The "info" URI Scheme
for Information Assets with Identifiers in Public Namespaces

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

Copyright (C) The Internet Society (2006).

Abstract

This document defines the "info" Uniform Resource Identifier (URI) scheme for information assets with identifiers in public namespaces. Namespaces participating in the "info" URI scheme are regulated by an "info" Registry mechanism.
Table of Contents

1. Introduction .................................................... 3
   1.1. Terminology ................................................ 3
   1.2. Information Assets ........................................ 3
2. Application of the "info" URI Scheme ............................ 4
3. The "info" Registry ............................................ 5
   3.1. Management Characteristics of the "info" Registry ......... 5
   3.2. Functional Characteristics of the "info" Registry ......... 5
   3.3. Maintenance of the "info" Registry ........................ 6
4. The "info" URI Scheme ........................................... 6
   4.1. Definition of "info" URI Syntax ............................ 6
   4.2. Allowed Characters Under the "info" URI Scheme .......... 8
   4.3. Examples of "info" URIs ................................ 9
5. Normalization and Comparison of "info" URIs .................... 10
6. Rationale ................................................... 12
   6.1. Why Create a New URI Scheme for Identifiers from Public Namespaces? ........................................ 12
   6.2. Why Not Use an Existing URI Scheme for Identifiers from Public Namespaces? ................................. 12
   6.3. Why Not Create a New URN Namespace ID for Identifiers from Public Namespaces? ................................. 12
7. Security Considerations ........................................ 13
8. IANA Considerations ............................................ 14
9. Acknowledgements ............................................... 14
10. References .................................................. 14
   10.1. Normative References ................................... 14
   10.2. Informative References ................................. 15
1. Introduction

This document defines the "info" Uniform Resource Identifier (URI) scheme for information assets that have identifiers in public namespaces but are not part of the URI allocation. By "information asset" this document intends any information construct that has identity within a public namespace.

1.1. Terminology

In this document, the keywords "MUST", "MUST NOT", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "MAY", "MAY NOT", and "RECOMMENDED" are to be interpreted as described in RFC 2119 [RFC2119] and indicate requirement levels for compliant implementations.

1.2. Information Assets

There exist many information assets with identifiers in public namespaces that are not referenceable by URI schemes. Examples of such namespaces include Dewey Decimal Classifications [DEWEY], Library of Congress Control Numbers [LCCN], NISO Serial Item and Contribution Identifiers [SICI], NASA Astrophysics Data System Bibcodes [BIBCODE], and National Library of Medicine PubMed identifiers [PMID]. Other candidate namespaces include Online Computer Library Center OCLC Numbers [OCLCNUM] and NISO OpenURL Framework identifiers [OFI].

The "info" URI scheme facilitates the referencing of information assets that have identifiers in such public namespaces by means of URIs. When referencing an information asset by means of its "info" URI, the asset SHALL be considered a "resource" as defined in RFC 3986 [RFC3986] and SHALL enjoy the same common syntactic, semantic, and shared language benefits that the URI presentation confers. As such, the "info" URI scheme enables public namespaces that are not part of the URI allocation to be represented within the allocation. The "info" URI scheme thus provides a bridging mechanism to allow public namespaces to become part of the URI allocation.

Namespaces declared under the "info" URI scheme are regulated by an "info" Registry mechanism. The "info" Registry allows a public namespace that is not part of the URI allocation to be declared in a registration process by the organization that manages it (the Namespace Authority). The "info" Registry supports the declaration of public namespaces that are not part of the URI allocation in a manner that facilitates the construction of URIs for information assets without imposing the burdens of independent URI registration and maintenance of resource representations on the Namespace Authority. Information assets identified within a registered...
namespace SHALL be added or deleted according to the business processes of the Namespace Authority, and yet MAY be referenced within network applications via the "info" URI in an open, standardized way without additional action on the part of the Namespace Authority.

The "info" URI scheme exists primarily for identification purposes. Implementations MUST NOT assume that an "info" URI can be dereferenced to a representation of the resource identified by the URI although Namespace Authorities MAY disclose in the registration record references to service mechanisms pertaining to identifiers from the registered namespace. Applications of the "info" URI scheme are restricted to the identification of information assets and the declaration of normalization rules for comparing identifiers of such information assets regardless of whether any services relating to such information assets are accessible via the Internet. References to such services MAY be disclosed within an "info" registration record, but these services SHALL NOT be regarded as authoritative. The "info" URI scheme does not support global resolution methods.

