Independent Submission
Request for Comments: 8574
Category: Informational

ISSN: 2070-1721

H. Van de Sompel
Data Archiving and Networked Services
M. Nelson
Old Dominion University
G. Bilder
Crossref
J. Kunze
California Digital Library
S. Warner
Cornell University
April 2019

cite-as: A Link Relation to Convey a Preferred URI for Referencing

Abstract

A web resource is routinely referenced by means of the URI with which it is directly accessed. But cases exist where referencing a resource by means of a different URI is preferred. This specification defines a link relation type that can be used to convey such a preference.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This is a contribution to the RFC Series, independently of any other RFC stream. The RFC Editor has chosen to publish this document at its discretion and makes no statement about its value for implementation or deployment. Documents approved for publication by the RFC Editor are not candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8574.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

1.	Introduction									3
2.	Terminology									3
3.	Scenarios									3
	.1. Persistent Identifiers .									3
	.2. Version Identifiers									5
	.3. Preferred Social Identifier									5
	.4. Multi-resource Publications									6
	The "cite-as" Relation Type for									
	for the Purpose of Referencing	_		_						6
5.	Distinction with Other Relation									8
	.1. The "bookmark" Relation Typ									9
	.2. The "canonical" Relation Ty									9
6.	Examples									10
	.1. Persistent HTTP URI									11
	.2. Version URIs									12
	.3. Preferred Profile URI									13
	.4. Multi-resource Publication									13
	IANA Considerations									14
8.	Security Considerations									14
9.	References									15
9	.1. Normative References									15
9	.2. Informative References .									15
	nowledgements									17
	nors' Addresses									17

1. Introduction

A web resource is routinely referenced (e.g., linked or bookmarked) by means of the URI with which it is directly accessed. But cases exist where referencing a resource by means of a different URI is preferred, for example, because the latter URI is intended to be more persistent over time. Currently, there is no link relation type to convey such an alternative referencing preference; this specification addresses this deficit by introducing a link relation type intended for that purpose.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

This specification uses the terms "link context" and "link target" as defined in [RFC8288]. These terms correspond with "Context IRI" and "Target IRI", respectively, as used in [RFC5988]. Although defined as IRIs (Internationalized Resource Identifiers), they are also URIs in common scenarios.

Additionally, this specification uses the following terms:

- o "access URI": A URI at which a user agent accesses a web resource.
- o "reference URI": A URI, other than the access URI, that should preferentially be used for referencing.

By interacting with the access URI, the user agent may discover typed links. For such links, the access URI is the link context.

3. Scenarios

3.1. Persistent Identifiers

Despite sound advice regarding the design of Cool URIS [CoolURIS], link rot ("HTTP 404 Not Found") is a common phenomena when following links on the Web. Certain communities of practice (see examples below) have introduced solutions to combat this problem. These solutions typically consist of:

o Accepting the reality that the web location of a resource -- the access URI -- may change over time.

- o Minting an additional URI for the resource -- the reference URI -- that is specifically intended to remain persistent over time.
- o Redirecting (typically with "HTTP 301 Moved Permanently", "HTTP 302 Found", or "HTTP 303 See Other") from the reference URI to the access URI.
- o Committing, as a community of practice, to adjust that redirection whenever the access URI changes over time.

This approach is, for example, used by:

- o Scholarly publishers that use DOIs (Digital Object Identifiers) [DOIs] to identify articles and DOI URLs [DOI-URLs] as a means to keep cross-publisher article-to-article links operational, even when the journals in which the articles are published change hands from one publisher to another, for example, as a result of an acquisition.
- o Authors of controlled vocabularies that use PURLs (Persistent Uniform Resource Locators) [PURLs] for vocabulary terms to ensure that the URIs they assign to vocabulary terms remain stable even if management of the vocabulary is transferred to a new custodian.
- o A variety of organizations (including libraries, archives, and museums) that assign ARK (Archival Resource Key) URLs [ARK] to information objects in order to support long-term access.

In order for the investments in infrastructure involved in these approaches to pay off, and hence for links to effectively remain operational as intended, it is crucial that a resource be referenced by means of its reference URI. However, the access URI is where a user agent actually accesses the resource (e.g., it is the URI in the browser's address bar). As such, there is a considerable risk that the access URI instead of the reference URI is used for referencing [PIDs-must-be-used].

The link relation type defined in this document makes it possible for user agents to differentiate the reference URI from the access URI.

3.2. Version Identifiers

Resource versioning systems often use a naming approach whereby:

- o The most recent version of a resource is always available at the same, generic URI.
- o Each version of the resource -- including the most recent one -- has a distinct version URI.

