
Workgroup: Network Working Group
Internet-Draft: draft-lemieux-doi-uri-scheme-00
Published: 29 March 2024
Intended: Informational
Status: 30 September 2024
Expires: P.-A. Lemieux, Ed.
Author: Sandflow Consulting LLC

The "doi" URI Scheme

Abstract

This document specifies the "doi" URI scheme, as specified in [RFC3986], for Digital Object Identifier (DOI) names.

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 <https://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 30 September 2024.

Copyright Notice

Copyright (c) 2024 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
2. Syntax	3
3. Resolution	3
4. References	5
4.1. Normative References	5
4.2. Informative References	6
Author's Address	6

1. Introduction

A DOI name is a global unique identifier of a referent, which can be any digital, physical or abstract entity, including inventions, literary and artistic works, ideas, symbols, names, images, designs, etc. DOI names are, for example, widely used to identify academics publications. The DOI system is specified in [\[iso26324\]](#) and [\[doi-handbook\]](#), with the former offering regular formal snapshot of the latter.

EXAMPLE 1: The DOI name "10.1103/PhysRevLett.59.381" refers to the article Per Bak, Chao Tang, and Kurt Wiesenfeld, "Self-organized criticality: An explanation of the 1/f noise", Phys. Rev. Lett. 59, 381.

A DOI name is persistent over time. This persistence is provided by the independence of the DOI name from the referent itself and its descriptive elements. These descriptive elements of a referent, including location and ownership, can change over time, and their current values are retrieved by resolving the DOI name. The set of elements retrieved by resolving a DOI name is called the DOI record. The DOI record usually contains a web address (or URL for Unified Resource Locator) representing an instance of the referent, and may contain services such as email, and one or more items of data about the referent (metadata). The DOI name resolution process uses the Handle System specified at [\[RFC3650\]](#).

This document specifies a URI scheme for DOI names. This scheme formalizes the notation "doi:<DOI name>", which is in widespread use. When resolved, the URI corresponding to a DOI name yields the DOI record associated with the name.

EXAMPLE 2: "doi:10.1103/PhysRevLett.59.381" is the URI corresponding to the DOI name above.

This document intended to satisfy the guidelines and registration procedures specified at [\[RFC7595\]](#).

2. Syntax

As specified at [iso26324], a DOI name consists of an ordered sequence of Unicode code points of the Graphic type.

The scheme-specific-part of the URI corresponding to a given DOI name, i.e., the DOI Name URI, SHALL be equal to the result of the following ordered sequence of steps:

1. express the ordered sequence of Unicode code points that comprise the DOI name as a UTF-8 String, as defined at [iso10646], without the byte order mark and without any normalization;
2. percent-encode any byte in the UTF-8 String that is neither unreserved nor equal to "/".

The URI shall contain neither a query component nor a fragment component.

EXAMPLE 1: The DOI name "10.5594/SMPTE.ST2067-21.2020" corresponds to the URI <doi:10.5594/SMPTE.ST2067-21.2020>.

EXAMPLE 2: The DOI name "10.26321/Á.GUTIÉRREZ.ZARZA.02.2018.03" with the code point sequence <U+0031, U+0030, U+002E, U+0032, U+0036, U+0033, U+0032, U+0031, U+002F, U+00C1, U+002E, U+0047, U+0055, U+0054, U+0049, U+00C9, U+0052, U+0052, U+0045, U+005A, U+002E, U+005A, U+0041, U+0052, U+005A, U+0041, U+002E, U+0030, U+0032, U+002E, U+0032, U+0030, U+0031, U+0038, U+002E, U+0030, U+0033> corresponds to the URI <doi:10.26321/%C3%81.GUTI%C3%89RREZ.ZARZA.02.2018.03>.

NOTE 1: The sequence of code points comprising a DOI name is not normalized and equivalence between DOI names is based on code points. For example, two DOI names that differ only in the abstract character "Á" being encoded as <U+00C1> in the first and as <U+0041, U+0301> in the second are not identical.

