Network Working Group W3C URI IG
Internet-Draft WorldWideWeb Consortium
Expires: March 5, 2002 September 4, 2001

URI Schemes and URN Namespaces
draft-mealling-uri-ig-00.txt

M.Mealling

Status of this Memo

Copyright Notice

Abstract

Ed. note (temporary): This paper comines the two earlier papers, "URI Partitioning" (Draft 2) and "Current State of Registration of URI Schemes and URN Namespaces" (Draft 2), both dated February 16.

1. Background

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.

1.1 Classical View

During the early years of discussion of web identifiers (early to mid 90s), people assumed that an identifers 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 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 and URN, and additional subspaces, to be defined. The only such additional space ever proposed was 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 or the other of these two classes.

1.2 Contemporary View

It seems that the prevailing view now (and for the past few years) is that no need exists for this additional level of hierarchy. An individual scheme does 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; 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 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 identifying the resource by name or by some other attributes of that resource. Thus as we noted, "http:" is a URI scheme. An http URI is a URL. However, "http:" is not a URL scheme, since the term "URL scheme" is no longer used.

1.3 Confusion

The body of documents (RFCs, etc) covering URI architecture, syntax, registration, etc., spans both the classical and contemporary periods. Therefore, experts (or in general, people who are well-versed in URI matters), tend to use "URL" and "URI" interchangeably. Among these experts, this isn't a problem. But among the Internet

community at large, it is. There has never been any official endorsement, from the IETF or W3C, of the contemporary view. People are not convinced that URI and URL mean the same thing, in documents where they (apparently) do. When one sees an RFC that talks about URI schemes (e.g. RFC 2396, "URI Syntax"), another that talks about URL schemes (e.g. RFC 2717, "Registration Procedures for URL Schemes"), and yet another that talks of URN schemes (e.g. RFC 2276, "Architectural Principles of URN Resolution") it is natural to wonder what's the difference, and how they relate to one another. (Note that section 1.2 of RFC 2396 addresses the distinction between URIs, URLs, and URNs, but does little to clear up this confusion.)

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

2. URI Schemes

2.1 Registered URI schemes

The official register of URI scheme names is maintained by IANA, at http://www.iana.org/assignments/uri-schemes (See notes below). For each scheme, the RFC that defines the scheme is listed, for example "http:" is defined by RFC2068. The table currently lists 30 schemes. In addition, there are a few "reserved" scheme names; presumably, these are, or were, intended to become registered schemes but no RFC has yet been approved.

Notes:

* A reader would likely not know that it is the official register, since there is nothing conveyed by that page about its status. It would be very helpful if a heading were added to this page to the effect "This is the Official IANA Register of URI Schemes".

* The heading on the register is "URL Schemes". We consider this to be the register of URI schemes (see part 1 under "Contemporary View"). It would be helpful if IANA would refer to these as URI schemes.

* If this register is intended to be a complete list of registered URI schemes then URN needs to be added (and the entry for URN should point to the list of registered namespaces).

2.2 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.

2.2.1 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 un-registered, a total of 84 schemes. 50 or so of these are unregistered (not listed in the IANA register). Some may be obsolete (for example, it appears that "phone", is obsolete, superceded by "tel"). Some have an RFC, but are not included in the IANA list.

2.2.2 Private Schemes

It's 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.

RFC 2717: "Registration Procedures for URL Scheme Names" [1]specifies procedures for registering scheme names, and points to RFC 2718: "Guidelines for new URL Schemes" [2] which supplies guidelines. RFC 2717 describes an organization of schemes into "trees".

2.3 IETF Tree

The IETF tree is intended for schemes of general interest to the Internet community, and which require a substantive review and approval process. Registration in the IETF tree requires publication of the scheme syntax and semantics in an RFC. Presumably the schemes listed in the IANA registry are all in the IETF tree (it seems no other tree has been defined). It is not clear how to tell what tree a given URI scheme belongs to.

2.4 Other Trees

Although RFC 2717 describes "alternative trees", it does not seem to describe how to register a tree, nor where is the official register of trees.

3. URN Namespaces

A URN namespace is identified by a "Namespace ID", NID, which is registered with IANA (see 2.2.4).

3.1 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 three registered in this category (urn-1, urn-2, and urn-3).

- o Formal: The official list of registered NIDs is kept by IANA at http://www.iana.org/assignments/urn-namespaces. Currently it lists eight registered NIDs:

  * 'ietf', defined by RFC 2648: "URN Namespace for IETF Documents" [3]

  * 'pin', defined by RFC 3043: " The Network Solutions Personal Internet Name (PIN): A URN Namespace for People and Organizations" [4]

  * 'issn' defined by RFC 3044: "Using The ISSN as URN within an ISSN-URN Namespace" [5]

  * 'oid' defined by RFC 3061: "A URN Namespace of Object Identifiers" [6]

  * 'newsml' defined by RFC 3085: "URN Namespace for NewsML Resources" [7]

  * 'oasis' defined by RFC 3121: "A URN Namespace for OASIS" [8]

  * 'xmlorg' defined by RFC 3120: "A URN Namespace for XML.org" [9]

  * 'publicid' defined by RFC 3151: "A URN Namespace for Public Identifiers" [10]

3.2 Pending URN NIDs

Michael Mealling maintains an informal list (not to be confused with the "informal" category) of known NIDs at http://www.uri.net/urn-nid-status.html. It includes registered as well as pending NIDs, and tracks their status. This informal document seems to be the only source of information on the status of NIDs other than those formally registered. For example, 'isbn' and 'nbn' have been approved by the IESG and are in the RFC Editor's queue. 'isbn', as a potential URN namespace (or URI scheme), in particular has been a source of much speculation and confusion over several years. It would be helpful if there were a better-known means to discover the status of a pending

URN NID.

3.3 Unregistered NIDs

In the "unregistered" category (besides the experimental case, not described in this paper) there are bona fide NIDs that just haven't bothered to even explore the process of 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.

3.4 Registration Procedures for URN NIDs

RFC 2611: "URN Namespace Definition Mechanisms" [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 explains 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. Recommendations

The W3C URI Interest Group recommends the following:

- The W3C should jointly develop and publish a model for URIs, URLs and URNs consistent with the '"Contemporary View" described in section 1.

- Public but unregistered schemes should become registered, where possible. Obsolete schemes should be purged.

- RFCs such as RFC 2717 ("Registration Procedures for URL Scheme Names") [1] and RFC 2718 ("Guidelines for new URL Schemes") [2] should both be generalized to refer to "URI schemes" rather than "URL schemes". Both documents should be moved to "Best Current

Practice" status and also documented as a W3C Recommendation.

- o 'urn' should be added to the list of registered URI schemes.

- o If there is, or will be, more than a single URI-scheme tree, there should be a way for people to discover: for a given scheme what tree the scheme belongs to; how to register a tree; where the official register is.

- o Status information about pending URN NID requests should be maintained.

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.

Author's Address

W3C URI Planning Interest Group
WorldWideWeb Consortium

Full Copyright Statement

Acknowledgement

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