Internet Draft
Document: draft-kindberg-tag-uri-00.txt
Expires: October 1, 2001

Tim Kindberg
Hewlett-Packard Corporation
Sandro Hawke
World Wide Web Consortium

April 2001

The tag: URI scheme (DRAFT)

STATUS OF THIS MEMO

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html.

This Internet-draft will expire on October 1, 2001.

Copyright Notice Copyright (C) The Internet Society (2001). All Rights Reserved.

DISCLAIMER. The views and opinions of authors expressed herein do not necessarily state or reflect those of the World Wide Web Consortium, and may not be used for advertising or product endorsement purposes. This proposal has not undergone technical review within the Consortium and must not be construed as a Consortium recommendation.

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 uses that are 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:' URI itself. Unlike UUIDs or GUIDs such as 'uuid:' and 'urn:oid' URIs, which also have some of the above properties, 'tag:' URIs 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.

0. TERMINOLOGY

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

1. INTRODUCTION

A 'tag:' identifier is a type of Uniform Resource Identifier (URI) [RFC2396] 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 mint (create), read, type etc.;
3) zero registration cost, at least to holders of domain names or email addresses; and negligible cost to mint new identifiers;
4) easy identification of the entity that has minted the identifier, should that be desirable;
5) independence of any particular resource-location or identifier-resolution scheme.

For example, the above requirements may apply in the case of a user who wants to place identifiers on their documents:

A) They want to be sure that the identifier is unique. Global uniqueness is valuable because it guarantees that one identifier cannot conflict with another, however identifiers become shared.
B) It is useful for the identifier to be tractable to humans: they should be able to mint new identifiers conveniently and to type them into forms; the identifiers should be able to contain a hint about how to categorise the document.
C) They do not want to have to communicate with anyone else in order to mint identifiers for their documents.
D) 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.
E) As a good net citizen, 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 [UUID, ISO-11578] are hard for humans to read and the assigning organisation is not explicit.

OIDs [OID, RFC3061] and Digital Object Identifiers [DOI] require naming authorities to register themselves, even if they already hold a domain name registration.

URNs [RFC2141] 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 entity that minted the identifier has not bound 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 drawbacks to URLs-as-identifiers:

1) Software might try to dereference a URL-as-identifier, even though there is no resource at the 'location'.
2) The new holder of a domain name can't be sure that they are minting new names. If Smith registers champignon.net and then Jones registers it, how can Jones know, in general, whether Smith has already used http://champignon.net/99?
3) We can't find out who minted a URL-as-identifier, if the domain has changed hands. Using the example from (B), no-one can tell who minted http://champignon.net/99.

Adding a fragment "#fragment" on the end of a URL (thus forming a URI reference) does not, of itself, remove the undesirable characteristics of URLs as identifiers.

2. THE 'TAG:' URI SCHEME

Examples of tag: URIs (also known as 'tags') are:

```
tag:hpl.hp.com/1:tst.1234567890
tag:exploratorium.edu/1:pi.99
tag:sandro@w3c.org/1:my-dog
tag:myIDs.com/1:TimKindberg/doc.101
tag:champignon.net/1
tag:champignon.net/1-3-22:99
tag:champignon.net/2-4:100
```

Each tag consists of a 'tag authority' followed, optionally, by a specific identifier. The tag authority consists of an 'authority name' -- a fully qualified domain name or an email address containing a fully qualified domain name -- followed by a date. The tag authority is globally unique because domain names and email addresses are assigned to at most one entity at a time and that entity can be sure of minting unique identifiers.

The date specifies any particular day on which the authority name was assigned to the minting entity. Depending on defaults, dates appear in one of three forms: 'year', 'year-month' or 'year-month-day'. Several abbreviations are mandated, in the interests of being able to transcribe

tags into identification technologies of limited capacity (e.g. barcodes), while ensuring that tags are single-valued, for easy comparison:

1) The year, which MUST be at least 2001, is abbreviated by subtracting 2000, so that 2001 is written '1', 11958 will be '9958', etc.
2) The month and day default to 1. A day value of 1 MUST be omitted. A month value of 1 MUST be omitted unless it is followed by a day value other than 1. For example, '1' is the date 2001/1/1, '3-4' is 2003/4/1. The date values '2-1' and '2-4-1' are not allowed but '2-1-4' is allowed.
3) Date components MUST NOT contain a leading zero.

Note that dates, such as '1' and '3-4', each specify a single day. They are not to be taken as 'the whole of 2001' and 'the whole of April 2003', respectively.

A tag authority mints specific identifiers that are unique within its context, in accordance with any internal scheme that uses only URI characters. Some tag authorities (e.g. corporations, mailing lists) consist of many people, in which case group decision-making and record-keeping procedures are required to achieve uniqueness.

Entities that were assigned an authority name on a given date MAY mint tags rooted at that date-qualified name. An entity MUST NOT mint tags under an authority name that was assigned to a different entity on the given date, and it MUST NOT mint tags under a future date. We take the date of assignment of an authority name to be the first day for which the assignment is held at midnight (00:00) UTC.

An entity that acquires an authority name immediately after a period during which the name was unassigned MAY mint tags as if the entity was assigned the name during the unassigned period. This practice has considerable potential for error and MUST NOT be used unless the entity has substantial evidence that the name was unassigned during that period. The authors are currently unaware of any mechanism that would count as evidence, other than daily polling of the 'whois' registry.

For example, Hewlett-Packard holds the domain registration for hpl.hp.com and may mint any tags rooted at that name with a current or past date when it held the registration (2001/1/1 or later). It must not mint tags such as tag:champignon.net/1 under domain names not registered to it. It must not mint tags dated in the future, such as tag:hpl.hp.com/999. If it obtains assignment of extremelyunlikelytobeassigned.org on 2001/5/1, then it must not mint tags under extremelyunlikelytobeassigned.org/1 unless it has found substantial evidence that that name was continuously unassigned between 2001/1/1 and 2001/5/1.

