Uniform Resource Names (urnbis)                                  J. Klensin
Internet-Draft                                           February 5, 2016
Updates: 3986 (if approved)
Intended status: Standards Track
Expires: August 8, 2016

URN Semantics Clarification
draft-ietf-urnbis-semantics-clarif-03.txt

Abstract

Experience has shown that identifiers associated with persistent names have properties and requirements that may be somewhat different from identifiers associated with the locations of objects. This is especially true when such names are expected to be stable for a very long time or when they identify large and complex entities. In order to allow Uniform Resource Names (URNs) to evolve to meet the needs of the Library, Museum, Publisher, and Information Science communities and other users, this specification separates URNs from the semantic constraints that many people believe are part of the specification for Uniform Resource Identifiers (URIs) in RFC 3986, updating that document accordingly. The syntax of URNs is still constrained to that of RFC 3986, so generic URI parsers are unaffected by this change.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 8, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.
1. Introduction

The Generic URI Syntax specification [RFC3986] covers both locators and names and mixtures of the two (See its Section 1.1.3) and describes Uniform Resource Locators (URLs) -- first documented in the IETF in RFC 1738 [RFC1738] -- as an embodiment of the locator concept and Uniform Resource Names (URNs), specifically those using the "urn" scheme [RFC2141], as an embodiment of the names that do not directly provide for resource location. This specification is concerned only about URNs of the variety described in RFC 2141 [RFC2141] and its...
successors [RFC2141bis] (i.e., those that use the "urn" scheme). URLs, other types of names, and any URI types that may not fall into one of the above categories are out of its scope and unaffected by it.

Experience with URNs since the publication of RFC 3986 has identified several ways in which their inclusion under the 3986 scope has hampered understanding, adoption, and especially extension (specifically extensions of types that were anticipated, but not defined, in RFC 2141). The need for extensions to the URN concept is now being felt in some communities, especially those that include libraries, museums, publishers, and other information scientists.

In particular, the Generic URI Syntax specification goes beyond syntax to specify the meaning and interpretation of various fields, especially the "query" and "fragment" ones and the various syntax forms and interpretations it allows for <hier-part>. This specification excludes URNs from those definitions of meaning and interpretation so that RFC 3986 applies to their syntax only. The meaning --and any more specific syntax rules-- for those fields for URNs are now defined in a URN-specific document [RFC2141bis]. URNs remain members of the URI family and parsers for generic URI syntax are not affected by this specification although parsers that make assumptions based on other URI schemes obviously might be.

Neither this specification nor the successor to RFC 2141 [RFC2141bis] discusses DDDS [RFC3401] resolution or conversion to (and interpretation of) URCs [RFC2483] or, with the exception of providing some syntax to cover some specific cases, URN "resolution" more generally. Any of those topics that do need to be addressed should be covered in other documents. The document also does not discuss alternatives to URNs, either those that might use a different scheme name within the RFC 3986 URI framework or those that might use a different framework entirely. In particular, some externally-defined content or object identification systems could be represented either by a URN namespace or through separate URI schemes. This specification does not offer advice on that choice other than to suggest that the two options not be confused (or both used in a way that would be confusing).

This document updates RFC 3986 to make the distinction between syntax and semantics clear for URNs and to isolate URNs from presumed URI semantic requirements. It is important to note that some readers of RFC 3986 are convinced that the separation is clear in that specification and therefore that no changes to that document are needed. For them, this specification is only a confirming clarification.
In the long term, as the expanded syntax and uses of URNs become commonplace and RFC 3986 is updated, this specification is likely to become of historical interest only, providing an extended rationale for decisions made and adjustment of the boundary between URN specifications and generic URI ones.

2. Pragmatic Goals

Despite the important background and rationale in the sections that follow, the change made (or clarification provided) by this specification is driven by a desire to avoid philosophical debates about terminology or ultimate truths. Instead, it is motivated by three very pragmatic principles and goals:

1. Accommodate all of those who think URNs are necessary, i.e., that they can and should be usefully distinguished from other URIs, at least location-oriented ones including URI schemes defined prior to the time work started on this document in August 2014. In particular, provide a foundation for extensions to the URN syntax (as allowed by and partially defined in RFC 2141) to support requirements encountered by some of those communities.

2. Provide a path to avoid getting bogged down in declarative statements about definitions and debates about what is and is not correct in the abstract.

3. Avoid a fork in the standard that would be likely to lead to multiple, conflicting, definitions or criteria for URNs.

In addition, this document is intended to move past debates about whether or not URNs are intended to be parsed at all (i.e., whether a "urn"-scheme URI is simply opaque to a URI parser once the scheme name is identified) and, if not, how much of it is actually expected to be understood and broken into identifiable parts by such a parser. It establishes a principle that, for the "urn" scheme, parsing into the components identified in RFC 3986 will be performed but that any meanings or interpretation assigned to those components (including that applicability of the normal English meanings of such terms as "query" or "fragment" are a matter for URN-specific specifications. It helps lay the foundation for the distinguishing terms "q-component", "r-component", and "f-component" in the accompanying URN definition specification [RFC2141bis].

