

Network Working Group
Request for Comments: 3305
Category: Informational

M. Mealling, Ed.
R. Denenberg, Ed.
W3C URI Interest Group
August 2002

Report from the Joint W3C/IETF URI Planning Interest Group:
Uniform Resource Identifiers (URIs), URLs, and Uniform Resource Names
(URNs): Clarifications and Recommendations

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 (2002). All Rights Reserved.

Abstract

This document, a product of the W3C Uniform Resource Identifier (URI) Interest Group, addresses and attempts to clarify issues pertaining to URIs. This document addresses how URI space is partitioned and the relationship between URIs, URLs, and URNs, describes how URI schemes and URN namespaces are registered, and presents recommendations for continued work on this subject.

Table of Contents

1.	The W3C URI Interest Group	2
2.	URI Partitioning	2
2.1	Classical View	3
2.2	Contemporary View	3
2.3	Confusion	3
3.	Registration	4
3.1	URI Schemes	4
3.1.1	Registered URI schemes	4
3.1.2	Unregistered URI Schemes	4
3.1.2.1	Public Unregistered Schemes	4
3.1.2.2	Private Schemes	5
3.1.3	Registration of URI Schemes	5
3.1.3.1	IETF Tree	5
3.1.3.2	Other Trees	5
3.2	URN Namespaces	5
3.2.1	Registered URN NIDs	5
3.2.2	Pending URN NIDs	6
3.2.3	Unregistered NIDs	7
3.2.4	Registration Procedures for URN NIDs	7
4.	Additional URI Issues	7
5.	Recommendations	8
6.	Security Considerations	8
7.	Acknowledgements	8
	References	9
	Authors' Addresses	10
	Full Copyright Statement	11

1. The W3C URI Interest Group

In October, 2000 the W3C formed a planning group whose mission was to evaluate the opportunities for W3C work in the area of Uniform Resource Identifiers (URIs) and to develop a proposal for continued work in this area. The Interest Group was composed of W3C members and invited experts from the IETF to participate as well. This document is a set of recommendations from this group, to the W3C and the IETF for work that can and should continue in this area.

2. URI Partitioning

There is some confusion in the web community over the partitioning of URI space, specifically, the relationship among the concepts of URL, URN, and URI. The confusion owes to the incompatibility between two different views of URI partitioning, which we call the "classical" and "contemporary" views.

2.1 Classical View

During the early years of discussion of web identifiers (early to mid 90s), people assumed that an identifier type would be cast into one of two (or possibly more) classes. An identifier might specify the location of a resource (a URL) or its name (a URN), independent of location. Thus a URI was either a URL or a URN. There was discussion about generalizing this by the addition of a discrete number of additional classes; for example, a URI might point to metadata rather than the resource itself, in which case the URI would be a URC (citation). URI space was thus viewed as partitioned into subspaces: URL, URN, and additional subspaces to be defined. The only such additional space ever proposed was Uniform Resource Characteristics (URC) and there never was any buy-in; so without loss of generality, it's reasonable to say that URI space was thought to be partitioned into two classes: URL and URN. Thus, for example, "http:" was a URL scheme, and "isbn:" would (someday) be a URN scheme. Any new scheme would be cast into one of these two classes.

2.2 Contemporary View

Over time, the importance of this additional level of hierarchy seemed to lessen; the view became that an individual scheme did not need to be cast into one of a discrete set of URI types, such as "URL", "URN", "URC", etc. Web-identifier schemes are, in general, URI schemes, as a given URI scheme may define subspaces. Thus "http:" is a URI scheme. "urn:" is also a URI scheme; it defines subspaces, called "namespaces". For example, the set of URNs, of the form "urn:isbn:n-nn-nnnnnn-n", is a URN namespace. ("isbn" is an URN namespace identifier. It is not a "URN scheme", nor is it a "URI scheme.")

Further, according to the contemporary view, the term "URL" does not refer to a formal partition of URI space; rather, URL is a useful but informal concept. A URL is a type of URI that identifies a resource via a representation of its primary access mechanism (e.g., its network "location"), rather than by some other attributes it may have. Thus, as we noted, "http:" is a URI scheme. An http URI is a URL. The phrase "URL scheme" is now used infrequently, usually to refer to some subclass of URI schemes which exclude URNs.

2.3 Confusion

The body of documents (RFCs, etc) covering URI architecture, syntax, registration, etc., spans both the classical and contemporary periods. People who are well-versed in URI matters tend to use "URL" and "URI" in ways that seem to be interchangeable. Among these experts, this isn't a problem, but among the Internet community at

large, it is a problem. People are not convinced that URI and URL mean the same thing, in documents where they (apparently) do. When one RFC talks about URI schemes (e.g. "URI Syntax" (RFC 2396) [12]), another talks about URL schemes (e.g. "Registration Procedures for URL Schemes" (RFC 2717) [1]), and yet another talks of URN schemes ("Architectural Principles of URN Resolution" (RFC 2276) [13]), it is natural to wonder how they differ, and how they relate to one another. While RFC 2396, section 1.2, attempts to address the distinction between URIs, URLs and URNs, it has not been successful in clearing up the confusion.

