; and 2) an official demotion of the httpRange-14 rule (replacing it with a slash 303 option on equal footing to other options), including a disavowal of the "information resource" terminology. I suspect if the TAG adopts this option, that subsequent scrutiny and input might address these issues and improve its clarity even further.

There are other alternatives submitted, prominently the one by Jeni Tennison with many co-signatories [\[8\]](#). This one, too, embraces multiple options and cow paths. However, it has the disadvantage of embedding itself into the same flawed terminology and structure as offered by httpRange-14.

---

[1] For my recent discussion about the history of these issues, see M.K. Bergman, 2012. "Give Me a Sign: What Do Things Mean on the Semantic Web?," in *AI3:::Adaptive Information* blog, January 24, 2012; see <http://www.mkbergman.com/994/give-me-a-sign-what-do-things-mean-on-the-semantic-web/>.

[2] In all fairness, this call was the result of [ISSUE-57](#), which had its own constraints. Not knowing all of the background that led to the httpRange-14 Pandora's Box being opened again, the benefit of the doubt would be that the form and approach prescribed by the TAG dictated the current approach. In any event, now that the Box is open, all pertinent issues should be addressed and the form of the final resolution should also not be constrained from what makes best sense and is most pragmatic.

[3] [David Booth](#)'s alternative proposal is for the "[URI Definition and Discovery Protocol](#)" (uddp). The actual submission according to form is found [here](#).

[4] See Bernard Vantant, 2012. "[Beyond httpRange-14 Addiction](#)," *the wheel and the hub* blog, March 27, 2012. See <http://blog.hubjects.com/2012/03/beyond-http-range-14-addiction.html>.

[5] M.K. Bergman, 2007. "More Structure, More Terminology and (hopefully) More Clarity," in *AI3:::Adaptive Information* blog, July 27, 2007; see <http://www.mkbergman.com/391/more-structure-more-terminology-and-hopefully-more-clarity/>. Subsequent to that piece, I have written further on semantic Web semantics in "[The Semantic Web and Industry Standards](#)" (January 26, 2008), "[The Shaky Semantics of the Semantic Web](#)" (March 12, 2008), "[Semantic Web Semantics: Arcane, but Important](#)," (April 8, 2008), "[The Semantics of Context](#)," (May 6, 2008), "[When Linked Data Rules Fail](#)" (November 16, 2009), "[The Semantic 'Gap'](#)" (October 24, 2010) and [1].

[6] Claude E. Shannon, 1948. "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, 1948.

[7] In the "[Call for proposals to amend the "httpRange-14 resolution"](#)" (February 29, 2012), Jonathan Rees (presumably on behalf of the TAG), stated this as one of the rules of engagement: "9. Kindly avoid arguing in the change proposals over the terminology that is used in the baseline document. Please use the terminology that it uses. If necessary discuss terminology questions on the list as document issues independent of the 303 question." The specific [template form](#) for alternative proposals was also prescribed. In response to interactions on this question on the mailing list, Jonathan stated:

If it were up to me I'd purge "information resource" from the document, since I don't want to argue about what it means, and strengthen the (a) clause to be about content or instantiation or something. But the document had to reflect the status quo, not things as I would have liked them to be.

I have not submitted this as a change proposal because it doesn't address ISSUE-57, but it is impossible to address ISSUE-57 with a 200-related change unless this issue is addressed, as you say, head on. This is what I've written in my TAG F2F preparation materials.

[8] Jeni Tennison, 2012. "[httpRange-14 Change Proposal](#)," submitted March 25, 2012. See the [mailing list notice](#) and [actual proposal](#).