2. Application of the "info" URI Scheme

Public namespaces that are used for the identification of information assets and that are not part of the URI allocation MAY be registered as namespaces within the "info" Registry. Namespace Authorities MAY register these namespaces in the "info" Registry, thereby making these namespaces available to applications that need to reference information assets by means of a URI. Registrations of public namespaces that are not part of the URI allocation by parties other than the Namespace Authority SHALL NOT be permitted, thereby ensuring against hostile usurpation or inappropriate usage of registered service marks or the public namespaces of others.

Registration of a public namespace under the "info" Registry implies no particular functionalities of the identifiers from the registered namespace other than the identification of information assets. No resolution mechanisms can be assumed for the "info" URI scheme, though for any particular namespace there MAY exist mechanisms for resolving identifiers to network services. The definition of such services falls outside the scope of the "info" URI scheme. Registration does not define namespace-specific semantics for identifiers within a registered namespace, though allowable character sets and normalization rules are specified in Sections 4 and 5 so as to ensure that the URIs created using such identifiers are compliant with applications that use URIs.
The registration of a public namespace in the "info" Registry SHALL NOT preclude further development of services associated with that namespace that MAY qualify the namespace for additional publication elsewhere within the URI allocation.

3. The "info" Registry

The "info" Registry provides a mechanism for the registration of public namespaces that are used for the identification of information assets and that are not part of the URI allocation.

NISO [NISO], the National Information Standards Organization, will act as the Maintenance Agency for the "info" Registry and will delegate the day-to-day operation of the "info" Registry to a Registry Operator. As the Maintenance Agency, NISO will ensure that the Registry Operator operates the "info" Registry in accordance with a publicly articulated policy document established under NISO governance and made available on the "info" website, <http://info-uri.info/>. The "info" Registry policy defines a review process for candidate namespaces and provides measures of quality control and suitability for entry of namespaces.

3.1. Management Characteristics of the "info" Registry

The "info" Registry will be managed according to policies established under the auspices of NISO. All such policies, as well as the namespace declarations in the "info" Registry, will be public.

3.2. Functional Characteristics of the "info" Registry

The "info" Registry will be publicly accessible and will support discovery (by both humans and machines) of:

- string literals identifying the namespaces for which the Registry provides a guarantee of uniqueness and persistence
- names and contact information of Namespace Authorities
- syntax requirements for identifiers maintained in such namespaces
- normalization methodologies for identifiers maintained in such namespaces
- network references to a description of service mechanisms (if any) for identifiers maintained in such namespaces
- ancillary documentation

Registry entries refer to the corresponding "namespace" and "identifier" components, which are defined in the ABNF given in Section 4.1 of this document.
3.3. Maintenance of the "info" Registry

The public namespaces that MAY be registered in the "info" Registry will be those of interest to the communities served by NISO, and therefore NISO is committed to act as Maintenance Authority for the "info" Registry and to assign a Registry Operator to operate it.

NISO, a non-profit association accredited by the American National Standards Institute (ANSI), identifies, develops, maintains, and publishes technical standards to manage information in the digital environment. NISO standards apply technologies to the full range of information-related needs, including retrieval, re-purposing, storage, metadata, and preservation.

Founded in 1939, incorporated as a not-for-profit education association in 1983, and assuming its current name the following year, NISO draws its support from the communities it serves. The leaders of over 70 organizations in the fields of publishing, libraries, IT, and media serve as its voting members. Hundreds of experts and practitioners serve on NISO committees and as officers of the association.

NISO has been designated by ANSI to represent US interests to the International Organization for Standardization’s (ISO) Technical Committee 46 on Information and Documentation.

The NISO headquarters office is located at 4733 Bethesda Ave., Bethesda, MD 20814, USA. (For further information, see the NISO website, <http://www.niso.org/>.)

4. The "info" URI Scheme

4.1. Definition of "info" URI Syntax

The "info" URI syntax presented in this document is conformant with the generic URI syntax defined in RFC 3986 [RFC3986]. This specification uses the Augmented Backus-Naur Form (ABNF) notation of RFC 4234 [RFC4234] to define the URI. The following core ABNF productions are used by this specification as defined by Appendix B.1 of RFC 4234: ALPHA, DIGIT, HEXDIG.