For example, Wikipedia uses generic URIs of the form https://en.wikipedia.org/wiki/John_Doe and version URIs of the form https://en.wikipedia.org/w/index.php?title=John_Doe&oldid=776253882.

While the current version of a resource is accessed at the generic URI, some versioning systems adhere to a policy that favors linking and referencing a specific version URI. To express this using the terminology of Section 2, these policies intend that the generic URI is the access URI and that the version URI is the reference URI. These policies are informed by the understanding that the content at the generic URI is likely to evolve over time and that accurate links or references should lead to the content as it was at the time of referencing. To that end, Wikipedia's "Permanent link" and "Cite this page" functionalities promote the version URI, not the generic URI.

The link relation type defined in this document makes it possible for user agents to differentiate the version URI from the generic URI.

3.3. Preferred Social Identifier

A web user commonly has multiple profiles on the Web, for example, one per social network, a personal homepage, a professional homepage, a FOAF (Friend Of A Friend) profile [FOAF], etc. Each of these profiles is accessible at a distinct URI. But the user may have a preference for one of those profiles, for example, because it is most complete, kept up to date, or expected to be long lived. As an example, the first author of this document has, among others, the following profile URIs:

- o <https://hvdsomp.info>
- o <https://twitter.com/hvdsomp>
- o <https://www.linkedin.com/in/herbertvandesompel/>
- o <https://orcid.org/0000-0002-0715-6126>

Of these, from the perspective of the person described by these profiles, the first URI may be the preferred profile URI for the purpose of referencing because the domain is not under the custodianship of a third party. When an agent accesses another profile URI, such as https://orcid.org/0000-0002-0715-6126, this preference for referencing by means of the first URI could be expressed.

The link relation type defined in this specification makes it possible for user agents to differentiate the preferred profile URI from the accessed profile URI.

3.4. Multi-resource Publications

When publishing on the Web, it is not uncommon to make distinct components of a publication available as different web resources, each with their own URI. For example:

- o Contemporary scholarly publications routinely consists of a traditional article as well as additional materials that are considered an integral part of the publication such as supplementary information, high-resolution images, or a video recording of an experiment.
- o Scientific or governmental open data sets frequently consist of multiple files.
- o Online books typically consist of multiple chapters.

While each of these components is accessible at its distinct URI -- the access URI -- they often also share a URI assigned to the intellectual publication of which they are components -- the reference URI.

The link relation type defined in this document makes it possible for user agents to differentiate the URI of the intellectual publication from the access URI of a component of the publication.

4. The "cite-as" Link Relation Type for Expressing a Preferred URI for the Purpose of Referencing

A link with the "cite-as" relation type indicates that, for referencing the link context, use of the URI of the link target is preferred over use of the URI of the link context. It allows the resource identified by the access URI (link context) to unambiguously link to its corresponding reference URI (link target), thereby expressing that the link target is preferred over the link context for the purpose of permanent citation.

The link target of a "cite-as" link SHOULD support protocol-based access as a means to ensure that applications that store them can effectively reuse them for access.

The link target of a "cite-as" link SHOULD provide the ability for a user agent to follow its nose back to the context of the link, e.g., by following redirects and/or links. This helps a user agent to establish trust in the target URI.

Because a link with the "cite-as" relation type expresses a preferred URI for the purpose of referencing, the access URI SHOULD only provide one link with that relation type. If more than one "cite-as" link is provided, the user agent may decide to select one (e.g., an HTTP URI over a mailto URI) based on the purpose that the reference URI will serve.

Providing a link with the "cite-as" relation type does not prevent using the access URI for the purpose of referencing if such specificity is needed for the application at hand. For example, in the case of the scenario in Section 3.4, the access URI is likely required for the purpose of annotating a specific component of an intellectual publication. Yet, the annotation application may also want to appropriately include the reference URI in the annotation.

Applications can leverage the information provided by a "cite-as" link in a variety of ways, for example:

- o Bookmarking tools and citation managers can take this preference into account when recording a URI.
- o Webometrics applications that trace URIs can trace both the access URI and the reference URI.
- o Discovery tools can support lookup by means of both the access and the reference URI. This includes web archives that typically make archived versions of web resources discoverable by means of the original access URI of the archived resource; they can additionally make these archived resources discoverable by means of the associated reference URI.

