

URIs and the Myth of Resource Identity

David Booth

HP Software

This version: 18-March-2006

Please read the latest version: <http://dbooth.org/2006/identity/>

The views expressed herein are those of the author, and not necessarily those of HP. These views are evolving, and comments are invited.

Abstract

The question is not "What does this URI mean?". The question is "What can you *do* with it?". The problem of resource identity is the problem of locating appropriate descriptive information about the resource -- information that enables you to make use of that URI in a particular application. However, the value of a URI is increased when descriptive information uniquely identifies the URI's resource, even if the resource is not completely described, because it encourages the network effect.

Table of Contents

- [Introduction](#)
- [The Problem of Resource Identity](#)
- [Descriptive Information](#)
- [Descriptive Information Is Inherently Partial](#)
- [OO Analogy](#)
- [Identifying Versus Describing: Distinguishing One Resource from All Others](#)
- [Resource Identity and the Network Effect](#)
- [Should a URI Identify Both a Person and an Information Resource?](#)
- [Serving Documents from a URI](#)
- [Conclusion](#)
- [References](#)

Introduction

The question is not "What does this URI mean?". It has no intrinsic meaning. The question is "What can you *do* with it?".

Well, what do you *want* to do with it?

- "I want to retrieve Web documents from it, and comment on their content."
<http://documents.example.com/foo.html> (HTTP GET, 200 response)
- "I want to use it to find Molly's friend Jack's phone number."
<http://foaf.example.com/mollys-contacts.rdf> (HTTP GET, 200 response; FOAF ontology)
- "I want to use it to retrieve my mobile phone status, model, configuration and settings, from a browser."
<http://acme-mobile.example.com/phones/id9988765/> (HTTP GET, 200 response)
- "I want to use it in a mobile phone service provider application, to unambiguously identify a particular phone, so that I know who to bill for the calls that are made from it."
<http://acme-mobile.example.com/phones/id9988765/>
- "I want to use it to unambiguously identify the person who was videotaped robbing the Willow Street Bank at 10:08AM on March 15, 2005. We do not yet know the person's name, but we know other things about the person."
<http://crimes.example.org/perpetrators/2005/willow-street-bank-robbery/>
- "I want to use it as a Web address for retrieving a description of the person who was videotaped robbing the Willow Street Bank at 10:08AM on March 15, 2005."
<http://crimes.example.org/perpetrators/2005/willow-street-bank-robbery/>

The Problem of Resource Identity

The association between a URI and its resource is contrived and imaginary. It exists only because we follow certain conventions and we believe it exists. We do so because it is a useful fiction: it allows us to symbolically manipulate that resource by **proxy**. We use the URI to manipulate a model of that resource instead of manipulating the actual resource.

Descriptive Information

Some URIs are merely Web addresses, and are used only to identify "[information resources](#)". Other URIs are used to identify things such as concepts in an ontology, or mobile phones, or people. In general, a URI identifies a "[resource](#)", but *what* resource? The problem of resource identity is the problem of understanding what a particular URI "means". Specifically, it is the problem of understanding what *resource* that URI identifies.

Suppose I [own](#) a URI that identifies a concept in an ontology.[1] How can I tell someone else what resource that URI identifies? I can provide **descriptive information** about it. In RDF, I can provide a set of assertions involving that URI; or in HTML I might provide an English description of the resource.

For example, the URI <http://t-d-b.org?http://dbooth.org/2005/dbooth/> identifies me, David Booth, the author of this paper. You can verify that by deferencing it to obtain **authoritative descriptive information** that I have provided (after following the 303-redirect). This descriptive information is expressed as a **set of assertions** about the URI. It tells you that I work for HP Software, I have email address dbooth@hp.com (as of 1-Jan-2005), etc. In this case the assertions are expressed in English in an HTML document, though I could also have expressed them in RDF or some other language. In conjunction with your knowledge of English and some shared background knowledge, this may be enough information to enable you to use my URI in some applications as an unambiguous identifier for me. In essence, these assertions that I have provided give you a **model** that I have endorsed, to enable you to make use of my URI in applications, whether this model is expressed formally in RDF or informal in English. Your application manipulates this model as a proxy for me.