3. The role of queries and fragments in URNs

Part of the concern that led to this document was a desire to accommodate URN components that would be analogous to the query and fragment components of generalized URNs but that might have different
properties. For many cases, the analogy cannot be exact. For example, RFC 3986 ties the interpretation of fragments to media types. Since media type is a function of specific content, URNs that are never resolved cannot have an associated media type, nor can URNs that resolve to, for example, other URIs that may then not be resolved further. Similarly, while the RFC 3986 syntax for queries (and fragments) may be entirely appropriate for URN use, terminology like "Service Request" (see Appendix B of the predecessor "URNs are not..." draft [Appendix B]”>ServiceRequests) for additional discussion) may be more suitable to the URN context than "query" (if, indeed, the portion of the URN that is syntactically equivalent to a URI query is where those requests belong).

4. Changes to RFC 3986

This specification removes URN semantics from the scope of RFC 3986. It makes no changes to the generic URI syntax. That syntax still applies to URNs as well as to other URI types. Even as regard to semantics, it has no practical effect for URNs defined in strict conformance to the prior URN specification [RFC2141] or the associated registration specification [RFC3406].

In particular (but without altering RFC 3986 in any way), the generic URI syntax for "queries" (strings starting with "?" and continuing to the end of the URI or to a "]"), and for "fragments" (strings starting with "]" and continuing to the end of the URI) is unchanged. For URNs, additional syntax is introduced to divide the URI "query" into two parts, referred to as "q-components" and "r-components". The syntax and general semantics of "fragments" (specified in RFC 3986 as scheme-independent) are unchanged, but a somewhat liberal interpretation may be needed in the context of URNs, so a fragment is referred to as an "f-component" as a term of convenience to highlight that distinction. [RFC2141bis].

5. Actions Occurring in Parallel with this Specification

The basic URN syntax specification [RFC2141] was published well before RFC 3986 and therefore does not depend on it. The successor to that specification [RFC2141bis], fully spells out, or references documents that spell out, the semantics and any required within-field syntax of URNs. It uses great care about generic or implicit reference to any URI specification and delegates further details to specific namespaces.

[[CREF1: Note in Draft: Perhaps this section can be dropped entirely.]]

Klensin Expires August 8, 2016 [Page 5]
6. Acknowledgments

This specification was inspired by a search in the IETF URNBIS WG for an approach that would both satisfy the needs of persistent name-type identifiers and still fully conform to the specifications and intent of RFC 3986. That search lasted several years and considered many alternatives. Discussions with Leslie Daigle, Juha Hakala, Barry Leiba, Keith Moore, Andrew Newton, and Peter Saint-Andre during the last quarter of 2013 and the first quarter of 2014 were particularly helpful in arriving at the conclusion that a conceptual separation of notions of location-based identifiers (e.g., URLs) and the types of persistent identifiers represented by URNs was necessary. Juha Hakala provided useful explanations and significant working text about the needs of the library community and their perception of identifiers and consequent implications for URN structure. Peter Saint-Andre provided significant text in a pre-publication review. The author also appreciates the efforts of several people, notably Tim Berners-Lee, Leslie Daigle, Larry Masinter, Keith Moore, Juha Hakala, Julian Reschke, Lars Svensson, Henry S. Thompson, and Dale Worely, to challenge text and ideas and demand answers to hard questions. Whether they agree with the results or not, their insights have contributed significantly to whatever clarity and precision appears in the present document.

The specification was changed considerably and its focus narrowed after an extended discussion at the WG meeting during IETF 90 in July 2014 [IETF90-URNBISWG] and subsequent comments and clarifications on the mailing list [URNBIS-MailingList]. The contributions of all of the participants in those discussions, only some of whose names appear above, are gratefully acknowledged.

7. Contributors

Juha Hakala contributed considerable text, some of which was removed from later versions of the document to streamline it.

Contact Information:
Juha Hakala
The National Library of Finland
P.O. Box 15, Helsinki University
Helsinki, MA FIN-00014
Finland
Email: juha.hakala@helsinki.fi
8. IANA Considerations

[[CREF2: RFC Editor: Please remove the first paragraph below before publication.]]

This memo is not believed to require any action on IANA’s part.

There is an existing (i.e. prior to the publication of this document) registry for "Uniform Resource Identifier (URI) Schemes" that already includes the "urn" scheme itself and a separate existing URN Namespace registry. None of those registrations have any specific dependencies on generic URI specifications.