5. Distinction with Other Link Relation Types

Some existing IANA-registered relationships intuitively resemble the relationship that "cite-as" is intended to convey. But a closer inspection of these candidates provided in the blog posts [identifier-blog], [canonical-blog], and [bookmark-blog] shows that they are not appropriate for various reasons and that a new link relation type is required. The remainder of this section provides a summary of the detailed explanations provided in the referenced blog posts.

It can readily be seen that the following link relation types do not address the requirements described in Section 3:

- o "alternate" [RFC4287]: The link target provides an alternate version of the content at the link context. These are typically variants according to dimensions that are subject to content negotiation, for example, the same content with varying Content-Type (e.g., application/pdf vs. text/html) and/or Content-Language (e.g., en vs. fr). The representations provided by the context URIs and target URIs in the scenarios in Sections 3.1 through 3.4 are not variants in the sense intended by [RFC4287], and, as such, the use of "alternate" is not appropriate.
- o "duplicate" [RFC6249]: The link target is a resource whose available representations are byte-for-byte identical with the corresponding representations of the link context, for example, an identical file on a mirror site. In none of the scenarios described in Sections 3.1 through 3.4 do the link context and the link target provide identical content. As such, the use of "duplicate" is not appropriate.
- o "related" [RFC4287]: The link target is a resource that is related to the link context. While "related" could be used in all of the scenarios described in Sections 3.1 through 3.4, its semantics are too vague to convey the specific semantics intended by "cite-as".

Two existing IANA-registered relationships deserve closer attention and are discussed in the remainder of this section.

5.1. The "bookmark" Link Relation Type

"bookmark" [W3C.REC-html52-20171214]: The link target provides a URI for the purpose of bookmarking the link context.

The intent of "bookmark" is closest to that of "cite-as" in that the link target is intended to be a permalink for the link context, for bookmarking purposes. The link relation type dates back to the earliest days of news syndication, before blogs and news feeds had permalinks to identify individual resources that were aggregated into a single page. As such, its intent is to provide permalinks for different sections of an HTML document. It was originally used with HTML elements such as <div>, <h1>, <h2>, etc.; more recently, HTML5 revised it to be exclusively used with the <article> element.

Moreover, it is explicitly excluded from use in the link> element in HTML <head> and, as a consequence, in the HTTP Link header that is semantically equivalent. For these technical and semantic reasons, the use of "bookmark" to convey the relationship intended by "cite-as" is not appropriate.

A more detailed justification regarding the inappropriateness of "bookmark", including a thorough overview of its turbulent history, is provided in [bookmark-blog].

5.2. The "canonical" Link Relation Type

"canonical" [RFC6596]: The meaning of "canonical" is commonly misunderstood on the basis of its brief definition as being "the preferred version of a resource." The description in the abstract of [RFC6596] is more helpful and states that "canonical" is intended to link to a resource that is preferred over resources with duplicative content. A more detailed reading of [RFC6596] clarifies that the intended meaning is that "canonical" is preferred for the purpose of content indexing. A typical use case is linking from each page in a multi-page magazine article to a single page version of the article provided for indexing by search engines: the former pages provide content that is duplicative to the superset content that is available at the latter page.

The semantics intended by "canonical" as preferred for the purpose of content indexing differ from the semantics intended by "cite-as" as preferred for the purpose of referencing. A further exploration of the various scenarios shows that the use of "canonical" is not appropriate to convey the semantics intended by "cite-as":

- o Scenario of Section 3.1: The reference URI that is intended to be persistent over time does not serve content that needs to be indexed; it merely redirects to the access URI. Since the meaning intended by "canonical" is that it is preferred for the purpose of content indexing, it is not appropriate to point at the reference URI (persistent identifier) using the "canonical" link relation type. Moreover, Section 6.1 shows that scholarly publishers that assign persistent identifiers already use the "canonical" link relation type for search engine optimization; it also shows how that use contrasts with the intended use of "cite-as".
- o Scenario of Section 3.2: In most common cases, custodians of resource versioning systems want search engines to index the most recent version of a page and hence would use a "canonical" link to point from version URIs of a resource to the associated generic URI. Wikipedia effectively does this. However, for some resource versioning systems, including Wikipedia, version URIs are preferred for the purpose of referencing. As such, a "cite-as" link would point from the generic URI to the most recent version URI (that is, in the opposite direction of the "canonical" link).
- o Scenario of Section 3.3: The content at the link target and the link context are different profiles for a same person. Each profile, not just a preferred one, should be indexed. But a single one could be preferred for referencing.
- o Scenario of Section 3.4: The content at the link target, if any, would typically be a landing page that includes descriptive metadata pertaining to the multi-resource publication and links to its component resources. Each component resource provides content that is different, not duplicative, to the landing page.