The general syntax of a 'tag:' URI, in BNF, is:

```
      tagURI         ::= "tag:" tagAuthority [":" specific]

Where:
      tagAuthority   ::= authorityName "/" date
      authorityName  ::= DNSname | emailAddress
```

```
DNSname        ::= DNScomp | DNSname "." DNScomp   ; [RFC 1035]
DNScomp        ::= lowAlphaNum *(lowAlphaNum | "-") lowAlphaNum
emailAddress   ::= 1*(lowAlphaNum |"-"|"."|"_") "@" DNSname
lowAlphaNum    ::= dig | "a"|"b"| ... "y"|"z"  ; all lwr case alphas
date           ::= year ["-" (monthNon1 | month "-" day)]
year           ::= digitNon0 [*dig]
monthNon1      ::= digit2+ | "10" | "11" | "12"
month          ::= "1" | monthNon1
day            ::= digit2+ | ("1"|"2") dig | "30" |"31"
dig            ::= "0" | digitNon0
digitNon0      ::= "1" | digit2+
digit2+        ::= "2"|"3"|"4"|"5"|"6"|"7"|"8"|"9"
specific       ::= 1*(URIchars)  ; [RFC 2396]
```

The component 'tagAuthority' is the name space part of the URI. This MUST be expressed in lower case. The domain name in 'authorityName' (whether an email address or a simple domain name) MUST be fully qualified.

Authority names could, in principle, belong to any syntactically distinct namespaces whose names are assigned to a unique entity at a time. Those include, for example, certain IP addresses, certain MAC addresses, and telephone numbers. However, to simplify the tag scheme, we restrict authority names to be assigned domain names and email addresses. Future standards efforts may allow use of such names following syntax that is disjoint from this syntax. To allow for such developments, software that processes tags MUST NOT reject tags on the grounds that they are outside the syntax defined above.

The component 'specific' is the name-space-specific part of the URI: it is any string of valid URI characters [RFC2396] chosen by the minter of the URI. Specific identifiers MUST be single-valued: that is, all syntactically distinct 'specific' strings must correspond to distinct identifiers. It is RECOMMENDED that specific identifiers should be human-friendly.

3. MEETING REQUIREMENTS 1-5

Requirement 2 of Section 1 -- convenience for humans -- is met by the URL-like syntax for tag authorities. However, the onus is on individual naming authorities to use human-friendly specific identifiers.

Requirement 3 -- negligible costs -- follows from use of domain names and email addresses. Those identifiers are already held by many individuals and organisations and are cheap to obtain. Specific identifiers may be minted without communication with any other entity.

Requirement 4 -- convenient identification of the minting entity, where desirable -- also follows from use of domain names and email addresses. An entity may use its authority name in a tag if it wishes to be so identified; alternatively, it could lease identifiers privately from another entity ('myTags.com').

Requirement 5 -- independence of resolution schemes -- is asserted by definition. However, this state of affairs is subject to actual usage conventions.

Requirement 1 specifies uniqueness over space and time. Tag URIs meet that requirement by using uniquely assigned authority names and by handling transfers of their assignment, e.g. the transfer of a domain name's registration from one entity to another. The date is used to guarantee uniqueness of 'tagAuthority' across assignments of the authority name.

For example, suppose that on April 2, 2001, the champignon.net domain registration becomes assigned to a new entity. That entity must qualify the domain name with a date on which it is or was assigned to it, to ensure that its tag authority is and will remain unique. In particular, it must take care not to use defaults in such a way as to specify an earlier date. For example, the new assignee of champignon.net may use '1-4-2', '1-5' or '2' (assuming it retains the assignment) but not '1' or '1-4'.

4. EQUALITY OF TAGS

The tag syntax rules in Section 2 uniquely determine tag authority identifiers for any particular authority and date. Furthermore, specific identifiers are mandated to be single-valued.

Therefore, two tag URIs are equal if and only if they are identical as character strings.

5. SECURITY CONSIDERATIONS

Minting a tag, by itself, is an operation internal to the minting entity with no external consequences. The consequences of using an improperly minted tag (due to malice or error) in a binding protocol or other protocol depend on the protocol, and must be considered in the design of any protocol that uses tags.

6. FURTHER INFORMATION

Further information about the tag URI scheme -- motivation, genesis and discussion -- can be obtained from http://www.taguri.org.

REFERENCES

[DOI] Norman Paskin (1997). Information Identifiers. Learned Publishing, Vol. 10, No. 2, pp. 135-156, April. See also www.doi.org.

[ISO-11578] ISO (International Organization for Standardization). ISO/IEC 11578:1996. "Information technology - Open Systems Interconnection - Remote Procedure Call (RPC)"

[OID] ITU-T recommendation X.208 (ASN.1). See also RFC 1778.

[RFC822] David H. Crocker (1982). Standard for the format of ARPA Internet text messages.

[RFC1035] P. Mocapetris (1987). Domain Names - implementation and specification.

[RFC2141] R. Moats (1997). URN syntax.

[RFC2396] T. Berners-Lee, R. Fielding, L. Masinter (1998). Uniform Resource Identifiers (URI): Generic Syntax.

[RFC3061] M. Mealling (2001). A URN Namespace of Object Identifiers.

[UUID] Paul Leach, Rich Salz (1997). UUIDs and GUIDs. Internet-Draft Draft-leach-uuids-01.

AUTHORS' ADDRESSES

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

Sandro Hawke
World Wide Web Consortium
200 Technology Square
Cambridge, MA 02139, USA
Tel: +1 617 253-7288
Email: sandro@w3.org