Network Working Group   H. Van de Sompel
Internet-Draft   Los Alamos National Laboratory
Intended status: Informational   M. Nelson
Expires: February 3, 2018   Old Dominion University
                                                       G. Bilder
                                                        Crossref
                                                        J. Kunze
                                      California Digital Library
                                                       S. Warner
                                             Cornell University
                                                 August 2, 2017

   Identifier: A Link Relation to Convey a Preferred URI for Referencing
                    draft-vandesompel-identifier-00

Abstract

   This specification defines a link relation type that is intended to
   convey that a URI, other than the URI that provides a link with the
   relation type, is preferred for the purpose of referencing.

Note to Readers

   Please discuss this draft on the ART mailing list
   (<https://www.ietf.org/mailman/listinfo/art>).

Table of Contents

[1](1). Introduction

   A web resource is routinely referenced (e.g. linked, bookmarked) by
   means of the URI where 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 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", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in RFC 2119 [RFC2119].

This specification uses the terms "link context" and "link target" as
defined in [I-D.nottingham-rfc5988bis]. These terms respectively
correspond with "Context IRI" and "Target IRI" as used in [RFC5988].
Although defined as IRIs, in common scenarios they are also URIs.

Additionally, this specification uses the following terms:

o   "access URI": A URI at which a user agent accesses a web resource.

o   "identifying 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 have introduced
solutions to combat this problem that 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 identifying URI -
    that is specifically intended to remain persistent over time.

o   Redirecting (typically "HTTP 301 Moved Permanently", "HTTP 302
    Found", or "HTTP 303 See Other") from the identifying URI to the
    access URI.

o   As a community, committing to adjust that redirection whenever the
    access URI changes over time.

This approach is, for example, used by:

o Scholarly publishers that use DOIs [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 [PURLs] for
   vocabulary terms to ensure that the term URIs remain stable even
   if management of the vocabulary is transfered to a new custodian.

o A variety of organizations, including libraries, archives, and
   museums that assign ARK URLs [draft-kunze-ark-18] 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 identifying 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 identifying URI is used for referencing
[PIDs-must-be-used].

The link relation type defined in this specification allows to convey
to user agents that the identifying URI is the preferred URI for
referencing. Applications such as bookmarking tools, citation
managers, and webometrics applications can take this preference into
account when recording a URI.

3.2. Version Identifiers

Resource versioning systems often use a naming approach whereby:

o the most recent version of a resource is at any time 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
<http://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 by means of the version URI that was minted for the
current version. 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 identifying 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 use the version URI, not the generic URI.

The link relation type defined in this specification allows to convey
to user agents that the version URI is preferred over the generic URI
for referencing.

[3.3](#). Preferred Social Identifier

A web user commonly has multiple profiles on the web, for example,
one per social network she takes part in, a personal homepage, a
professional homepage, a FOAF 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.

The link relation type defined in this specification allows to convey
to user agents that a profile URI - the identifying URI - other than
the one the agent is accessing - the access URI - is preferred for
referencing.

[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 its 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, 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 are accessible at their distinct URI -
the access URI - they often also share a URI assigned to the
intellectual publication of which they are components - the
identifying URI.

The link relation type defined in this specification allows to convey to user agents that, for the purpose of referencing, the identifying URI of the intellectual publication is preferred over an access URI of a component of the publication.

4. **The "identifier" Relation Type for Expressing a Preferred URI for the Purpose of Referencing**

A link with the "identifier" relation type indicates that the link target - the identifying URI - is preferred over the link context for the purpose of referencing.

An identifying URI SHOULD support protocol-based access as a means to ensure that applications that store identifying URIs can effectively re-use them for access.

An identifying URI SHOULD provide the ability for a user agent to follow its nose back to the access URI, e.g. by following redirects and/or links. This helps a user agent to establish trust in the identifying URI.

Because a link with the "identifier" 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 "identifier" link is provided, the user agent may decide to select one (e.g. an HTTP URI over a mailto URI), for example, based on the purpose that the identifying URI will serve.

Providing a link with the "identifier" 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 scenario 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 identifying URI in the annotation.

5. **Distinction with Other Relation Types**

The following existing IANA-registered relationships are similar to the relationship that "identifier" is intended to convey, but are not appropriate for various reasons:

o "alternate" [RFC4287], used to link to an alternate version of the content at the link context, 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).

o   "bookmark" [W3C.REC-html5-20151028], used to convey a permanent
    link to use for bookmarking purposes.

o   "canonical" [RFC6596], used to identify content that is either
    duplicative or a superset of the content at the link context, for
    example a single page version of a magazine article, provided for
    indexing by search engines, of an article that is spread over
    several pages for human use.

o   "duplicate" [RFC6249], used to link to 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.

o   "related" [RFC4287], used to link to a related resource.

A closer inspection of these candidates [identifier-blog] shows that
they are not appropriate and that a new relation type is required.

In the scenario of Section 3.1 there is no content available at the
identifying URI as it merely redirects to the access URI. In the
scenario of Section 3.3, the content at the identifying URI is a
profile that is different than the profile at the access URI. In the
scenario of Section 3.4 the content at the identifying URI, if any,
would typically be a sort of table of contents with links to
component resources and possibly a summary. These considerations
exclude "alternate", "canonical", and "duplicate" as possible
relation types.