9. Security Considerations

This specification changes the semantics of URNs to make them self-contained (as specified in other documents), relying on the generic URI syntax specification for syntax only. It should have no effect on Internet security unless the use of a definition, syntax, and semantics that are more clear reduces the potential for confusion and consequent vulnerabilities.

10. References

10.1. Normative References


10.2. Informative References

Appendix A. Background on the URN – URI relationship

The Internet community now has many years of experience with both name-type identifiers and location-based identifiers (or "references" for those who are sensitive to the term "identifier" such as many members of the library and information science communities). The primary examples of these two categories are Uniform Resource Names (URNs [RFC2141] [RFC2141bis]) and Uniform Resource Locators (URLs) [RFC1738]). That experience leads to the conclusion that it is impractical to constrain URNs to the high-level semantics of URLs. The generic syntax for URIs [RFC3986] is adequately flexible to accommodate the perceived needs of URNs, but the specific semantics associated with the URI syntax definition -- what particular constructions "mean" and how and where they are interpreted -- appear to not be. Generalization from URLs to generic Uniform Resource Identifiers (URIs) [RFC3986], especially to name-based, high-stability, long-persistence, identifiers such as many URNs, has failed because the assumed similarities do not adequately extend to all forms of URNs. Ultimately, locators, which typically depend on particular accessing protocols and a specification relative to some physical space or network topology, are simply different creatures from long-persistence, location-independent, object identifiers. The syntax and semantic constraints that are appropriate for locators are either irrelevant to or interfere with the needs of resource names as a class. That was tolerable as long as the URN system didn’t need additional capabilities (over those specified in RFC 2141) but experience since RFC 2141 was published has shown that they are, in fact, needed.

Appendix B. Change Log

[[CREF3: RFC Editor: Please remove this appendix before publication.]]


- Revised Section 1 slightly and added some new material to try to address questions raised on the mailing list.
- Added Section 2, reflecting an email exchange.
- Added a Security Considerations section, replacing the placeholder in the previous version.
- Added later-deleted Appendix B and inserted a note in the material titled "A Perspective on Locations and Names" pointing to it (that
material was removed from draft-ietf-urnbis-semantics-clarif-01, but was Section 2 and then Section 3 in earlier versions).

o Added temporary Appendix B for this version only.

o Enhanced and updated the Acknowledgments section.

o The usual small clarifications and editorial changes.

B.2. Changes from draft-ietf-urnbis-urns-are-not-uris-01 to draft-ietf-urnbis-semantics-clarif-00 (2014-08-25)

o Changed title and file name to better reflect changes summarized below. Note that the predecessor of this document was draft-ietf-urnbis-urns-are-not-uris-01.

o Revised considerably as discussed on the mailing list and at IETF 90. In particular, the document has been narrowed to change semantics only without affecting the relationship to URI syntax and the document title and other details changed to match.

o Dropped much of the original Introduction (moving it temporarily to an appendix) and trimmed the abstract to be consistent with the new, more limited scope.

o Revised an earlier version of Appendix B to make "perceived requirement" more clear.

o Removed the former Appendix B, as promised in the previous draft, moved considerably more text into appendices, and added some new appendix text.

o Added new material to discuss the next round of decisions the WG will have to make, assuming this provisions of this specification are approved. That material was removed from draft-ietf-urnbis-semantics-clarif-01.

B.3. Changes from draft-ietf-urnbis-semantics-clarif-00 (2014-08-25) to -01

o Removed some appendices and the topic discussion material, as discussed in the previous draft.

o Aligned the document and its terminology somewhat better with draft-ietf-urnbis-rfc2141bis-urn-09 including providing for p-components and using the p-/q-/f-component terminology.
o Made several clarifying changes to reflect mailing list
discussions (mostly of 2141bis) since the earlier version was
posted.

o Revised earlier portions of this change tracking appendix to
remove referenced to deleted material. It is not possible to
reconstruct what earlier versions of this document contained by
examining these change summaries.

o Moved specific comments about the IETF 90 discussions to
Acknowledgments and removed or edited some material that was only
appropriate for a discussion piece.

o Made several small editorial changes as usual.

B.4. Changes from draft-ietf-urnbis-semantics-clarif-01 (2015-02-14) to
-02

o Reissued to keep draft alive; no substantive changes.

o Updated references, including some that were already outdated in
-01.

B.5. Changes from draft-ietf-urnbis-semantics-clarif-02 (2015-08-10) to
-03

o Made a few substantive updates to reflect evolution in 2141bis,
particularly the elimination of p-component as a separate category
and the added distinction between q-component and r-component..

o Updated references and made several minor editorial changes to
improve clarity.

Author’s Address

John C Klensin
1770 Massachusetts Ave, Ste 322
Cambridge, MA 02140
USA

Phone: +1 617 245 1457
Email: john-ietf@jck.com