Workgroup: Internet Engineering Task Force
Published: 9 February 2021
Intended Status: Informational
Expires: 13 August 2021
Authors: M. Čermák, Ed.

Description of Entities Identified by the 'tag' URI Scheme

Abstract

This document specifies automated description resolution mechanism for Uniform Resource Identifiers (URIs) in the "tag" scheme. Tag URIs (also known as "tags") are designed to be non-dereferenceable, however it may be useful for a tag minter to optinally provide a public easily accessible description of the entity associated with a tag.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on 13 August 2021.

Table of Contents

1. Introduction

The 'tag' Uniform Resource Identifier (URI) scheme is described by RFC 4151 [RFC4151]. URIs in this scheme (tags) are human-friendly identifiers, internally consisting of 2 parts: the tagging entity (a domain name or an e-mail address, followed by a date) and a specific identifier (optional). The combination chosen for the tagging entity guarantees uniqueness of tags across time and space, as long as no unauthorized entity mints tags under an entity with a different ownership.

Using tags as opaque identifiers in some cases is convenient for a couple of reasons: they are simple and easily rememberable by humans, the date component tells about the date of creation of the entity and thus carries significant information about its relevance, and the option of using e-mail addresses opens the possibility of minting new tags to virtually everyone.

Use of tags in other contexts, such as in semantic web applications, could be problematic. While dereferenceability is not a strict requirement in those cases, it might be advantageous to have a standardized mechanism for querying information about the entity referred to by a specific tag, since it already contains enough information to locate its tagging entity.

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 [RFC2119].

2. Querying Information about a Tag

A tag URI has the following structure, in ABNF (RFC 5234 [RFC5234]):

```
tagURI = "tag:" authorityName "," date ":" specific [ "#" fragment ]
```

The tag specification mandates that any software that processes tags MUST NOT reject authorities outside the original syntax. Therefore, in case of future additions to the specification, we will consider the following broader definition instead:

```
authorityName = ( web-host / addr-spec )
web-host = ( host / userinfo-host-post )
userinfo-host-port = [ userinfo "@" ] host ":" port
```

<userinfo>, <host> and <port> are specified in RFC 3986 [RFC3986]; <addr-spec> is specified in RFC 6068 [RFC6068]. As the querying mechanisms specified here differ substantially between those two forms, we will separate them into two groups of authorities, host-based and mail-based. Mechanisms for other types of authorities are not specified in this document.

Note that this document doesn't attempt to widen the definition of the tag authority from its original specification, but only provides a mechanism in case such an authority is used in practice. Due to the fact that a host-based authority with <userinfo> could be mistaken for an e-mail address, only authorities with a port can be considered. Thus a tag beginning with "tag:user@example.org:80," has a host-based authority, while a tag beginning with "tag:user@example.org," has a mail-based authority.

2.1. Host-based Authority and the .well-known "tag" URI

Tags with a host-based authority are mapped to a .well-known URI (RFC 8615 [RFC8615]) with the "tag" suffix. It is RECOMMENDED for a minter of such tags to use this service to publish a description of the entity identified by the tag, or to redirect to one. When an application attempts to dereference a host-based tag URI, it MAY attempt to use the .well-known URI created via the mapping instead.

For a tag in the form of "tag:web-host,date:specific#fragment", the corresponding HTTP(S) URL produced by the mapping is "http://web-host/.well-known/tag/specific#fragment". The client, as well as the server, MAY use "https" instead of "http", as well as additional HTTP mechanisms such as content negotation (RFC 7231 [RFC7231]), when the description is accessed.

There is no specific set of content types in which the description should be accessible, however it is RECOMMENDED to provide a description in at least "text/html" and either or both of "application/rdf+xml" and "text/turtle". If the communication ends in a success (code 2xx), the response body SHOULD contain the full tag URI in any position. The fragment portion of the URI, if

present, MAY be used to select the relevant portion of the description, in an application or content type-dependent manner.

For example, when a URI "tag:yaml.org,2002:int" is encountered by an application, it could attempt to dereference a URL "http://yaml.org/.well-known/tag/int". If such a page is available, its content could indicate that <tag:yaml.org,2002:int> is a datatype.

The date and fragment portions of the tag SHOULD NOT be a part of the HTTP request, thus it follows that the response body MAY describe all entities that differ only in the date and fragment, with the specific portion of the tag fixed. This is intentional, as the date portion of a tag serves only to anchor the ownership of the authority in time, while this mapping can only be used to query the present version of any site. It is however possible to use a mapping that uses the date portion and thus offers a higher level of verifiability, as described in the following section.

2.1.1. Obtaining the Archived Version of a Description

It is possible to devise an alternative mapping that incorporates a so-called archiving authority. This is dependent on the choice of such an authority, and ensures that the credibility of the description is no lesser than that of the archiving authority.

In the case that the tagging entity wishes to use an archiving authority to preserve the description of the tag, it SHOULD publish the description of the tag at the location as described above as close as possible to the time instant specified by its date portion, and save the URL via the archiving authority. This description MAY then be queried at its archived location instead of the present one. This ensures its credibility does not diminish over time.

