

Network Working Group
Internet-Draft
Intended status: Informational
Expires: February 6, 2016

J. Levine
Taughannock Networks
August 5, 2015

Assigning Digital Object Identifiers to RFCs
draft-iab-doi-05

Abstract

We describe the way that Digital Object Identifiers (DOIs) are assigned to past and future RFCs. The DOI is a widely used system that assigns unique identifiers to digital documents that can be queried and managed in a consistent fashion.

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 February 6, 2016.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Structure and resolution of DOIs	3
3.	DOIs for RFCs	4
4.	The process of assigning DOIs	4
4.1.	Getting a DOI prefix	4
4.2.	Retroactively assigning DOIs	5
4.3.	Assigning DOIs to new RFCs	5
4.4.	Use of DOIs in RFCs	5
5.	Internationalization	6
6.	Informative References	6
Appendix A.	Change Log	7
A.1.	Changes from -04 to -05	7
A.2.	Changes from -03 to -04	7
A.3.	Changes from -02 to -03	7
A.4.	Changes from -01 to -02	7
A.5.	Changes from -00 to -01	7
	Author's Address	7

[1.](#) Introduction

The Digital Object Identifier (DOI) system assigns unique identifiers to digital documents that can be queried and managed in a consistent fashion. The structure of DOIs is defined by ISO 26324:2012 [[ISO-DOI](#)] and is implemented by a group of registration agencies coordinated by the International DOI Foundation.

Each DOI is associated with bibliographic metadata about the object, including one or more URIs where the object can be found. The metadata is stored in a public database with entries retrieved via HTTP.

DOIs are widely used by publishers and consumers of technical journals and other technical material published online.

``Typical web addresses are unreliable for locating online resources, because they can move, change or disappear entirely. But persistent identifiers are fixed, with an infrastructure that allows for the location of the item to be updated. The result is that the identifier can provide persistent access to the data. DataCite provides such a service, and DOIs (used by DataCite) were by far the identifier most commonly mentioned by interviewees, closely followed by Handles (on which the DOI system is built). There was a keen preference for DOIs from interviewees because this is a system already used and understood by publishers for traditional publications and so the barrier to uptake would presumably be lower than for an entirely novel system.''

Levine

Expires February 6, 2016

[Page 2]

(From [CITABILITY], page 15, citations omitted.)

Some scholarly publishers accept DOIs as references in published documents, and some versions of bibtex can automatically retrieve the bibliographic data for a DOI and format it. DOIs may have other advantages, such as making it easier to find the free online versions of RFCs rather than paywalled copies when following references or using some document indexes.

The benefits of DOIs apply equally to documents from all of the RFC submission streams, so all RFCs are assigned DOIs.

2. Structure and resolution of DOIs

DOIs are an application of the Handle system defined by RFCs [RFC3650], [RFC3651], and [RFC3652]. A DOI for an RFC might be

10.17487/rfc1149

The first part of a DOI is the number 10, which means a DOI within the handle system, a dot, and a unique number assigned to a publisher, in this case 17487. This part is the DOI prefix. Following that is a slash and a text string assigned by the publisher, called the DOI suffix.

DOIs are treated as opaque identifiers. The DOI suffixes assigned to RFCs are currently based on the doc-id field of the XML index rfc-index.xml, but those for future RFCs might be based on something else if circumstances change. Hence, the reliable way to find the DOI for an RFC is to not to guess, but to look it up in the RFC index, or in the bibxml entries generated from the index.

Although the handle system has its own protocol described in [RFC3652], the usual way to look up a DOI is to use web lookup. A proposed "doi:" URN was never widely implemented, so the standard way to look up a DOI is to use the public http proxy at <https://dx.doi.org>. The example DOI above could be looked up at:

<https://dx.doi.org/10.17487/rfc1149>

Whenever a publisher assigns a DOI, it provides the bibliographic metadata for the object (henceforth called a document, since that is what they are in this context) to its registration agency which then makes it available to clients that look up DOIs. The document's metadata is typically uploaded to the registration agency in XML using an HTTP based API. Users or publishing software can retrieve the metadata by fetching the DOI's URL and using standard HTTP

content negotiation to request application/citeproc+json, application/rdf+xml, or other bibliographic formats.

Publishers have considerable flexibility as to what resides at the URI(s) that a DOI refers to. Sometimes it's the document itself, while for commercial publishers it's typically a page with the abstract and bibliographic information, and some way to buy the actual document. Since some RFCs are in multiple formats (e.g., Postscript and text), an appropriate URI is that of the RFC Editor's info page that has the document's abstract and links to the document(s) in various formats. Hence the URI above when requested as text/html redirects to:

<https://www.rfc-editor.org/info/rfc1149>

More information on the structure and use of DOIs is in the DOI Handbook [[DOI-HB](#)].

3. DOIs for RFCs

With DOIs assigned to each RFC, it is useful to include DOI information in the XML bibliography as a "seriesInfo" item, so that rendering engines can display it if desired. Online databases and indexes that include RFCs should be updated to include the DOI, e.g., the ACM Digital Library. (A practical advantage of this is that the DOI would link directly to the RFC Editor, rather than perhaps to a copy of an RFC behind a paywall.)

Since RFCs are immutable, existing RFCs still wouldn't mention their own DOIs within the RFC itself, but putting their DOIs into indexes would provide value.