3. Registration

This section examines the state of registration of URI schemes and URN namespaces and the mechanisms by which registration currently occurs.

3.1 URI Schemes

3.1.1 Registered URI schemes

The official register of URI scheme names is maintained by IANA, at <http://www.iana.org/assignments/uri-schemes>. For each scheme, the RFC that defines the scheme is listed; for example "http:" is defined by RFC2616 [14]. The table lists 34 schemes (at time of publication of this RFC). In addition, there are a few "reserved" scheme names; at one point in time, these were intended to become registered schemes but have since been dropped.

3.1.2 Unregistered URI Schemes

We distinguish between public (unregistered) and private schemes. A public scheme (registered or not) is one for which there is some public document describing it.

3.1.2.1 Public Unregistered Schemes

Dan Conolly's paper, at <http://www.w3.org/Addressing/schemes>, provides a list of known public URI schemes, both registered and unregistered, a total of 85 schemes at time of publication of this RFC. 50 or so of these are unregistered (not listed in the IANA register). Some of these URI schemes are obsolete (for example, "phone" is obsolete, superseded by "tel"), while some have an RFC, but are not included in the IANA list.

3.1.2.2 Private Schemes

It is probably impossible to determine all of these, and it's not clear that it's worthwhile to try, except perhaps to get some idea of their number. In the minutes of the August 1997 IETF meeting is the observation that there may be 20-40 in use at Microsoft, with 2-3 being added a day, and that WebTV has 24, with 6 added per year.

3.1.3 Registration of URI Schemes

"Registration Procedures for URL Scheme Names" (RFC 2717) [1] specifies procedures for registering scheme names and points to "Guidelines for new URL Schemes" (RFC 2718) [2], which supplies guidelines. RFC 2717 describes an organization of schemes into "trees". It is important to note that these two documents use the historical term 'URL' when in fact, they refer to URIs in general. In fact, one of the recommended tasks in Section 5 is for these documents to be updated to use the term 'URI' instead of 'URL'.

3.1.3.1 IETF Tree

The IETF tree is intended for schemes of general interest to the Internet community, and for those which require a substantive review and approval process. Registration in the IETF tree requires publication of the scheme syntax and semantics in an RFC.

3.1.3.2 Other Trees

Although RFC 2717 describes "alternative trees", no alternative trees have been registered to date, although a vendor-supplied tree ("vnd") is pending. URI schemes in alternative trees will be distinguished because they will have a "." in the scheme name.

3.2 URN Namespaces

A URN namespace is identified by a "Namespace ID" (NID), which is registered with IANA (see Section 3.2.4).

3.2.1 Registered URN NIDs

There are two categories of registered URN NIDs:

- o Informal: These are of the form, "urn-`<number>`", where `<number>` is assigned by IANA. There are four registered (at time of publication of this RFC) in this category (urn-1, urn-2, urn-3, and urn-4).

- o Formal: The official list of registered NIDs is kept by IANA at <http://www.iana.org/assignments/urn-namespaces>. At the time of publication of this RFC it lists ten registered NIDs:
 - * 'ietf', defined by "URN Namespace for IETF Documents" (RFC 2648) [3]
 - * 'pin', defined by "The Network Solutions Personal Internet Name (PIN): A URN Namespace for People and Organizations" (RFC 3043) [4]
 - * 'issn' defined by "Using The ISSN as URN within an ISSN-URN Namespace" (RFC 3043) [4]
 - * 'oid' defined by "A URN Namespace of Object Identifiers" (RFC 3061) [6]
 - * 'newsml' defined by "URN Namespace for NewsML Resources" (RFC 3085) [7]
 - * 'oasis' defined by "A URN Namespace for OASIS" (RFC 3121) [8]
 - * 'xmlorg' defined by "A URN Namespace for XML.org" (RFC 3120) [9]
 - * 'publicid' defined by "A URN Namespace for Public Identifiers" (RFC 3151) [10]
 - * 'isbn' defined by "Using International Standard Book Numbers as Uniform Resource Names" (RFC 3187) [15]
 - * 'nbn' defined by "Using National Bibliography Numbers as Uniform Resource Names" (RFC 3188) [16]

3.2.2 Pending URN NIDs

There are a number of pending URN NID registration requests, but there is no reliable way to discover them, or their status. It would be helpful if there were some formal means to track the status of NID requests such as 'isbn'.

3.2.3 Unregistered NIDs

In the "unregistered" category (besides the experimental case, not described in this paper), there are entities that maintain namespaces that, while completely appropriate as URNs, just haven't bothered to explore the process of NID registration. The most prominent that comes to mind is 'hdl'. In the case of 'hdl', it has been speculated that this scheme has not been registered because it is not clear to the owners whether it should be registered as a URI scheme or as a URN namespace.