A more detailed justification regarding how the use of "canonical" is inappropriate to address the requirements described in this document, including examples, is provided in [canonical-blog].

6. Examples

Sections 6.1 through 6.4 show examples of the use of links with the "cite-as" relation type. They illustrate how the typed links can be used in a response header and/or response body.

6.1. Persistent HTTP URI

PLOS ONE is one of many scholarly publishers that assigns DOIs to the articles it publishes. For example, https://doi.org/10.1371/journal.pone.0171057 is the persistent identifier for such an article. Via the DOI resolver, this persistent identifier redirects to https://journals.plos.org/plosone/doi?id=10.1371/journal.pone.0171057, which delivers the actual article in HTML.

The HTML article contains a <link> element with the "canonical" link relation type pointing at itself, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167475>. As per Section 5.2, this indicates that the article content at that URI should be indexed by search engines.

PLOS ONE can additionally provide a link with the "cite-as" relation type pointing at the persistent identifier to indicate it is the preferred URI for permanent citation of the article. Figure 1 shows the addition of the "cite-as" link in both the HTTP header and the HTML that results from an HTTP GET on the article URI https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167475.

Figure 1: Response to HTTP GET on the URI of a Scholarly Article

6.2. Version URIs

The preprint server arXiv.org has a versioning approach like the one described in Section 3.2:

- o The most recent version of a preprint is always available at the same, generic URI. Consider the preprint with generic URI https://arxiv.org/abs/1711.03787.
- o Each version of the preprint -- including the most recent one -- has a distinct version URI. The considered preprint has two versions with respective version URIs: https://arxiv.org/abs/1711.03787v1 (published 10 November 2017) and https://arxiv.org/abs/1711.03787v2 (published 24 January 2018).

A reader who accessed https://arxiv.org/abs/1711.03787 between 10 November 2017 and 23 January 2018, obtained the first version of the preprint. Starting 24 January 2018, the second version was served at that URI. In order to support accurate referencing, arXiv.org could implement the "cite-as" link to point from the generic URI to the most recent version URI. In doing so, assuming the existence of reference manager tools that consume "cite-as" links:

- o The reader who accesses https://arxiv.org/abs/1711.03787> between 10 November 2017 and 23 January 2018 would reference https://arxiv.org/abs/1711.03787v1.
- o The reader who accesses https://arxiv.org/abs/1711.03787 starting 24 January 2018 would reference https://arxiv.org/abs/1711.03787v2.

Figure 2 shows the header that arXiv.org would have returned in the first case, in response to a HTTP HEAD on the generic URI https://arxiv.org/abs/1711.03787.

```
HTTP/1.1 200 OK
```

Date: Sun, 24 Dec 2017 16:12:43 GMT
Content-Type: text/html; charset=utf-8
Link: <https://arxiv.org/abs/1711.03787v1> ; rel="cite-as"

Vary: Accept-Encoding, User-Agent

Figure 2: Response to HTTP HEAD on the Generic URI of the Landing Page of an arXiv.org Preprint

6.3. Preferred Profile URI

If the access URI is the home page of John Doe, John can add a link with the "cite-as" relation type to it, in order to convey that he would prefer to be referenced by means of the URI of his FOAF profile. Figure 3 shows the response to an HTTP GET on the URI of John's home page.

Figure 3: Response to HTTP GET on the URI of John Doe's Home Page

6.4. Multi-resource Publication

The Dryad Digital Repository at datadryad.org specializes in hosting and preserving scientific datasets. Each dataset typically consists of multiple resources. For example, the dataset "Data from: Climate, demography, and lek stability in an Amazonian bird" consists of an Excel spreadsheet, a csv file, and a zip file. Each of these resources have different content and are accessible at their respective URIs. In addition, the dataset has a landing page at https://datadryad.org/resource/doi:10.5061/dryad.5d23f.

Each of these resources should be permanently cited by means of the persistent identifier that was assigned to the entire dataset as an intellectual publication, i.e., https://doi.org/10.5061/dryad.5d23f. To that end, the Dryad Digital Repository can add "cite-as" links pointing from the URIs of each of these resources to https://doi.org/10.5061/dryad.5d23f. This is shown in Figure 4 for the csv file that is a component resource of the dataset, through use of the HTTP Link header.

