Network Working Group	L. Masinter
Internet-Draft	Adobe
Intended status: Experimental	January 12, 2012
Expires: July 15, 2012	

TOC

The 'tdb' and 'duri' URI schemes, based on dated URIs
draft-masinter-dated-uri-09

Abstract

This document defines two URI schemes. The first, 'duri' (standing for "dated URI"), identifies a resource as of a particular time. This allows explicit reference to the "time of retrieval", similar to the way in which bibliographic references containing URIs are often written.

The second scheme, 'tdb' (standing for "Thing Described By"), provides a way of minting URIs for anything that can be described, by the means of identifying a description as of a particular time. These schemes were posited as "thought experiments" and therefore these schemes are designated as Experimental.

Note

This document is not a product of any working group. Many of the ideas here have been discussed since 2001. Versions of this document have been discussed on the mailing list <uri@w3.org>. Previous versions have couched 'tdb' and 'tdb' as URN namespaces, used different syntax for timestamps, and handled URIs with fragment identifiers differently.

Table of Contents

---

1. Overview and Requirements

TOC

This document is not a product of any working group. Many of the ideas here have been discussed since 2001. The practical application of the URI schemes defined here is uncertain, but enough interest has been expressed in having a stable reference for these concepts that it seems worthwhile to publish these, if only as "experimental".

Versions of this document have been discussed on the mailing list <uri@w3.org> and <www-tag@w3.org>. Previous versions have couched 'tdb' and 'tdb' as URN namespaces, used different syntax for timestamps, and handled URIs with fragment identifiers differently. The variations are discussed in line.

The URI schemes defined here attempted to demonstrate ways of addressing several related problems:

1.1. Persistent identifiers

TOC

[RFC1737] defined several requirements for Uniform Resource Names. In particular, it requires "persistence":

> Persistence: It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.

Many people have wondered how to create globally unique and persistent identifiers. There are a number of URI schemes and URN namespaces already registered which are intended to provide persistence, as well as discussions of how to make DNS-based naming systems persistent through some allocation of persistent DNS names. However, guarantees of both uniqueness and persistence are difficult.

In most cases, the assurance of persistence is provided through a promise of good management practice, such as is encouraged in **"Cool URIs don't change"** [COOL].

A primary design goal for URIs is that they are intended to mean the same thing, no matter in what context they appear (the "Uniform" of "Uniform Resource Identifier"). However, even when URIs have "Uniform" meaning independent of the context of use, they don't usually guarantee stability over time. Despite best efforts and intentions, the mechanisms of resolution are subject to change in unpredictable ways: domain names can disappear or be reassigned, name resolving organizations can change in structure or responsibility, can disappear, merge, or change in other unpredictable ways.

The interpretation of most URNs depend significantly on the reliable behavior of name assignment and resolution authorities. The authorities are usually individuals or organizations trusted initially first to insure the uniqueness of assignment (so that the same name is not latter assigned for a different resource), and secondly to reliably maintain the link between the name and the named.

However, assignment and resolution authorities (whether individuals or organizations) all have a lifetime. The functioning of identifiers as unique holders of meaning depends on having a reliable infrastructure for the authority to maintain records, and for anyone to contact the authority or the authority's records to determine the thing identified.

TOC

1.2. URIs for anything

The description of URIs **[RFC3986]** describes a range for 'Resource' that is quite broad:

> This specification does not limit the scope of what might be a resource; rather, the term "resource" is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with a consistent purpose (e.g., "today's weather report for Los Angeles"), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., "parent" or "employee"), or numeric values (e.g., zero, one, and infinity).

However, no means is given for constructing URIs with such a range. How, then, might one construct a URI that identifies a human being, a corporation, or the value 'zero'?

One might wish to use a URI such as 'mailto' URI to identify a person, or use a 'http' URI to identify an abstract concept, with the indirection determined by context. Doing so, however, leaves the open the question of how one might identify, within the same context, both the system mailbox and the person to which it is assigned; both the resource reached via the HTTP protocol as determined by the 'http' URI and also the concept that resource describes.

The idea behind the 'tdb' URI scheme was to provide a ready assignment of URIs for things, in a way that clearly distinguished the URI for the thing from the URI of the media content that described it. The 'tdb' URI scheme provides a mechanism which is, at the same time:

persistent:
: The URI is not subject to reinterpretation over time.

objective:
: What's meant by the URI, or how it's to be interpreted, is explicit in the URI, and does not require an authority to adjudicate meaning.

useful for non-networked things:
: The scheme allows identification of resources outside the network: people, organizations, places, things, even abstract concepts.

without long-term administration requirement:
: The mechanism does not depend on administrative processes of authorities for reliable interpretation over time.