3.2.4 Registration Procedures for URN NIDs

"URN Namespace Definition Mechanisms" (RFC 2611) [11] describes the mechanism to obtain an NID for a URN namespace, which is registered with IANA.

A request for an NID should describe features including: structural characteristic of identifiers (for example, features relevant to caching/shortcuts approaches); specific character encoding rules (e.g., which character should be used for single-quotes); RFCs, standards, etc, that explain the namespace structure; identifier uniqueness considerations; delegation of assignment authority, including how to become an assigner of identifiers; identifier persistence considerations; quality of service considerations; process for identifier resolution; rules for lexical equivalence; any special considerations required for conforming with the URN syntax (particularly applicable in the case of legacy naming systems); validation mechanisms (determining whether a given string is currently a validly-assigned URN); and scope (for example, "United States social security numbers").

4. Additional URI Issues

There are additional unresolved URI issues not considered by this paper, which we hope will be addressed by a follow-on effort. We have not attempted to completely enumerate these issues, however, they include (but are not limited to) the following:

- o The use of URIs as identifiers that don't actually identify network resources (for example, they identify an abstract object, such as an XML namespace, or a physical object such as a book or even a person).
- o IRIs (International Resource Identifiers): the extension of URI syntax to non-ASCII.

5. Recommendations

We recommend the following:

1. The W3C and IETF should jointly develop and endorse a model for URIs, URLs, and URNs consistent with the "Contemporary View" described in section 1, and which considers the additional URI issues listed or alluded to in section 3.
2. RFCs such as 2717 ("Registration Procedures for URL Scheme Names") and 2718 ("Guidelines for new URL Schemes") should both be generalized to refer to "URI schemes", rather than "URL schemes" and, after refinement, moved forward as Best Current Practices in the IETF.
3. The registration procedures for alternative trees should be clarified in RFC 2717.
4. Public, but unregistered schemes, should become registered, where possible. Obsolete schemes should be purged or clearly marked as obsolete.
5. IANA registration information should be updated:
 - * Add 'urn' to the list of registered URI schemes with a pointer to the URN namespace registry.
 - * Maintain status information about pending registrations (URI schemes and URN NID requests).
 - * Insure that it is clear that the page is the official registry, e.g., by adding a heading to the effect "This is the Official IANA Registry of URI Schemes".

6. Security Considerations

This memo does not raise any known security threats.

7. Acknowledgements

The participants in the URI Planning Interest Group are:

- o Tony Coates
- o Dan Connolly
- o Diana Dack

- o Leslie Daigle
- o Ray Denenberg
- o Martin Duerst
- o Paul Grosso
- o Sandro Hawke
- o Renato Iannella
- o Graham Klyne
- o Larry Masinter
- o Michael Mealling
- o Mark Needleman
- o Norman Walsh

References

- [1] Petke, R. and I. King, "Registration Procedures for URL Scheme Names", BCP 35, RFC 2717, November 1999.
- [2] Masinter, L., Alvestrand, H., Zigmond, D. and R. Petke, "Guidelines for new URL Schemes", RFC 2718, November 1999.
- [3] Moats, R., "A URN Namespace for IETF Documents", RFC 2648, August 1999.
- [4] Mealling, M., "The Network Solutions Personal Internet Name (PIN): A URN Namespace for People and Organizations", RFC 3043, January 2001.
- [5] Rozenfeld, S., "Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace", RFC 3044, January 2001.
- [6] Mealling, M., "A URN Namespace of Object Identifiers", RFC 3061, February 2001.
- [7] Coates, A., Allen, D. and D. Rivers-Moore, "URN Namespace for NewsML Resources", RFC 3085, March 2001.

- [8] Best, K. and N. Walsh, "A URN Namespace for OASIS", RFC 3121, June 2001.
- [9] Best, K. and N. Walsh, "A URN Namespace for XML.org", RFC 3120, June 2001.
- [10] Walsh, N., Cowan, J. and P. Grosso, "A URN Namespace for Public Identifiers", RFC 3151, August 2001.
- [11] Daigle, L., van Gulik, D., Iannella, R. and P. Faltstrom, "URN Namespace Definition Mechanisms", BCP 33, RFC 2611, June 1999.
- [12] Berners-Lee, T., Fielding, R. and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", RFC 2396, August 1998.
- [13] Sollins, K., "Architectural Principles of Uniform Resource Name Resolution", RFC 2276, January 1998.
- [14] Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P. and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.
- [15] Hakala, J. and H. Walravens, "Using International Standard Book Numbers as Uniform Resource Names", RFC 3187, October 2001.
- [16] Hakala, J., "Using National Bibliography Numbers as Uniform Resource Names", RFC 3188, October 2001.

Authors' Addresses

Michael Mealling
VeriSign, Inc.
21345 Ridgetop Circle
Sterling, VA 20166
US

EEmail: michael@verisignlabs.com

Ray Denenberg
Library of Congress
Washington, DC 20540
US

EEmail: rden@loc.gov

Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS 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.

Acknowledgement

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