Network Working Group
Request for Comments: 4151
Category: Informational

T. Kindberg
Hewlett-Packard Corporation
S. Hawke
World Wide Web Consortium

October 2005

The 'tag' URI Scheme

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

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. Tag URIs (also known as "tags") are designed to be unique across space and time while being tractable to humans. They are distinct from most other URIs in that they have no authoritative resolution mechanism. A tag may be used purely as an entity identifier. Furthermore, using tags has some advantages over the common practice of using "http" URIs as identifiers for non-HTTP-accessible resources.

Table of Contents

1. Introduction

A tag is a type of Uniform Resource Identifier (URI) [1] designed to meet the following requirements:

1. Identifiers are likely to be unique across space and time, and come from a practically inexhaustible supply.

2. Identifiers are relatively convenient for humans to mint (create), read, type, remember etc.

3. No central registration is necessary, at least for holders of domain names or email addresses; and there is negligible cost to mint each new identifier.

4. The identifiers are independent of any particular resolution scheme.

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

a. The user wants to be reasonably sure that the identifier is unique. Global uniqueness is valuable because it prevents identifiers from becoming unintentionally ambiguous.

b. The identifiers should be tractable to the user, who should, for example, be able to mint new identifiers conveniently, to memorise them, and to type them into emails and forms.

c. The user does not want to have to communicate with anyone else in order to mint identifiers for their documents.

d. The user wants to avoid identifiers that might be taken to imply the existence of an electronic resource accessible via a default resolution mechanism, when no such electronic resource exists.

Existing identification schemes satisfy some, but not all, of the requirements above. For example:

UUIDs [5], [6] are hard for humans to read.

OIDs [7], [8] and Digital Object Identifiers [9] require entities to register as naming authorities, even in cases where the entity already holds a domain name registration.

URLs (in particular, "http" URLs) are sometimes used as identifiers that satisfy most of the above 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:

- An attempt may be made to resolve a URL-as-identifier, even though there is no resource accessible at the "location".
- Domain names change hands and the new assignee of a domain name can't be sure that they are minting new names. For example, if example.org is assigned first to a user Smith and then to a user Jones, there is no systematic way for Jones to tell whether Smith has already used a particular identifier such as http://example.org/9999.
- Entities could rely on purl.org or a similar service as a (first-come, first-served) assigner of unique URIs; but a solution without reliance upon another entity such as the Online Computer Library Center (OCLC, which runs purl.org) may be preferable.

Lastly, many entities -- especially individuals -- are assignees of email addresses but not domain names. It would be preferable to enable those entities to mint unique identifiers.

1.1. 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.

1.2. Further Information and Discussion of this Document

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

Earlier versions of this document have been discussed on uri@w3.org. The authors welcome further discussion and comments.

2. Tag Syntax and Rules

This section first specifies the syntax of tag URIs and gives examples. It then describes a set of rules for minting tags that is designed to make them unique. Finally, it discusses the resolution and comparison of tags.

2.1. Tag Syntax and Examples

The general syntax of a tag URI, in ABNF [2], is:

```
tagURI = "tag:" taggingEntity ":" specific [ "#" fragment ]
```

Where:

```
taggingEntity = authorityName "," date
authorityName = DNSname / emailAddress
date = year ["-" month ["-" day]]
year = 4DIGIT
month = 2DIGIT
day = 2DIGIT
DNSname = DNScomp *( "."  DNScomp ) ; see RFC 1035 [3]
DNScomp = alphaNum [*(alphaNum /"-") alphaNum]
emailAddress = 1*(alphaNum /"-"/"."/"_") "@" DNSname
alphaNum = DIGIT / ALPHA
specific = *( pchar / "/" / "?" ) ; pchar from RFC 3986 [1]
fragment = *( pchar / "/" / "?" ) ; same as RFC 3986 [1]
```

The component "taggingEntity" is the name space part of the URI. To avoid ambiguity, the domain name in "authorityName" (whether an email address or a simple domain name) MUST be fully qualified. It is RECOMMENDED that the domain name should be in lowercase form. Alternative formulations of the same authority name will be counted as distinct and, hence, tags containing them will be unequal (see Section 2.4). For example, tags beginning "tag:EXAMPLE.com,2000:" are never equal to those beginning "tag:example.com,2000:", even though they refer to the same domain name.