The "info" URI syntax is presented in two parts. Part A contains productions specific to the "info" URI scheme, while Part B contains generic productions from RFC 3986 [RFC3986], which are repeated here both for completeness and for reference. The following set of productions (Part A) is specific to the "info" URI scheme:
; Part A:
; productions specific to the "info" URI scheme

info-URI = info-scheme :: info-identifier [ "#" fragment ]
info-scheme = "info"
info-identifier = namespace "/" identifier
namespace = scheme
identifier = *( pchar / "/" )

; Note that "info" URIs containing dot-segments (i.e., segments
; whose full content consists of "." or ".") MAY NOT be suitable
; for use with applications that perform dot-segment normalization

This next set of productions (Part B) are generic productions
reproduced from RFC 3986 [RFC3986]:

; Part B:
; generic productions from RFC 3986 [RFC3986]
scheme = ALPHA *( ALPHA / DIGIT / "+" / "-" / "." )
pchar = unreserved / pct-encoded / sub-delims / ":" / "@"
fragment = *( pchar / "/" / "?" )
unreserved = ALPHA / DIGIT / "-" / "." / ";" / "_" / ";"
pct-encoded = "%" HEXDIG HEXDIG
sub-delims = ";" / ";" / ";" / ";" / ";" / ";" / ";" / ";" / ";"

An "info" URI has an "info-identifier" as its scheme-specific part
and MAY take an optional "fragment" component. An "info-identifier"
is constructed by appending an "identifier" component to a
"namespace" component separated by a slash "/" character. The "info"
URI scheme is supportive of hierarchical processing as indicated by
the presence of the slash "/" character, although the slash "/"
character SHOULD NOT be interpreted as a strict hierarchy delimiter.

Values for the "namespace" component of the "info" URI are name
tokens composed of URI scheme characters only (cf. the "scheme"
production). They identify the public namespace in which the
(unescape) value for the "identifier" component originates, and are
registered in the "info" Registry, which guarantees their uniqueness.
and persistence. Although the "namespace" component is case-insensitive, the canonical form is lowercase and documents that specify values for the "namespace" component SHOULD do so using lowercase letters. An implementation SHOULD accept uppercase letters as equivalent to lowercase in "namespace" names, for the sake of robustness, but SHOULD only generate lowercase "namespace" names, for consistency.

Values for the "identifier" component of the "info" URI MAY be viewed as being hierarchical strings composed of path segments built from path segment characters (cf. the "pchar" production), the segments being separated by slash "/" characters, although any semantic interpretation of the "/" character as a hierarchy delimiter MUST NOT be assumed. In their originating public namespace, the (unescaped) values for the "identifier" component identify information assets. The values for the "identifier" component MUST be %-escaped as required by this syntax. The "identifier" component SHOULD be treated as case-sensitive, although the "info" Registry MAY record the case-sensitivity of identifiers from particular registered public namespaces. The "info" Registry MAY also disclose additional normalization rules regarding the treatment of punctuation characters and the like.

Values for the "fragment" component of the "info" URI are strings composed of path segment characters (cf. the "pchar" production) plus the slash "/" character and the question mark "?" character. No semantic role is assigned to the slash "/" character and the question mark "?" character within the "fragment" component. The (unescaped) values for the "fragment" component identify secondary information assets with respect to the primary information asset, which is referenced by the "info-identifier". The values for the "fragment" component MUST be %-escaped as required by this syntax. The "fragment" component MUST be treated as being case-sensitive.

4.2. Allowed Characters Under the "info" URI Scheme

The "info" URI syntax uses the same set of allowed US-ASCII characters as specified in RFC 3986 [RFC3986] for a generic URI. An "info" URI string SHOULD be represented as a Unicode [UNICODE] string and be encoded in UTF-8 [RFC3629] form. Reserved characters as well as excluded US-ASCII characters and non-US-ASCII characters MUST be %-escaped before forming the URI. Details of the %-escape encoding can be found in RFC 3986 [RFC3986], Section 2.4.
4.3. Examples of "info" URIs

Some examples of syntactically valid "info" URIs are given below:

a) info:ddc/22/eng//004.678

where "ddc" is the "namespace" component for a Dewey Decimal Classification [DEWEY] namespace and "22/eng//004.678" is the "identifier" component for an identifier of an information asset within that namespace.

The information asset identified by the identifier "22/eng//004.678" in the namespace for (22nd Ed.) English-language Dewey Decimal Classifications is the classification "Internet"

b) info:lccn/2002022641

where "lccn" is the "namespace" component for a Library of Congress Control Number [LCCN] namespace and "2002022641" is the "identifier" component for an identifier of an information asset within that namespace.

The information asset identified by the identifier "2002022641" in the namespace for Library of Congress Control Numbers is the metadata record "Newcomer, Eric. Understanding Web services: XML, WSDL, SOAP, and UDDI. Boston: Addison-Wesley, 2002."

c) info:sici/0363-0277(19950315)120:5%3C%3E1.0.TX;2-V

where "sici" is the "namespace" component for a Serial Item and Contribution Identifier [SICI] namespace and "0363-0277(19950315)120:5%3C%3E1.0.TX;2-V" is the "identifier" component for an identifier of an information asset in that namespace in %-escaped form, or in unescaped form "0363-0277(19950315)120:5<>1.0.TX;2-V".