Descriptive Information Is Inherently Partial

If the URI is used to identify a real life entity or other **non-idealized entity**, such as an actual person, then it would be impossible to describe that resource completely: it is not possible to describe everything about a person. (In contrast, an idealized entity would be something that can be completely characterized in a finite description, such as a mathematical formula.)

For example, it is not possible to completely describe David Booth, the person, because such a description would have to include all possible information about me, and that isn't feasible to collect or anticipate. For example, if the descriptive information includes my name, email address, phone number and physical address, that might be enough for one application, but it won't be enough for other applications that need my blood type, my mass, my educational history, or my philosophical viewpoints.

Thus, **descriptive information is inherently partial**. [1] What matters is that you are able to find *sufficient* descriptive information about me for the purpose of your application.

Fortunately, a resource does not need to be completely described in order to productively use its URI in a model as a proxy for the resource. It only needs to be described enough for the purpose of the application or task at hand. Thus, the problem of trying to completely describe or understand the resource is not necessarily relevant.

However, if I give you descriptive information that you cannot relate to anything else you already know, then you will not be able to make use of the URI, even if that descriptive information may be perfectly adequate for some other application. For example, if you receive a set of assertions expressed using the FOAF ontology[2], but you are unable to relate that ontology to the ontology you are using, the URI will be meaningless: you will not be able to make use of it as a proxy for its associated resource. **Thus, the problem of resource identity is the problem of locating appropriate descriptive information.** The identity of a URI is not always irrelevant except as a convenient shorthand for the set of ways that URI can be used.

Does the descriptive information describe the resource, or does it really just describe the use of the URI? The descriptive information is expressed in terms of the URI, since it has no way of directly referring to the resource. For example, it may provide a set of assertions involving that URI. You can think of those assertions as indirectly conveying information about the associated resource, but in some sense that is also a useful fiction: the assertions are really just telling you how you may use the URI in manipulating a model that acts as a proxy for the resource, i.e., they tell you can *do* with that URI.

OO Analogy

In essence, when I (authoritatively) give you descriptive information about me, I am endorsing the use of my URI within a particular model (or application context). That model is a crude symbolic representation of certain aspects of the real me.

Another way to put it is that the authoritative descriptive information that I publish licenses the use of my URI in certain models. This is analogous to publishing some interfaces for an object in an OO system. You can never be sure that I won't (monotonically) publish an additional interface at some point in the future, just as you cannot be sure I won't publish more descriptive information about me.

Identifying Versus Describing: Distinguishing One Resource from All Others

Even if it is not possible to completely *describe* a resource, it may be possible to unambiguously *identify* that resource, in the sense of conveying what that resource is, as distinct from all other possible resources.

For example, if I provide descriptive information telling you that the URI <http://t-d-b.org?http://dbooth.org/2005/dbooth/> identifies all and only the actual, living person with email address dbooth@hp.com as of 1-Jan-2005, that is sufficient to unambiguously identify me, distinct from all other possible resources. There is a finite number of living people and only one with that email address on that date. Assuming we have a shared understanding of what an "actual, living person" means, this information allows you to understand exactly which resource I meant in the universe of all possible resources.

Uniquely identifying a resource is different from merely naming that resource. The URI <http://t-d-b.org?http://dbooth.org/2005/dbooth/>

acts as a recognizable name for me, but by itself it gives you no information about what resource it names. You need descriptive information to know what resource it identifies. Caution: I may slip sometimes and use the word "identify" when I meant "name".

Resource Identity and the Network Effect

The ability to uniquely identify a resource -- in the sense of conveying the distinction between this resource and all other resources -- is important because it enables others to publish additional descriptive information about the resource, beyond what the URI owner provides.

If a URI's resource is not uniquely identified, then others who wish to make statements about it run the risk that they may have guessed wrong about what resource the URI owner was intending to identify. For example, if I only said that <http://t-d-b.org?http://dbooth.org/2005/dbooth/> identifies an individual who works at HP with the name "David Booth", then it could be identifying any of four "David Booth"s who currently work for HP. And if I didn't say that it identifies an actual, living person then it could have been identifying a fictional person who bears my name. And if I didn't say that it identifies "all and only" me, then in theory it might be identifying only a part of me, or the union of me and something else. (Of course, in colloquial language people are not usually so legalistic as to say that they are identifying "all and only" a particular person. So in practice it is generally assumed that if I publish information saying that my URI identifies "the person who works at HP with email address dbooth@hp.com as of 1-Jan-2005", I am referring to an actual person and all and only that person. It would be misleading to mean otherwise.)