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 domain names and email addresses. Future standards efforts may allow use of other authority names following syntax that is disjoint from this syntax. To allow for such developments, software that processes tags MUST NOT reject them 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 a string of URI characters (see restrictions in syntax specification) chosen by the minter of the URI. Note that the "specific" component allows for "query" subcomponents as defined in RFC 3986 [1]. It is RECOMMENDED that specific identifiers should be human-friendly.

Tag URIs may optionally end in a fragment identifier, in accordance with the general syntax of RFC 3986 [1].

In the interests of tractability to humans, tags SHOULD NOT be minted with percent-encoded parts. However, the tag syntax does allow percent-encoded characters in the "pchar" elements (defined in RFC 3986 [1]).

Examples of tag URIs are:

```
tag:timothy@hpl.hp.com,2001:web/externalHome
tag:sandro@w3.org,2004-05:Sandro
tag:my-ids.com,2001-09-15:TimKindberg:presentations:UBath2004-05-19
tag:blogger.com,1999:blog-555
tag:yaml.org,2002:int
```

2.2. Rules for Minting Tags

As Section 2.1 has specified, each tag includes a "tagging entity" followed, optionally, by a specific identifier. The tagging entity is designated by an "authority name" -- a fully qualified domain name or an email address containing a fully qualified domain name -- followed by a date. The date is chosen to make the tagging entity globally unique, exploiting the fact that domain names and email addresses are assigned to at most one entity at a time. That entity then ensures that it mints unique identifiers.

The date specifies, according to the Gregorian calendar and UTC, any particular day on which the authority name was assigned to the tagging entity at 00:00 UTC (the start of the day). The date MAY be a past or present date on which the authority name was assigned at

that moment. The date is specified using one of the "YYYY", "YYYY-MM" and "YYYY-MM-DD" formats allowed by the ISO 8601 standard [4] (see also RFC 3339 [10]). The tag specification permits no other formats. Tagging entities MUST ascertain the date with sufficient accuracy to avoid accidentally using a date on which the authority name was not, in fact, assigned (many computers and mobile devices have poorly synchronised clocks). The date MUST be reckoned from UTC, which may differ from the date in the tagging entity's local timezone at 00:00 UTC. That distinction can generally be safely ignored in practice, but not on the day of the authority name's assignment. In principle it would otherwise be possible on that day for the previous assignee and the new assignee to use the same date and, thus, mint the same tags.

In the interests of brevity, the month and day default to "01". A day value of "01" MAY be omitted; a month value of "01" MAY be omitted unless it is followed by a day value other than "01". For example, "2001-07" is the date 2001-07-01 and "2000" is the date 2000-01-01. All date formulations specify a moment (00:00 UTC) of a single day, and not a period of a day or more such as "the whole of July 2001" or "the whole of 2000". Assignment at that moment is all that is required to use a given date.

Tagging entities should be aware that alternative formulations of the same date will be counted as distinct and, hence, tags containing them will be unequal. For example, tags beginning "tag:example.com,2000:" are never equal to those beginning "tag:example.com,2000-01-01:", even though they refer to the same date (see Section 2.4).

An entity MUST NOT mint tags under an authority name that was assigned to a different entity at 00:00 UTC on the given date, and it MUST NOT mint tags under a future date.

An entity that acquires an authority name immediately after a period during which the name was unassigned MAY mint tags as if the entity were 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 hp.com and may mint any tags rooted at that name with a current or past date when it held the registration. It must not mint tags, such as "tag:champignon.net,2001:", under domain names not registered to it. It must not mint tags dated in the future, such as

"tag:hp.com,2999:". If it obtains assignment of "extremelyunlikelytobeassigned.org" on 2001-05-01, then it must not mint tags under "extremelyunlikelytobeassigned.org,2001-04-01" unless it has evidence proving that name was continuously unassigned between 2001-04-01 and 2001-05-01.

