Internet Architecture Board (IAB)
Request for Comments: 7669
Category: Informational
ISSN: 2070-1721

J. Levine
Taughannock Networks

October 2015

Assigning Digital Object Identifiers to RFCs

Abstract

This document describes 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 document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Architecture Board (IAB) and represents information that the IAB has deemed valuable to provide for permanent record. It represents the consensus of the Internet Architecture Board (IAB). Documents approved for publication by the IAB are not a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7669.

Table of Contents

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.

Page 15 of [CITABILITY] indicates that (note that citations have been omitted):

> 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.

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]. For example, a DOI for an RFC might be as follows:

10.17487/rfc1149

The first part of a DOI is the number 10, which means a DOI within the Handle System, followed by 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 RFC index in XML (rfc-index.xml), but the suffix of future RFCs might be based on something else if circumstances change. Hence, the reliable way to find the DOI for an RFC is not to guess, but to look it up in the RFC index or on the RFC Editor website <https://www.rfc-editor.org/>. RFC references created from entries in the usual bibxml libraries will have DOIs included automatically.

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 that 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) to which a DOI refers. Sometimes it's the document itself, while for commercial publishers it's typically a page with the abstract, bibliographic information, and some way to buy the actual document. Because 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 fetched via an HTTP request that accepts 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 don't mention their own DOIs within the RFCs themselves, 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. All registration agencies charge for DOIs to defray the cost of maintaining the metadata databases.

The RFC Editor chose CrossRef, an agency widely used by journal publishers. The prices associated with CrossRef membership are on the order of $660.00 per year for membership, deposit fees of $0.15 cents per document for a bulk upload of the backfile (the existing RFCs), and $1.00 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. The RFC Production Center's internal database was updated to include a DOI field for each RFC, the schema for rfc-index.xml was updated to include a DOI field, and the scripts that create the XML and text indexes were updated 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. Where the citation libraries include DOIs, the output (references created from those citation libraries) will include DOIs.

The RFC Style Guide [RFC-STYLE] has been updated to describe the rules for including DOIs in the References sections of RFCs.

4.5. Possible Future Work

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. Security Considerations

The DOI system adds a new way to locate RFCs and a bibliographic database containing a description of each RFC. The existing locations and bibliographic info are essentially unchanged, so there is no new dependency on the DOI system.

Were CrossRef or the DOI database to suffer a security breach, it is hypothetically possible that users would be directed to locations other than the RFC Editor's web site or would retrieve incorrect bibliographic data, but the actual RFCs would remain intact.

7. 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 2015, <http://www.doi.org/registration_agencies.html>.

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

[RFC-STYLE] RFC Editor, "RFC Editor Style Guide", <https://www.rfc-editor.org/styleguide/>.

[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>.

IAB Members at the Time of Approval

Jari Arkko (IETF Chair)
Mary Barnes
Marc Blanchet
Ralph Droms
Ted Hardie
Joe Hildebrand
Russ Housley
Erik Nordmark
Robert Sparks
Andrew Sullivan (IAB Chair)
Dave Thaler
Brian Trammell
Suzanne Woolf

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