The Semantic Web is all about the [network effect](#) created by the use of URIs as universal identifiers. When a URI's resource is uniquely identified, it enables "anyone to say anything" about that resource[4]. If a URI's resource is not uniquely identified -- if users must rely solely on authoritative descriptive information about that resource -- it hampers others' ability to make statements about that resource, thus diminishing the network effect.

Should a URI Identify Both a Person and an Information Resource?

Since **anything can be a resource**, and http URIs can be used for any purpose [6], I could mint a URI, *u1*, that identifies **the union of me and my Web page**. Clearly this URI may not be useful for your particular application, which may need to discriminate more finely between me and my Web page. But it may meet my needs perfectly well, and it is, after all, *my* URI. Would this be anti-Web or anti-social? It depends.

First of all, since this URI doesn't distinguish between me and my Web page, in order to be a good Web citizen I should *also* create a second URI, *u2*, to separately identify me and a third URI, *u3*, to separately identify my Web page, because every important resource should have its own URI[3], and presumably both I and my Web page are important resources. :) (In other words, someone else may want to refer to them individually.) So the resource of *u1* would be the union of the resource of *u2* and the resource of *u3*.

Second, even if I do supply these additional URIs to permit you to talk separately about me and my Web page, this alone may not be enough to make me a good Web citizen. It may also depend on what is returned when *u1* is dereferenced using an HTTP GET request.

Serving Documents from a URI

If an HTTP GET on URI *u1* results in a 2xx response code, that response code represents prima facie evidence that the URI's resource can act as an "[information resource](#)". (Duh! That's what it just did!) Does that mean it must *only* act as an information resource and should not also act as some other kind of resource? I don't know.

One view would be to say that it's fine if a resource that acts as an "information resource" can also act as some other kind of resource, such as a person. Just because you know a URI can be used in one way, that doesn't mean it cannot also be used in some other way. This is somewhat analogous to an object in object oriented programming, in which an object implements multiple interfaces. What matters is not what that object (or resource) *is*, what matters is what you can *do* with it, i.e., the interfaces through which you can interact with it (or the ways you can make use of it). Each interface is analogous to the use of the URI in one application or "processing context". A single interface may not completely characterize the potential uses of the object, just as a single set of descriptive information may not completely characterize the potential uses of that URI. This is like making an open world assumption about the identity of the resource.

In some ways this view seems reasonable, but it also seems peculiar, because in the vast majority of cases (for normal Web pages), the URI for accessing a page is synonymous with the URI that unambiguously names that page: if you know how to access a page, you know how to name it. But in this case, even though the URI, *u1*, of the union could be used to access the Web page of *u3*, *u1* is a different URI than *u3*.

Conclusion

Although the identity of a URI's resource is a useful fiction, what really matters is what you can *do* with the URI. It is more useful if its resource is unambiguously identified, even if it is not completely described. Is it acceptable to have a URI that identifies the union of a person and a Web page? I don't know.

References

1. Pat Hayes, on Semantic Web Best Practices mailing list, 26 Jan 2006: <http://lists.w3.org/Archives/Public/public-swbp-wg/2006Jan/0153>
2. Friend Of A Friend (FOAF) ontology project: <http://www.foaf-project.org/>
3. "Architecture of the World Wide Web Volume 1", W3C Technical Architecture Group, <http://www.w3.org/TR/webarch/#uri-benefits>
4. "Resource Description Framework (RDF): Concepts and Abstract Data Model" W3C Working Draft 29 August 2002: <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/>
5. "Four Uses of a URL: Name, Concept, Web Location and Document Instance", David Booth, http://www.w3.org/2002/11/dbooth-names/dbooth-names_clean.htm
6. Roy Fielding, on TAG mailing list, 18 June 2005: <http://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html>