NOTE 2: Presenting a DOI name by rendering its sequence of code points to glyphs can be ambiguous since multiple code points or sequences of code points can result in the same glyphs. For example, U+002D HYPHEN-MINUS, U+2212 MINUS SIGN and U+2013 EN DASH are rendered as similar glyphs. As another example, the abstract character "á" can be represented by either the code point U+00E1 or the sequence of code points <U+0061, U+0301>. Presenting a DOI name in its URI form resolves this ambiguity.

NOTE 3: DOI names are case-insensitive only when testing for equivalence and only with respect to the Basic Latin Unicode block.

3. Resolution

Resolving a DOI name means retrieving its DOI record, which contains the descriptive elements associated with the referent identified by the DOI name.

A DOI name URI can be resolved by performing an HTTP GET request at the following URL (expressed using ABNF syntax as defined at [RFC5234]):

url-scheme "://" "doi.org" "/api/handles/" scheme-specific-part

where

- scheme-specific-part is the scheme-specific-part of the DOI name URI, as defined at [Section 2](#); and
- scheme is either "https" or "http".

The body of the response is a JSON object, as defined at [\[RFC8259\]](#), that contains the following members:

responseCode

The property is a Number. The following values are defined:

- 1 The resolution completed successfully. The HTTP response status code is 200.
- 2 The resolution did not complete successfully because of a server error. The HTTP response status code is 500.
- 100 The DOI name was not found. The HTTP response status code is 404.
- 200 No descriptive elements were found for the requested DOI name. The HTTP response status code is 200.

handle

The property is a String. It is equal to the DOI name for which resolution was requested.

values

The property is an Object. It contains the descriptive elements for the referent identified by the DOI name. The contents of the property are specified at [\[RFC3651\]](#).

[Figure 1](#) illustrates the DOI record, at the time of this writing, for the DOI name corresponding to the URI <doi:10.1000/182>. The DOI record was retrieved by performing an HTTP GET request to <https://doi.org/10.1000/182>.

```
{
  "responseCode": 1,
  "handle": "10.1000/182",
  "values": [
    {
      "index": 1,
      "type": "URL",
      "data": {
        "format": "string",
        "value": "http://www.doi.org/hb.html"
      },
      "ttl": 86400,
      "timestamp": "2004-01-21T14:14:17Z"
    },
    {
      "index": 100,
      "type": "HS_ADMIN",
      "data": {
        "format": "admin",
        "value": {
          "handle": "0.na/10.1000",
          "index": 200,
          "permissions": "011111110010",
          "legacyByteLength": true
        }
      },
      "ttl": 86400,
      "timestamp": "2000-06-23T15:17:46Z"
    }
  ]
}
```

Figure 1: DOI record for the DOI name "10.1000/182" (at the time of this writing)

4. References

4.1. Normative References

- [iso26324]** ISO, "ISO 26324, Information and documentation, Digital object identifier system".
- [iso10646]** ISO, "ISO/IEC 10646, Information technology, Universal coded character set (UCS)".
- [RFC3986]** Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, DOI 10.17487/RFC3986, January 2005, <<https://www.rfc-editor.org/info/rfc3986>>.
- [RFC5234]** Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <<https://www.rfc-editor.org/info/rfc5234>>.

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

[RFC8259] Bray, T., Ed., "The JavaScript Object Notation (JSON) Data Interchange Format", STD 90, RFC 8259, DOI 10.17487/RFC8259, December 2017, <<https://www.rfc-editor.org/info/rfc8259>>.

4.2. Informative References

[RFC7595] Thaler, D., Ed., Hansen, T., and T. Hardie, "Guidelines and Registration Procedures for URI Schemes", BCP 35, RFC 7595, DOI 10.17487/RFC7595, June 2015, <<https://www.rfc-editor.org/info/rfc7595>>.

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

[doi-handbook] DOI Foundation, "DOI Handbook", DOI 10.1000/182, <<https://www.doi.org/the-identifier/resources/handbook/>>.

Author's Address

Pierre-Anthony Lemieux (editor)

Sandflow Consulting LLC
San Mateo, CA
United States of America
Email: pal@sandflow.com