HTTP/1.1 200 OK

Date: Tue, 12 Jun 2018 19:19:22 GMT

Last-Modified: Wed, 17 Feb 2016 18:37:02 GMT Content-Type: text/csv; charset=ISO-8859-1

Content-Length: 25414

Link: https://doi.org/10.5061/dryad.5d23f; rel="cite-as"

DATE, Year, PLOT/TRAIL, LOCATION, SPECIES CODE, BAND NUM, COLOR, SEX, AGE, TAIL, WING, TARSUS, NARES, DEPTH, WIDTH, WEIGHT 6/26/02, 2002, DANTA, 325, PIPFIL, 969, B/O, M, AHY, 80, 63, 16, 7.3, 3.9, 4.1, 14.4 ... 2/3/13, 2013, LAGO, PIPFIL, BR-5095, O/YPI, M, SCB, 78, 65.5, 14.2, 7.5, 3.8,

Figure 4: Response to HTTP GET on the URI of a csv File That Is a Component of a Scientific Dataset

7. IANA Considerations

3.7,14.3

The link relation type has been registered by IANA per Section 2.1.1 of [RFC8288] as follows:

Relation Name: cite-as

Description: Indicates that the link target is preferred over the link context for the purpose of permanent citation.

Reference: RFC 8574

8. Security Considerations

In cases where there is no way for the agent to automatically verify the correctness of the reference URI (cf. Section 4), out-of-band mechanisms might be required to establish trust.

If a trusted site is compromised, the "cite-as" link relation could be used with malicious intent to supply misleading URIs for referencing. Use of these links might direct user agents to an attacker's site, break the referencing record they are intended to support, or corrupt algorithmic interpretation of referencing data.

9. References

9.1. 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.

- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,
 May 2017, https://www.rfc-editor.org/info/rfc8174.
- [W3C.REC-html52-20171214]
 Faulkner, S., Eicholz, A., Leithead, T., Danilo, A., and
 S. Moon, "HTML 5.2", World Wide Web
 Consortium Recommendation REC-html52-20171214, December
 2017, https://www.w3.org/TR/2017/REC-html52-20171214/.

9.2. Informative References

[ARK] Kunze, J. and R. Rodgers, "The ARK Identifier Scheme", Work in Progress, draft-kunze-ark-18, April 2013.

[bookmark-blog]

Nelson, M. and H. Van de Sompel, "rel=bookmark also does not mean what you think it means", August 2017, http://ws-dl.blogspot.com/2017/08/2017-08-26-relbookmark-also-does-not.html.

[canonical-blog]

Nelson, M. and H. Van de Sompel, "rel=canonical does not
mean what you think it means", August 2017,
<http://ws-dl.blogspot.nl/2017/08/
2017-08-07-relcanonical-does-not-mean.html>.

[CoolURIs]

Berners-Lee, T., "Cool URIs don't change", World Wide Web
Consortium Style, 1998,
<https://www.w3.org/Provider/Style/URI.html>.

[DOI-URLs]

Hendricks, G., "Display guidelines for Crossref DOIs",
March 2017,
<https://blog.crossref.org/display-guidelines/>.

[FOAF] Brickley, D. and L. Miller, "FOAF Vocabulary Specification 0.99", January 2014, http://xmlns.com/foaf/spec/.

[identifier-blog]

Nelson, M. and H. Van de Sompel, "Linking to Persistent
Identifiers with rel=identifier", November 2016,
<http://ws-dl.blogspot.com/2016/11/
2016-11-07-linking-to-persistent.html>.

[PIDs-must-be-used]

Van de Sompel, H., Klein, M., and S. Jones, "Persistent URIs Must Be Used To Be Persistent", February 2016, https://arxiv.org/abs/1602.09102.

[PURLs] Wikipedia, "Persistent uniform resource locator", September 2018, https://en.wikipedia.org/w/index.php?title=Persistent_uniform_resource_locator&oldid=858558072>.

Acknowledgements

The authors would like to thank the following individuals for their comments and suggestions: Martin Klein, Harihar Shankar, Peter Williams, John Howard, Mark Nottingham, and Graham Klyne.

Authors' Addresses

Herbert Van de Sompel Data Archiving and Networked Services

Email: herbert.van.de.sompel@dans.knaw.nl URI: https://orcid.org/0000-0002-0715-6126

Michael Nelson Old Dominion University

Email: mln@cs.odu.edu

URI: http://www.cs.odu.edu/~mln/

Geoffrey Bilder Crossref

Email: gbilder@crossref.org

URI: https://www.crossref.org/authors/geoffrey-bilder/

John Kunze California Digital Library

Email: jak@ucop.edu

URI: https://orcid.org/0000-0001-7604-8041

Simeon Warner Cornell University

Email: simeon.warner@cornell.edu

URI: https://orcid.org/0000-0002-7970-7855