TOC

2. Syntax

A 'duri' URI takes the form:

```
    duri:<timestamp>:<embeddedURI>
```

and A 'tdb' URI takes a similar form:

```
    tdb:<timestamp>:<embeddedURI>
```

<embeddedURI> is an absoluteURI (as defined in **[RFC3986]**).

Note that the <embeddedURI> in duri and tdb URIs does not allow embedded fragment identifiers.

A <timestamp> in these URI schemes consists of a restricted subset of date times, as per **[RFC3339]**.

```
timestamp = date [ "T" time "Z" ]
date       =date-fullyear [ "-" date-month [ "-" date-mday ]]
time       = time-hour  [ ":" time-minute
             [ ":" time-second [ time-secfrac ]]]
```

where non-terminals "date-fullyear", "date-month", "date-mday", "time-hour", "time-minute", "time-second", "time-secfrac" are taken from **[RFC3339]**. The goal is to allow relatively short expressions with no ambiguity, but also allow arbitrary precision. While shorter forms are available (e.g., year-only timestamps), it is possible to use forms that are exactly compatible with **[RFC3339]** "date-time" non-terminal. (Note that earlier versions of this document proposed timestamps relative to International Atomic Time **[TAI]**, using a syntax without any puncutation at all, based on **[RFC2550]**. The syntactic variations don't seem to matter much.)

---

3. Semantics

TOC

3.1. 'duri' Semantics

TOC

The meaning of a 'duri' URI is "the resource that was identified by the <embeddedURI> at the the time given".

For example, "duri:2001:http://www.ietf.org" is a persistent identifier to the resource identified by "http://www.ietf.org" as of (the end of) 2001. (Note that during the many years of discussion around times within time intervals, various alternatives were proposed for whether a timestamp meant "as was stable during the entire period" or "at the beginning of the implied interval" vs. "at the end". Part of the question was whether it was reasonable to coin a URI using a year number at any time during that year, since the resource being identified may not have been active or established at the beginning of the year.)

3.2. 'tdb' Semantics

TOC

The 'tdb' URI scheme is intended to be useful for describing entities, concepts, abstractions, and other things which may not themselves be network accessible resources, but are (or at least have been at some point) described by network accessible resources.

Thus, a 'tdb' URI would be most useful when the <embeddedURI> identifies a 'document' of some sort (something a person could read, peruse, view, understand), and where the document thus identified describes some thing or concept. The 'tdb' URI itself then identifies

the subject of that document. This is similar to the common practice of giving a reference for a concept by including a pointer to a document phrase that defines the concept.

For example, one might use "tdb:2009:http://en.wikipedia.org/wiki/IETF" as a persistent identifier for the Internet Engineering Task Force (at least as described by the Wikipedia article as of 2009).

The 'tdb' URI scheme can be thought of as giving a way to invoke a level of semantic indirection to URI resolution.

Expressed in RDF, one might consider

```
    <duri:T:U;> foaf:primaryTopic <tdb:T:U>
```

where '-- foaf:primaryTopic --' is read '-- has, as its primary topic, --' (**[FOAF]** term "primaryTopic").

TOC

3.3. Timestamp Semantics

It is conventional in references and citations in printed works to include the date of publication; this practice provides important context. **[MANSTYLE]**.

While one could imagine using 'tdb' without a date, it would leave the possibility that a reference that is unambiguous at one time might become ambiguous at some other time. There are two ways that the date is useful for 'tdb' URIs: it fixes the time of access of the resource (and thus time variations of the description), and it fixes the time of interpretation, for descriptions whose meaning might vary. Thus, timestamps are useful in 'tdb' even when the resource identification does not vary (as with 'data' URIs).

While normally, in a literary work in natural language which makes a reference to another work, both the reference itself and the work referenced are dated, e.g., a footnote in an article written in 1967 might talk about a "private communication" which itself had a date. The difference between a URI and a conventional literary reference is the desire to be able to extract the URI from its context and still retain its meaning.

The meaning of a timestamp is the interval specified by the granularity of the time range indicated, in the UTC time zone, as described in **[RFC3339]**. If necessary, timestamps can include times and even fractional times, so that a generator of 'duri' or 'tdb' URIs can be arbitrarily precise.