A tagging entity mints specific identifiers that are unique within its context, in accordance with any internal scheme that uses only URI characters. Tagging entities SHOULD use record-keeping procedures to achieve uniqueness. Some tagging entities (e.g., corporations, mailing lists) consist of many people, in which case group decision-making SHOULD also be used to achieve uniqueness. The outcome of such decision-making could be to delegate control over parts of the namespace. For example, the assignees of example.com could delegate control over all tags with the prefixes "tag:example.com,2004:fred:" and "tag:example.com,2004:bill:", respectively, to the individuals with internal names "fred" and "bill" on 2004-01-01.

2.3. Resolution of Tags

There is no authoritative resolution mechanism for tags. Unlike most other URIs, tags can only be used as identifiers, and are not designed to support resolution. If authoritative resolution is a desired feature, a different URI scheme should be used.

2.4. Equality of Tags

Tags are simply strings of characters and are considered equal if and only if they are completely indistinguishable in their machine representations when using the same character encoding. That is, one can compare tags for equality by comparing the numeric codes of their characters, in sequence, for numeric equality. This criterion for equality allows for simplification of tag-handling software, which does not have to transform tags in any way to compare them.

3. Security Considerations

Minting a tag, by itself, is an operation internal to the tagging entity, and has no external consequences. The consequences of using an improperly minted tag (due to malice or error) in an application depends on the application, and must be considered in the design of any application that uses tags.

There is a significant possibility of minting errors by people who fail to apply the rules governing dates, or who use a shared (organizational) authority-name without prior organization-wide agreement. Tag-aware software MAY help catch and warn against these

errors. As stated in Section 2, however, to allow for future expansion, software MUST NOT reject tags which do not conform to the syntax specified in Section 2.

A malicious party could make it appear that the same domain name or email address was assigned to each of two or more entities. Tagging entities SHOULD use reputable assigning authorities and verify assignment wherever possible.

Entities SHOULD also avoid the potential for malicious exploitation of clock skew, by using authority names that were assigned continuously from well before to well after 00:00 UTC on the date chosen for the tagging entity -- preferably by intervals in the order of days.

4. IANA Considerations

The IANA has registered the tag URI scheme as specified in this document and summarised in the following template:

URI scheme name: tag

Status: permanent

URI scheme syntax: see Section 2

Character encoding considerations: percent-encoding is allowed in 'specific' and 'fragment' components (see Section 2)

Intended usage: see Section 1 and Section 2.3

Applications and/or protocols that use this URI scheme name: Any applications that use URIs as identifiers without requiring dereference, such as RDF, YAML, and Atom.

Interoperability considerations: none

Security considerations: see Section 3

Relevant publications: none

Contact: Tim Kindberg (timothy@hpl.hp.com) and Sandro Hawke (sandro@w3.org)

Author/Change controller: Tim Kindberg and Sandro Hawke

5. References

5.1. Normative References

[1] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

[2] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997.

[3] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

[4] "Data elements and interchange formats -- Information interchange -- Representation of dates and times", ISO (International Organization for Standardization) ISO 8601:1988, 1988.

5.2. Informative References

[5] Leach, P. and R. Salz, "UUIDs and GUIDs", Work in Progress, 1997.

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

[7] "Specification of abstract syntax notation one (ASN.1)", ITU-T recommendation X.208, (see also RFC 1778), 1988.

[8] Mealling, M., "A URN Namespace of Object Identifiers", RFC 3061, February 2001.

[9] Paskin, N., "Information Identifiers", Learned Publishing Vol. 10, No. 2, pp. 135-156, (see also www.doi.org), April 1997.

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

Authors' Addresses

Tim Kindberg
Hewlett-Packard Corporation
Hewlett-Packard Laboratories
Filton Road
Stoke Gifford
Bristol BS34 8QZ
UK

Phone: +44 117 312 9920
EMail: timothy@hpl.hp.com

Sandro Hawke
World Wide Web Consortium
32 Vassar Street
Building 32-G508
Cambridge, MA 02139
USA

Phone: +1 617 253-7288
EMail: sandro@w3.org

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.