There is no single specific mapping of this kind mandated by this document, but as an example, the Wayback Machine located at https://web.archive.org/ will be used to show a possible mapping. In this case, "tag:domain,date:specific" is saved by navigating to the URL "https://web.archive.org/save/http://domain/.well-known/tag/specific". Any client looking for a historical description of the entity can navigate to "https://web.archive.org/web/datetime/http://domain/.well-known/tag/specific", where <datetime> is the time instant represented by the full canonical date portion of the tag, in the yyyyMMddHHmmss format (ISO 8601 [ISO8601]).

In case the archiving service does not support content negotiation, it is RECOMMENDED to use "text/html" as the primary content type of the description, and embed other data into it (e.g. using HTML extensions such as RDFa).

2.2. Mail-based Authority

In order to resolve tags based on an e-mail address, the application needs to be able to create, send and receive e-mail messages (RFC 5322 [RFC5322]), i.e. act like a normal mail server (using SMTP (RFC 5321 [RFC5321])). The mapping is defined in terms of creating a URI in the "mailto:" scheme (RFC 6068 [RFC6068]), which the application MAY load and process as-is. If the application does not support loading "mailto:" URIs, it MAY compose the message directly, but the end result MUST be equivalent.

For a tag in the form of "tag:addr-spec,date:specific#fragment", the corresponding URI produced by the mapping is "mailto:addr-spec?subject=About%20tag%20%3Cspecific%3E". The message SHOULD include the relevant fields that allow the recipient of the message to reply to it, such as "From:" and "Message-ID:", and it MAY populate other fields or add a body (to add a human-friendly text in case the recipient is not a machine).

For a receiver of such a message, it is RECOMMENDED to reply with a description of the entity identified by the tag in question with the same content types of the individual parts of the reply as specified for host-based authorities, and at least one of the parts SHOULD contain the full tag URI in any position. The reply SHOULD have the "In-Reply-To:" field to identify the original request.

Due to the nature of e-mail messages, the application MUST be prepared for the case that a reply to the message will never be received, and, to comply with mail etiquette, it MUST NOT send a message asking for a description of a tag with the same authority and specific portion more than once when it is waiting for a reply or when the reply was already received and is still accessible.

When processing a reply, the application is free to interpret the contents of the message as it sees fit, and MAY use the fragment portion of the URI to select the relevant entities. It is RECOMMENDED not to make any difference between interpretations of the descriptions of host-based tags and mail-based tags, provided they are accepted by the application.

3. IANA Considerations

3.1. Assignment of .well-known 'tag' URI

The following assignment of a well-known URI is made, per RFC 8615 [RFC8615]:

URI suffix: tag

Change controller: IETF

Specification document(s): This document

Related information: None

4. Security Considerations

In addition to the security considerations of the underlying technologies, such as tags (RFC 4151 [RFC4151]), URIs (RFC 3986 [RFC3986]), and the "mailto:" scheme (RFC 6068 [RFC6068]), the most notable security consideration is the case when a specific domain or e-mail address starts providing malicious or errorneous description of a tag, possibly by violation of its own security or simply due to a change of owners. This is however something that linked data applications must already acknowledge and be ready to face for every usual dereferenced URL.

Due to the advantage of containing a date portion, tags offer higher security than standard URLs for identifying entities, as it is possible to use that information and compare certificates or "whois" records with historical data, or simply skip tags that are too "old". The possibility of using an archiving authority requires additional trust, but prevents attacks on the original site from affecting the application. Additionally, archive records that are not close enough to the date of the tag can also be ignored by the application.

Mail-based tags are not as secure as host-based tags, since the ownership of a particular e-mail address is usually completely governed by its provider. However, some providers do not allow re-registering an e-mail address or may implement other security measures, such as exposing the age of a particular mailbox. This knowledge could be used when estimating the credibility of the provided description, in addition to verifying the identity of the domain.

In case case, the application MAY refuse to query the description of a tag or ignore the result if it doesn't conform to its own security requirements.

5. References

5.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC

3986, DOI 10.17487/RFC3986, January 2005, <https://www.rfc-editor.org/info/rfc3986>.

[RFC4151] Kindberg, T. and S. Hawke, "The 'tag' URI Scheme", RFC 4151, DOI 10.17487/RFC4151, October 2005, <https://www.rfc-editor.org/info/rfc4151>.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <https://www.rfc-editor.org/info/rfc5234>.

[RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, DOI 10.17487/RFC5322, October 2008, <https://www.rfc-editor.org/info/rfc5322>.

[RFC6068] Duerst, M., Masinter, L., and J. Zawinski, "The 'mailto' URI Scheme", RFC 6068, DOI 10.17487/RFC6068, October 2010, <https://www.rfc-editor.org/info/rfc6068>.

[RFC8615] Nottingham, M., "Well-Known Uniform Resource Identifiers (URIs)", RFC 8615, DOI 10.17487/RFC8615, May 2019, <https://www.rfc-editor.org/info/rfc8615>.

5.2. Informative References

[ISO8601] ISO, "Data elements and interchange formats -- Information interchange -- Representation of dates and times", ISO 8601:1988, 1988.

[RFC5321] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, DOI 10.17487/RFC5321, October 2008, <https://www.rfc-editor.org/info/rfc5321>.

[RFC7231] Fielding, R., Ed. and J. Reschke, Ed., "Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content", RFC 7231, DOI 10.17487/RFC7231, June 2014, <https://www.rfc-editor.org/info/rfc7231>.

Author's Address

Marek Čermák (editor)

Email: standards@is4.site