The intent of "bookmark" is closest to that of "identifier" in that
the link target of a link with this relation type is intended for
bookmarking, which is a case of referencing. However, "bookmark" is
specifically defined for use in conjunction with the HTML <article>
element and is explictly excluded from use in the <link> element in
HTML <head>. Since a link in <link> and a link in the HTTP Link
header are semantically equivalent, "bookmark" is also excluded from
use in HTTP Link.

While "related" could be used, its semantics are too vague to convey
the specific nature of "identifier" as a means to convey a URI for
the purpose of referencing.

6. Examples

Sections Section 6.1 and Section 6.2 show examples of the use of
links with the "identifier" relation type. One example shows its use
in a response header and body, the other in a response body only.

6.1. **Persistent HTTP URI**

   If the access URI is a landing page for a scholarly article for which
   the persistent HTTP URI <http://persistence.example.org/738207472>
   was minted, then the response to an HTTP GET on the landing page's
   URI could be as shown in Figure 1.

```
HTTP/1.1 200 OK
Link: <http://persistence.example.org/738207472> ; rel="identifier"
Content-Type: text/html;charset=utf-8

<html>
 <head>
 ...
  <link rel="identifier" href="http://persistence.example.org/738207472" />
 ...
 </head>
 <body>
  ...
 </body>
</html>
```

       Figure 1: Response to HTTP GET on the URI of the landing page of a
                            scholarly article

6.2. **Preferred Profile URI**

   If the access URI is the home page of John Doe, John can add a link
   with the "identifier" relation type to it, as a means to convey that
   he would preferably be referenced by means of the URI of his FOAF
   profile. Figure 2 shows the response to an HTTP GET on the URI of
   John's home page.

```
HTTP/1.1 200 OK
Content-Type: text/html;charset=utf-8

<html>
 <head>
 ...
   <link rel="identifier" href="http://johndoe.example.com/foaf" type="text/
ttl"/>
 ...
 </head>
 <body>
  ...
 </body>
</html>
```

        Figure 2: Response to HTTP GET on the URI of John Doe's home page

7. IANA Considerations

7.1. Link Relation Type: identifier

   The link relation type below has been registered by IANA per
   Section 2.1.1 of [I-D.nottingham-rfc5988bis]:

      Relation Name: identifier

      Description: A link with the "identifier" relation type indicates
      that the link target is preferred over the link context for the
      purpose of referencing.

      Reference: [[This document]]

8. Security Considerations

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

   If a trusted site is compromised, the "identifier" 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

   [I-D.nottingham-rfc5988bis]
              Nottingham, M., "Web Linking", draft-nottingham-
              rfc5988bis-06 (work in progress), June 2017.

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

   [RFC4287] Nottingham, M., Ed. and R. Sayre, Ed., "The Atom
              Syndication Format", RFC 4287, DOI 10.17487/RFC4287,
              December 2005, <http://www.rfc-editor.org/info/rfc4287>.

   [RFC5988] Nottingham, M., "Web Linking", RFC 5988,
              DOI 10.17487/RFC5988, October 2010,
              <http://www.rfc-editor.org/info/rfc5988>.

   [RFC6249] Bryan, A., McNab, N., Tsujikawa, T., Poeml, P., and H.
              Nordstrom, "Metalink/HTTP: Mirrors and Hashes", RFC 6249,
              DOI 10.17487/RFC6249, June 2011,
              <http://www.rfc-editor.org/info/rfc6249>.

   [RFC6596] Ohye, M. and J. Kupke, "The Canonical Link Relation",
              RFC 6596, DOI 10.17487/RFC6596, April 2012,
              <http://www.rfc-editor.org/info/rfc6596>.

   [W3C.REC-html5-20151028]
              Hickson, I., Berjon, R., Faulkner, S., Leithead, T., Doyle
              Navara, E., O'Connor, E., and S. Pfeiffer, "HTML5", World
              Wide Web Consortium Recommendation REC-HTML5-20141028,
              October 2014, <https://www.w3.org/TR/2014/REC-
              html5-20141028/>.

9.2. Informative References

   [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",
              June 2017, <https://blog.crossref.org/display-
              guidelines/>.

[DOIs]   "Information and documentation - Digital object identifier
            system", ISO 26324:2012(en), 2012,
            <https://www.iso.org/obp/ui/#iso:std:iso:26324:ed-
            1:v1:en>.

[draft-kunze-ark-18]
            Kunze, J. and R. Rodgers, "The ARK Identifier Scheme",
            Internet Draft draft-kunze-ark-18, April 2013,
            <https://datatracker.ietf.org/doc/html/draft-kunze-ark>.

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

[identifier-blog]
            Nelson, M., "Linking to Persistent Identifiers with
            rel="identifier"", July 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]   "Persistent uniform resource locator", April 2017,
            <https://en.wikipedia.org/wiki/
            Persistent_uniform_resource_locator>.

Appendix A. Acknowledgements

Authors' Addresses

Herbert Van de Sompel
Los Alamos National Laboratory

Email: herbertv@lanl.gov
URI: http://public.lanl.gov/herbertv/

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: http://www.cdlib.org/contact/staff_directory/jkunze.html

Simeon Warner
Cornell University

Email: simeon.warner@cornell.edu
URI: https://orcid.org/0000-0002-7970-7855