Tim Kindberg
Internet Draft Hewlett-Packard Corp.
Document: draft-kindberg-tag-uri-00.txt
Expires: September 1, 2001 March 2001

The tag: URI scheme

STATUS OF THIS MEMO

ABSTRACT

This document describes the 'tag:' Uniform Resource Identifier (URI) scheme for identifiers that are unique across space and time. Identifiers belonging to this scheme are distinct from most other URIs in that they are intended for use that is independent of any particular method for resource location or name resolution. A 'tag:' URI may be used purely as an identifier that distinguishes one entity from another. It may also be presented to services for resolution into a web resource or into one or more further URIs, but no particular resolution scheme is implied or

preferred by a 'tag:' identifier itself. Unlike UUIDs or GUIDs such as 'uuid:' and 'urn:oid' URIs, which also have some of the above properties, 'tag:' identifiers are designed to be tractable to humans. Furthermore, they have many of the desirable properties that 'http:' URLs have when used as identifiers, but none of the drawbacks.

Table of Contents

INTRODUCTION

A 'tag:' identifier is a type of Uniform Resource Identifier (URI) [X] designed to meet the following requirements:

1) Identifiers are unique across space and time and come from a practically inexhaustible supply;
2) identifiers are convenient for humans to read, type etc.;
3) zero registration cost, at least to current holders of a registered domain name, and negligible cost to mint new identifiers;
4) easy identification of the organisation that has minted the identifier, should that be desirable;
5) no commitment to any particular resource-location or identifier-resolution scheme.

The above requirements obtain in the case that a user wants to place identifiers on their documents.

1) They want to be sure that the identifier is unique. Global uniqueness is valuable because it guarantees that one identifier cannot conflict with another, whatever the scope of future sharing.
2) The user would like the identifier to be tractable to humans: they should be able to type it into a form; it could contain a hint about how to categorise the document, or the date of issue.

3) They do not want to have to communicate with anyone else in order to create identifiers for their documents.
4) It is natural to use a name associated with the user or their organisation within the identifier, since that is the origin of the identifier.
5) As good net citizens, the user does not want to use an identifier that might be assumed by software to imply the existence of a corresponding resource in a default binding scheme – so that an attempt to retrieve that resource is likely but doomed to failure. Of course, this leaves them free to exploit the identifier in particular applications and services, where the context is clear.

Existing identification schemes satisfy some but not all of the general requirements 1-5. For example:

UUIDs [X] are hard for humans to read and the assigning organisation is not explicit.

OIDs [x] and DOIs [x] both require naming authorities to register themselves, even if they already hold a domain name registration.

URNs [x] are intended to be resolvable in a default naming context. Software encountering a URN in a document is liable to attempt to resolve it, even though the identifier has not registered any resource in that context.

URLs (in particular, 'http:' URLs) are sometimes used as ersatz identifiers that satisfy most of our requirements. Many users and organisations have already registered a domain name, and the use of the domain name to mint identifiers comes at no additional cost. But there are several drawbacks to URLs-as-identifiers:
A) Many pieces of software might try to dereference a URL-as-identifier, even though there is no resource at the 'location'.
B) We can't find out who minted a URL-as-identifier, if the domain has changed hands. If Smith registers champignon.net and then Jones registers it, no-one can tell who minted http://champignon.net/99.
C) The new holder of a domain name can't be sure that they are minting new names. Using the example from (B), how can Jones know, in general, whether Smith has already used http://champignon.net/99?

THE 'TAG:' URI SCHEME

The general form of a 'tag:' URI is:

```
          tag:nameSpace:specific
```

Where:

```
          nameSpace = domainName</DateQualifier>
          domainName          = any well-formed domain name [x]
          dateQualifier       = year | year.month | year.month.day
          year                = [2-9][0-9][0-9][0-9]
          month               = 0[1-9]|[1-12]
          day                 = 0[1-9]|[1-31]
          specific            = any string that makes the URI well-formed
```

[x]

'nameSpace' is the name space part of the URI: it is a well-formed, domain name registered to the entity that has minted the URI, optionally post-fixed by a date qualifier.

'specific' is the name-space-specific part of the URI: it is any string of valid URI characters chosen by the minter of the tag.

For example:

```
          tag:hpl.hp.com:tst.1234567890
          tag:exploratorium.edu:pi.99
          tag:myIDs.com:TimKindberg/doc.101
          tag:champignon.net:99
          tag:champignon.net/2001.3.2:99
          tag:champignon.net/2001.04:100
          tag:champignon.net/2002:docs/research/99
```

Anyone who holds the current registration to a domain name has the right to mint identifiers rooted at that name, as long as the domain name is date-qualified, should that be necessary for uniqueness (see 'transfers of domain names').

For example, Hewlett-Packard Laboratories holds the registration for hpl.hp.com and can mint any tag URIs rooted at that name; but they may not mint names under domain names not registered to them, such as champignon.net.

TRANSFERS OF DOMAIN NAMES

The 'tag:' scheme copes with transfers of a domain name's registration from one party to another. The 'dateQualifier' is used to guarantee uniqueness of 'nameSpace' across several registrations of the domain.

For example, an organisation or individual that knows itself to be the first registrant of champignon.net can mint names of the form tag:champignon.net:...

However, in some cases the domain will change hands. For example, suppose that on March 2, 2001, the champignon.net domain registration becomes held by a new entity. That entity must date-qualify the domain name to ensure that its tag name space is unique. It must postfix the full date (day, month, year) during the month of March; it must postfix at lease the month and year during the remainder of that year; it must postfix at least the year subsequently.

An alternative to date qualification would be ordinal-qualification: the second registrant of champignon.net could use the name space champignon.net.2; the third champignon.net.3, etc. However, no mechanism exists for tracking that ordinality. The advantage of date qualification is that each registrant knows its responsibilities and relies on no other agency for correct namespace designation.

EQUALITY OF TAGS

Two tag URIs are equal if and only if:
their name space identifiers match and
their specific identifiers are identical, including their case.

Two tag name space identifiers match if their canonical forms match, where their canonical form is obtained by reducing upper case characters to lower case and removing leading zeros from date components.

Thus, tag:champignon.net.2.3.2001:99 and tag:champignon.net.4.2001:99 are unequal, as are tag:champignon.net.2002:test and tag:champignon.net.2002:Test.

But the following two tags are equal: tag:CHAMPIGNON.NET:99 and tag:champignon.net:99

SECURITY CONSIDERATIONS

No mechanism can prevent an organisation from using another's domain name. A malicious party could pollute another party's tag name space. Only the threat of legal action counts against that.

REFERENCES

AUTHOR'S ADDRESS

Tim Kindberg
Hewlett-Packard Laboratories Phone: 1-650-857-5609
1501 Page Mill Road
Palo Alto
CA 94304
USA Email: timothy@hpl.hp.com