If there is any ambiguity of the resource within the range of time indicated (for example, if the timestamp consists only of a year, and the resource changes over the course of the year), then the last available resource state within the the range indicated should be used.

Timestamps are allowed to be specified with as much precision as needed. This keeps most 'duri' and 'tdb' URIs relatively short.

TOC

4. Use as a Locator

A 'duri' URI is not directly useful as a resource locator, since many resources vary their content over time.

A 'tdb' URI is not a resource locator in a practical sense, since it explicitly requires human interpretation. However, it allows one to know that a resource was described at some point in time; whether the description is still available, or whether that description is still meaningful, is not guaranteed.

TOC

5. Hierarchy

For 'tdb', the "thing described by" a resource may bear little relationship to the "thing described by" a relative pointer, so the 'tdb' URI scheme seems to have no use cases for using "/" as a hierarchical delimiter.

However, 'duri' URIs can often be used with relative URI references with some amount of reliability.

---

6. Additional Considerations

TOC

6.1. Embedded URI schemes

TOC

The 'tdb' scheme is primarily useful when the <embeddedURI> identifies an "information resources".

For example, a 'http' URI might refer to a web page or the subject of a web as it was described at the given time. This can be a way of referring to a web site at some time in the past, or an organization that has changed, merged, split, or disappeared.

A 'file' URI with a known-to-be unique host name might also be used within a 'tdb' URI, for example,

```
tdb:2001-08-14T14:23:27Z:file://this.example.com/c|/temp/test.txt
```

This use is primarily focused on providing a unique way of identifying something, even if the referent is not widely known. (Using 'file' URIs in this way without a fully qualified domain name as the authority would not be appropriate, because the interpretation is not uniform even at any particular instant.)

One might consider using 'tdb' with a 'data' URI to designate concepts that can be described uniquely briefly inline. For example,

```
tdb:2001:data:,The%20US%20president
```

names the concept described by the (text/plain) string "The US president" at the end of 2001. Of course, this practice is only useful if the referent of the data is (or was at the time) unique. Since a 'data' URI does not contain a way to designate content-language, the string in question should be unambiguous as to its language. In the case of 'data' URIs, there is no assigning authority at all; the interpretation of the 'tdb' depend on the interpreting community.

Using 'duri' with an embedded 'urn' might not seem to be too useful, but it might be useful where the assignment of names in a URN namespace are not, in practice, permanent, and one wants to refer to the assignment as of a given date. In this case, it is possible to use a "urn" within a 'duri', e.g.,

```
duri:2000:urn:ietf:std:50
```

might be used to refer to "the document that the IETF considered to be STD 50, at the end of 2000".

For 'tdb', many URIs identify resources which do not clearly describe anything at all. Even so, some care should be given; for example, the home page for an organization might not be as good a resource to use to describe an organization as the organization's "about" page or an external authority's description of the organization. It is up to the minter of the 'tdb' URI to choose wisely.

6.2. Useful timestamps

TOC

Timestamps in the future are suspect, because the future content of a description resource cannot usually be reliably predicted. Timestamps which preceed the availability of the description resource should not be used either. For example, using a http URI with a timestamp before the description resource has been created is also not recommended.

However, although these practices are not recommended, there is no assurance that they haven't been used; by itself, a 'tdb' URI by itself does not constitute an assertion that the description resource was available or assigned at the date specified.

Note that the use of the "last available state" allows for the conventional bibliographic convention that a work published in 2009 can use "2009" as the date string, to refer to the work in the year of publication.

6.3. Free assignment

TOC

Because of the many possible schemes that can be used in the embedded URI, there should be no difficulty in almost any computational process being able to assign 'duri' or 'tdb' URIs at will. Of course, it is necessary for there to be some resource which is available at some point in time, and to have a clock which is accurate to the granularity of the frequency of assignment.

6.4. Resolution

TOC

There are no direct resolution servers or processes for 'duri' or 'tdb' URIs. However, a 'duri' URI might be "resolvable" in the sense that a resource that was accessed at a point in time might have the result of that access cached or archived in an Internet archive service. See, for example, the "Internet Archive" project **[archive]**. And a 'tdb' URI is "resolvable" to the extent that the description resource can be accessed and interpreted.

Clients without access to an Internet archive service might take the embedded URI of a 'duri' and attempt resolution of that identifier. This will give an approximation whose reliability depends on the what has happened in the time since the date indicated.