4. The process of assigning DOIs

There are three phases to assigning DOIs to RFCs: getting a DOI prefix, retroactively assigning DOIs to existing documents, and updating the publication process to assign DOIs as new RFCs are published.

4.1. Getting a DOI prefix

There are ten registration agencies [[DOI-RA](#)] that assign DOI prefixes. Most of them serve specialized audiences or limited geographic areas, but there are a few that handle scholarly and technical materials. The RFC Editor chose Crossref, an agency widely used by journal publishers. All registration agencies charge for DOIs to defray the cost of maintaining the metadata databases. Crossref publishes its price list; the prices are on the order of

\$660/year for membership, and deposit fees of 15 cents per document for a bulk upload of the backfile (the existing RFCs), and \$1/per document to deposit them as they are published.

The RFC Editor's DOI prefix is 10.17487.

4.2. Retroactively assigning DOIs

Other than paying the deposit fees, assigning DOIs to all of the existing RFCs was primarily a software problem. We updated the RFC Production Center's internal database to include a DOI field for each RFC, changed the schema for the XML index rfc-index.xml to include a DOI field, and updated the scripts that create the XML and text indexes to include the DOI for each RFC. A specialized DOI submission script extracted the metadata for all of the RFCs from the XML index and submitted it to the registration agency using the agency's online API.

4.3. Assigning DOIs to new RFCs

As RFCs are published, the publication software assigns a DOI to each new RFC. The submission script extracts the metadata for new RFCs from the XML index and submits the information for new RFCs to the registration agency.

4.4. Use of DOIs in RFCs

The DOI agency requests that documents that are assigned DOIs in turn include DOIs when possible when referring to other organizations' documents. DOIs can be listed using the existing seriesInfo field in the xml2rfc reference entity, and authors are requested provide DOIs for non-RFC documents when possible. The RFC production center might add missing DOIs when it's easy to do so, e.g., when the same reference with a DOI has appeared in a prior RFC, or a quick online search finds the DOI. With DOIs in the xml2rfc reference databases, DOIs in references from citation libraries can appear in the RFCs automatically.

The RFC Style Guide will be updated to describe the rules for including DOIs in the References sections of RFCs.

Since it is usually possible to retrieve the bibliographic information for a document from its DOI (as bibtex can do, described above), it might also be worth adding this feature to xml2rfc, so a reference with only a DOI could be automatically fetched and expanded.

5. Internationalization

Adding DOIs presents no new internationalization issues.

Since DOIs are opaque, the characters used in any particular DOI are unimportant beyond ensuring that they can be represented where needed. The Handle system says they are UTF-8 encoded Unicode, but in practice all DOIs appear to use only printable ASCII characters. The metadata for each RFC is uploaded as UTF-8 encoded XML.

6. Informative References

[CITABILITY]

Kotarski, R., Reilly, S., Schrimpf, S., Smit, E., and K. Walshe, "Report on best practices for citability of data and on evolving roles in scholarly communication", 2012, <http://www.stm-assoc.org/2012_07_10_STM_Research_Data_Group_Data_Citation_and_Evolving_Roles_ODE_Report.pdf>.

[DOI-HB] International DOI Foundation, "DOI Handbook", DOI 10.1000/182, April 2012, <<http://www.doi.org/hb.html>>.

[DOI-RA] International DOI Foundation, "DOI Registration Agencies", July 2013, <http://www.doi.org/registration_agencies.html>.

[ISO-DOI] International Organization for Standardization (ISO), "ISO 26324:2012 Information and documentation -- Digital object identifier system", 2012, <http://www.iso.org/iso/catalogue_detail?csnumber=43506>.

[RFC3650] Sun, S., Lannom, L., and B. Boesch, "Handle System Overview", [RFC 3650](#), DOI 10.17487/RFC3650, November 2003, <<http://www.rfc-editor.org/info/rfc3650>>.

[RFC3651] Sun, S., Reilly, S., and L. Lannom, "Handle System Namespace and Service Definition", [RFC 3651](#), DOI 10.17487/RFC3651, November 2003, <<http://www.rfc-editor.org/info/rfc3651>>.

[RFC3652] Sun, S., Reilly, S., Lannom, L., and J. Petrone, "Handle System Protocol (ver 2.1) Specification", [RFC 3652](#), DOI 10.17487/RFC3652, November 2003, <<http://www.rfc-editor.org/info/rfc3652>>.

Appendix A. Change Log

Remove this section before publication, please.

A.1. Changes from -04 to -05

Editorial clarifications, reorganize first part and add quote from [\[CITABILITY\]](#). Add i18n section.

A.2. Changes from -03 to -04

Make the rest of everything present tense. Fix typos, note that RSE style guide will include use of DOIs.

A.3. Changes from -02 to -03

Make everything present tense, minor adjustments to reflect reality.

A.4. Changes from -01 to -02

Clarify submission process, multi-document DOIs. Note all streams treated the same. Remove unused reference.

A.5. Changes from -00 to -01

DOI in the xml, not necessarily in the text

Use of DOI in RFCs section.

Author's Address

John Levine
Taughannock Networks
PO Box 727
Trumansburg, NY 14886

Phone: +1 831 480 2300
Email: standards@taugh.com
URI: <http://jl.ly>