The information asset identified by the identifier "0363-0277(19950315)120:5<>1.0.TX;2-V" in the namespace for Serial Item and Contribution Identifiers is the journal issue "Library Journal, Vol. 120, no. 5. March 15, 1995."
d) <rdf:Description about="info:bibcode/2003Icar..163..263Z"/>

where "bibcode" is the "namespace" component for a NASA Astrophysics Data System (ADS) Bibcode [BIBCODE] namespace and "2003Icar..163..263Z" is the "identifier" component for an identifier of an information asset within that namespace. This example further shows an application of an "info" URI as the subject of a Resource Description Framework (RDF) statement.

The information asset identified by the identifier "2003Icar..163..263Z" in the namespace for NASA ADS Bibcodes is the metadata record in the ADS system that describes the journal article


e) info:pmid/12376099

where "pmid" is the "namespace" component for a PubMed Identifier [PMID] namespace and "12376099" is the "identifier" component for an identifier of an information asset in that namespace.

The information asset identified by the identifier "12376099" in the namespace for PubMed Identifiers is the metadata record in the PubMed database that describes the journal article


5. Normalization and Comparison of "info" URIs

In order to facilitate comparison of "info" URIs, a sequence of normalization steps SHOULD be applied as detailed below. After normalizing the URI strings, comparison of two "info" URIs is then applied on a character-by-character basis as prescribed by RFC 3986 [RFC3986], Section 6.2.1.

The following generic normalization steps SHOULD anyway be applied by applications processing "info" URIs:

   a) Normalize the case of the "scheme" component to be lowercase
   b) Normalize the case of the "namespace" component to be lowercase
   c) Unescape all unreserved %-escaped characters in the "namespace" and "identifier" components
d) Normalize the case of any %-escaped characters in the "namespace" and "identifier" components to be uppercase.

Further normalization steps MAY be applied by applications to "info" URIs based on rules recorded in the "info" Registry for a registered public namespace, but such normalization steps remain outside of the scope of the "info" URI definition.

Since the "info" URI SHOULD be treated as being case-sensitive, a canonical form MAY only be arrived at by consulting the "info" Registry for possible information on the case-sensitivity for identifiers from a registered public namespace, and any case normalization step to apply. The "info" Registry MAY also disclose additional normalization rules regarding the treatment of punctuation characters and the like.

In cases, however, where no single canonical form of the "identifier" component exists, it is nevertheless RECOMMENDED that a Namespace Authority nominate a preferred form, which SHOULD be used wherever possible within an "info" URI so that applications MAY have an increased chance of successful comparison of two "info" URIs.

Note that "info" URIs containing dot-segments (i.e., segments whose full content consists of "." or ".") MAY NOT be suitable for use with applications that perform dot-segment normalization.

The following unnormalized forms of an "info" URI

   U1. INFO:PII/S0888-7543(02)96852-7
   U2. info:PII/S0888754302968527
   U3. info:pii/S0888%2D7543%2802%2996852%2D7
   U4. info:pii/s0888-7543(02)96852-7

are normalized to the following respective forms

   N1. info:pii/S0888-7543(02)96852-7
   N2. info:pii/S0888754302968527
   N3. info:pii/S0888-7543(02)96852-7
   N4. info:pii/s0888-7543(02)96852-7

The "info" URI definition does not prescribe further normalization steps, although applications MAY apply additional normalization steps according to any rules recorded in the "info" Registry for a registered public namespace.
6. Rationale

6.1. Why Create a New URI Scheme for Identifiers from Public Namespaces?

Under RFC 4395, "Guidelines and Registration Procedures for New URI Schemes" [RFC4395], it is stated in Section 2.1 "Demonstrable, New, Long-Lived Utility" that "New URI schemes SHOULD have clear utility to the broad Internet community, beyond that available with already registered URI schemes". The "info" URI scheme allows identifiers within public namespaces, used for the identification of information assets, to be referred to within the URI allocation. Once a namespace is registered in the "info" Registry, the "info" URI scheme enables an information asset with an identifier in that namespace to be referenced by means of a URI. As a result, the information asset SHALL be considered a resource as defined in RFC 3986 [RFC3986] and SHALL enjoy the same common syntactic, semantic, and shared language benefits that the URI presentation confers.