6.5. Ambiguous Resources

TOC

There are many URIs which are, unfortunately, not particularly "uniform", in the sense that two clients can observe completely different content for the same resource, at exactly the same time. These resources are not so useful with 'tdb' URIs, since time alone is not adequate to identify precisely because the results of access depends on other details of the observation (e.g., IP address, cookies, HTTP request headers, which physical server responded, etc.).

When using 'duri' with URIs for which result of access varies depending on other conditions of access, all a 'duri' URI really says is that someone observed something at the given URI at a certain time.

6.6. Other Ambiguities

TOC

Unfortunately, the scope of a URI is always ambiguous as a reference point for both documents and things described by them. When I point to a web site (http://www.ietf.org), do I mean just the home page or the whole site? Do I mean just the HTML there, or also the embedded images and other things that are displayed? While "tdb" and "duri" attempt to nail down two of the more important areas of ambiguity (use/mention and time varying), other dimensions of ambiguity remain.

6.7. Why Names with Semantics?

TOC

The 'tdb' URI scheme differs from other URI or URN methods for identifying abstractions because the designation of what is actually identified by the 'tdb' doesn't depend on knowing the intention of the assigner of the identifier. Unlike the 'tag' **[RFC4151]**, 'info' **[RFC4452]**, 'cid' or 'mid' **[RFC2392]** schemes, for example, the identification does not depend on any authority or process not reusing the same identifier at some later point for a different concept, or maintaining any records or meaning. In these other schemes, the assigning authority only insures uniqueness at the time of minting, with some other agent or process or context providing the authority to interpret the meaning of the identifier in the future. In this sense, 'duri' and 'tdb' are different, in that it is the agreement between the describer (the agent creating the URI) and the receiver of the URI (the agent interpreting the URI) to agree upon the semantics without any reference to any third party.

6.8. Avoiding MetaData

TOC

One might consider the timestamp in a 'duri' or 'tdb' URI to be just one piece of additional metadata about the URI, and consider adding other pieces of metadata as annotation.

However, the use of the timestamp is intended primarily as a mechanism of accomplishing uniqueness over time. No other bit of metadata or description readily fills that purpose. Further, the date is not descriptive (an assertion about the URI) but merely refining.

6.9. Avoiding 'duri' and 'tdb'

TOC

Many applications of URIs already provide a context of timestamp. For example, one could imagine a hypertext system where the URIs contained within a document were intended to refer to the resources as of the date of the enclosing document. This would be a reasonable interpretation of URIs within an Internet archive system, for example.

Some applications of URIs already implicitly use the level of interpretive indirection that is explicit with 'tdb', For example, within an ontology language definition, the URIs used for abstract concepts, individuals and so forth are generally considered the "thing described by" the URI.

In addition, the 'application/rdf+xml' Media Type **[RFC3870]** uses the fragment identifier resolution as an explicit way of identifying things that are described by an RDF document.

6.10. 'tdb' and levels of indirection

TOC

The 'tdb' scheme introduces a level of semantic indirection. The puzzles and confusions about use and mention, name and reference, and levels of indirection have been puzzling and amusing for quite a while.

> "It's long," said the Knight, "but it's very, very beautiful. Everybody that hears me sing it--either it brings tears into their eyes, or else--"
> "Or else what?" said Alice, for the Knight had made a sudden pause.
> "Or else it doesn't, you know. The name of the song is called 'Haddock's Eyes.'"
> "Oh, that's the name of the song, is it?" Alice said, trying to feel interested.
> "No, you don't understand," the knight said, looking a little vexed. "That's what the name is called. The name really is 'The Aged Aged Man.'"
> "Then I ought to have said 'That's what the song is called'?" Alice corrected herself.
> "No, you oughtn't: that's quite another thing! The song is called 'Ways and Means': but that's only what it's called, you know!"
> "Well, what is the song, then?" said Alice, who was by this time completely bewildered.

"I was coming to that," the Knight said. "The song really is 'A-sitting On A Gate': and the tune's my own invention." **[LOOK]**

---

7. URI Specification Templates

TOC

7.1. 'duri' Scheme Template

TOC

URI scheme name:
duri
Status:
permanent
URI scheme syntax:
The syntax is described in detail in **Section 2** of this document. Briefly, the syntax is duri:<timestamp>:<embeddedURI>
where <timestamp> is year, year-month or date taken from **[RFC3339]**, and <embeddedURI> is an <absoluteURI> from **[RFC3986]**.
URI scheme semantics:
A URI as of a particular time. Semantics are described in detail in this document.
Encoding considerations:
'duri' URIs consist of a prefix followed by another URI, and should have the same encoding considerations as others.
Applications/protocols that use this URI scheme name:
Limited: this scheme was originally developed as a "thought experiment".
Interoperability considerations:
The actual interoperability with Internet archiving services needs further exploration.
Security considerations:
See **Section 9** of this document.
Contact:
Larry Masinter tdb:2012:http://larry.masinter.net
Author/Change controller:
Contact, as above.
References:
See References of this document.

7.2. 'tdb' Scheme Template

TOC

URI scheme name:
tdb
Status:
permanent
URI scheme syntax:
The syntax is described in detail in **Section 2** of this document. Briefly, the syntax is tdb:<timestamp>:<embeddedURI>
where <timestamp> is year, year-month or date taken from **[RFC3339]**, and <embeddedURI> is an <absoluteURI> from **[RFC3986]**.
URI scheme semantics:
Semantic indirection at indicated date. Semantics are described in detail in this document.
Encoding considerations:
'tdb' URIs consist of a prefix followed by another URI, and should have the same encoding considerations as others.
Applications/protocols that use this URI scheme name:
Limited: This scheme was originally designed as a "thought experiment", as a way resolve some of the use/mention ambiguities in semantic web applications that wish to "denote" concepts and other ideas and not just access resources over the Internet.
Interoperability considerations:
Existing semantic web applications may have other means of fixing meaning at a

particular time or semantic indirection, and do not fix description by time.

Security considerations:

See **Section 9** of this document.

Contact:

Larry Masinter tdb:2012:http://larry.masinter.net

Author/Change controller:

as above

References:

See References of this document.

---

8. IANA considerations

TOC

This document includes two URI scheme registrations (**Section 7** that should be entered into the IANA registry of URI schemes as a permanent registration (once approved).

---

9. Security Considerations

TOC

'tdb' identifiers are not any more reliable because they have dates. URIs don't contain enough information to supply the authority for deciding what was or wasn't at a given URI at a given date.

---

10. Acknowledgements

TOC

There have been many discussions over several years on the relationship of URLs, URNs, URIs, resources and resource identifiers, with many contributions. Particular thanks to those who have severely critiqued various versions of this document in the past, including Jonathan Rees, Nathan, Alan Ruttenberg, Alfred Hines, Herbert Van de Sompel, Al Gilman, Aaron Swartz, Brian McBride, Stuart Williams, Michael Mealling, Ray Denenberg and Pat Hayes.

---

11. References

TOC

11.1. Normative References

TOC

[RFC3339] Klyne, G. and C. Newman, "**Date an Time on the Internet: Timestamps**," RFC 3339, July 2002.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "**Uniform Resource Identifiers (URI): Generic Syntax**," RFC 3986, January 2005.

11.2. Informative References

TOC

[COOL] Berners-Lee, T., "**Cool URIs don't change**," 1998.

[FOAF] Brickley, D., "**FOAF Vocabulary Specification 0.98**," August 2010.

[LOOK] Carroll, L., "**Through the Looking Glass**," 1872.

[MANSTYLE] The University of Chicago, "**The Chicago Manual Of Style Online: Chicago-Style Citation Quick Guide**," 2012.

[MEMENTO] VandeSompel, H., Nelson, M., and R. Sanderson, "**HTTP framework for time-based access to resource states -- Memento**," December 2011.

[RFC1737] Sollins, K., "**Functional Requirements for Uniform Resource Names**," RFC 1737, December 1994.

[RFC2392] Levinson, E., "**Content-ID and Message-ID Uniform Resource Locators**," RFC 2392, August 1998 (**TXT**, **HTML**, **XML**).

[RFC2550] Glassman, S., Manasse, M., and J. Mogul, "**Y10K and Beyond**," April 1999.

[RFC3870] Swartz, A., "**application/rdf+xml Media Type Registration**," RFC 3870, September 2004.

[RFC4151] Kindberg, T. and S. Hawke, "**The 'tag' URI Scheme**," RFC 4151, October 2005 (**TXT**).

[RFC4452] Van de Sompel, H., Hammond, T., Neylon, E., and S. Weibel, "**The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces**," RFC 4452, April 2006 (**TXT**).

[TAI] Bureau International des Poids et Mesures, "International Atomic Time."

[archive] Kahle, B., "**Preserving the Internet**," Scientific American , March 1997.

Author's Address

Larry Masinter
Adobe
345 Park Ave
San Jose, CA 95110
US
Phone: +1 408 536 3024
Email: **LMM@acm.org**
URI: **http://larry.masinter.net**