6.2. Why Not Use an Existing URI Scheme for Identifiers from Public Namespaces?

Existing URI schemes are not suitable for employment as the "info" URI scheme admits of no global dereference mechanism. While examples of resource identifiers minted under other URI schemes MAY not always be dereferenceable, nevertheless there is always a common expectation that such URIs can be dereferenced by various resolution mechanisms, whether they be location-dependent or location-independent resource identifiers. The "info" URI scheme applies to a class of resource identifiers whose Namespace Authorities MAY or MAY NOT choose to disclose service mechanisms. Nevertheless, Namespace Authorities are encouraged to disclose in the "info" registration record references to any such service mechanisms in order to provide a greater utility to network applications.

6.3. Why Not Create a New URN Namespace ID for Identifiers from Public Namespaces?

RFC 2141 [RFC2141] states that "Uniform Resource Names (URNs) are intended to serve as persistent, location-independent, resource identifiers". The "info" URI scheme, on the other hand, does not assert the persistence of the identifiers created under this scheme but rather of the public namespaces grandfathered under this scheme. It exists primarily to disclose the identity of information assets and to facilitate a lightweight registration mechanism for public namespaces of identifiers managed according to the policies and business models of the Namespace Authorities. The "info" URI scheme is neutral with respect to identifier persistence. Moreover, for
"info" to operate as a URN Network Identifier (NID) would require that "info" be constituted as a delegated naming authority. It is not clear that a URN NID would be an appropriate choice for naming authority delegation.

Further, the "info" URI scheme is not globally dereferenceable in contrast to the specific recommendation given in RFC 1737, "Functional Requirements for Uniform Resource Names" [RFC1737] that "It is strongly recommended that there be a mapping between the names generated by each naming authority and URLs". Individual Namespace Authorities registered in the "info" Registry MAY, however, disclose references to service mechanisms and are encouraged to do so.

An extra consideration is that the "urn" URI syntax explicitly excludes generic URI hierarchy by reserving the slash "/" character. An "info" URI, on the other hand, admits of hierarchical processing, while remaining neutral with respect to supporting actual hierarchy, and thus allows the slash "/" character (as well as more liberally allowing the ampersand "&" and tilde "~" characters). It therefore represents a lower barrier to entry for Namespace Authorities in keeping with its intention of acting as a bridging mechanism to allow public namespaces to become part of the URI allocation. In sum, an "info" URI is more widely supportive of "human transcribability" as discussed in RFC 3986 [RFC3986] than is a "urn" URI.

Additionally, the "urn" URI syntax does not support "fragment" components as does the "info" URI syntax for indirect identification of secondary resources.

7. Security Considerations

The "info" URI scheme syntax is subject to the same security considerations as the generic URI syntax described in RFC 3986 [RFC3986].

While some "info" Namespace Authorities MAY choose to disclose service mechanisms, any security considerations resulting from the execution of such services fall outside the scope of this document. It is strongly recommended that the registration record of an "info" namespace include any such considerations.
8. IANA Considerations

The IANA registry for URI schemes
<http://www.iana.org/assignments/uri-schemes.html> SHOULD be updated
to include an entry for the "info" URI scheme when the "info" URI
scheme is accepted for publication as an RFC. This entry SHOULD
contain the following values:

Scheme Name: info

Description: Information Assets with Identifiers in Public
Namespaces

Reference: RFC 4452

9. Acknowledgements

The authors acknowledge the contributions of Michael Mealling,
Verisign, and Patrick Hochstenbach, Ghent University.

10. References

10.1. Normative References

[RFC1737] Sollins, K. and L. Masinter, "Functional Requirements for

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate


[RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO

Resource Identifier (URI): Generic Syntax", STD 66,


[RFC4395] Hansen, T., Hardie, T., and L. Masinter, "Guidelines and
Registration Procedures for New URI Schemes", BCP 115, RFC
4395, February 2006.
10.2. Informative References


Authors’ Addresses

Herbert Van de Sompel
Los Alamos National Laboratory
Research Library, MS-P362
PO Box 1663
Los Alamos, NM  87545-1362
USA

EMail: herbertv@lanl.gov

Tony Hammond
Nature Publishing Group
Macmillan House
4 Crinan Street
London   N1 9XW
UK

EMail: t.hammond@nature.com

Eamonn Neylon
Manifest Solutions
Bicester, Oxfordshire   OX26 2HX
UK

EMail: eneylon@manifestsolutions.com

Stuart L. Weibel
OCLC Online Computer Library Center, Inc.
6565 Frantz Road
Dublin, OH  43017-3395
USA

EMail: weibel@oclc.org
Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).