doi Foundation

CANDIDATE SPECIFICATION

This specification is offered for public review until October 17, 2024. Please direct your feedback to info@doi.org, including [DOI URI Scheme] in the subject line.

DOI URI SCHEME

COPYRIGHT LICENSE AGREEMENT

TRADEMARK AND LOGO USAGE POLICY

DOI®, DOI.ORG® and shortDOI® are registered trademarks of the DOI® Foundation. The DOI® Foundation authorizes users who correctly implement International Standard ISO 26324 to use the trademark free of charge to indicate such implementation; however, this authorization applies only and exclusively in the context of such use. The DOI® Foundation is willing to authorize anyone who creates software or other products or services implementing ISO 26324 to use the trademark DOI free of charge provided that:

- the software, product or service is accurately described
- DOI is identified as a trademark of the DOI® Foundation
- the superscript symbol ® is placed following the letters "DOI" at its first occurrence in any printed or electronic document describing or marketing the software, product or service

A re-sizable version of the DOI logo for Internet use under these conditions can be downloaded from the DOI® web site as a convenience for companies, organizations and the press. Inclusion of this logo should not be used to imply DOI® Foundation's endorsement of the company, product or services.

For more information, see the Trademark Policy (https://www.doi.org/resources/130718-trademark-policy.pdf).

Publication date: DRAFT

CONTENTS

PREFACE

This document specifies the "doi" URI scheme. It has been prepared and published by the DOI Foundation (https://www.doi.org/), and was subject to consensus review by DOI Foundation members, including its registration agencies.

Chapter 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 academic 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 "10.5240/7481-838B-59CA-63D0-B9A8-E" refers to My Neighbor Totoro, a 1988 film directed by Hayao Miyazaki.

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 name resolution process uses the Handle System specified at [RFC3650], [RFC3651] and [RFC3652], as updated by [DOI-RP].

This document specifies a URI scheme for DOI names. This scheme conforms to the syntax specified at [RFC3986] and formalizes the notation "doi:<DOI name>", which is in widespread use. When dereferenced as detailed in Section 4, the URI corresponding to a DOI name yields the DOI record associated with the name.

> EXAMPLE 2: "doi:10.5240/7481-838B-59CA-63D0-B9A8-E" is the URI corresponding to the DOI name above.

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

Chapter 2
SYNTAX

As specified at [iso26324], a DOI name consists of an ordered sequence of Unicode code points of the Graphic type. The code points are arranged in a DOI prefix and a DOI suffix separated by U+002F SOLIDUS.

A DOI Name URI is a URI that corresponds to a given DOI name. As defined at [RFC7595], its scheme name SHALL be `"doi"` and its scheme-specific-part SHALL conform to the following syntax (expressed using ABNF syntax as defined at [RFC5234]):

```
scheme-specific-part = doi-encoded-prefix "/" doi-encoded-suffix
```

`doi-encoded-prefix` and `doi-encoded-suffix` SHALL be the result of applying the following ordered sequence of steps to, respectively, the DOI prefix and DOI suffix:

1. express the ordered sequence of Unicode code points as a UTF-8 String, as defined at [iso10646], without the byte order mark and without any normalization;
2. for every byte in the UTF-8 String:
    a. output the byte unmodified if the byte is either in the `unreserved` set, in the `sub-delims` set, equal to `":"` or equal to `"@"`;
    b. otherwise, replace the byte with the US-ASCII byte triplet resulting from percent-encoding the byte.

The `unreserved` and `sub-delims` sets are specified at [RFC3986].

A DOI Name 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 "10.6338/JDA.202212/SP_17(4).0000" corresponds to the URI <doi:10.6338/JDA.202212%2FSP_17(4).0000>.

EXAMPLE 3: 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.

Chapter 3

EQUIVALENCE

The following procedure SHALL be performed to determine whether two DOI Name URIs are equivalent:

1. the scheme-specific-part of each of the two URIs is percent-decoded into a UTF-8 String;
2. the two UTF-8 Strings are interpreted as two DOI names;
3. the two DOI Name URIs are equivalent if the two DOI names are equivalent, as defined at [doi-handbook].

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

Chapter 4

DOI NAME 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 used to resolve its corresponding DOI name by performing an HTTP GET request at the following URL (expressed using ABNF syntax as defined at [RFC5234]):

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

where `scheme-specific-part` is defined at Section 2, and the `"https"` scheme is specified at [RFC9110].

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/api/handles/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).

Chapter 5

RETRIEVING THE REFERENT IDENTIFIED BY A DOI NAME

While Section 4 specifies the procedure for retrieving the DOI record associated with DOI name, the steps necessary to retrieve the actual referent described by the record depend on the nature of the referent, e.g., a referent can be a physical object.

Some, but not all, referents can be retrieved by dereferencing an HTTP/HTTPS URI found in their respective DOI records, as illustrated in Figure 1 where the referent identified by the DOI name "10.1000/182" can be retrieved at "http://www.doi.org/hb.html".

The single DOI resolution and multiple doi resolution functions at [doi-handbook] specify the process of retrieving a referent that is available by dereferencing an HTTP/HTTPS URI.

Chapter 6

SECURITY CONSIDERATIONS

A DOI name is an opaque string, which does not have a discernible meaning on its own and is for use by humans and machines alike. It consists of a sequence of Unicode codepoints and the security considerations at [UNICODE-TR36] apply. In particular, and as noted at Section 2, presenting a DOI name by rendering its sequence of code points to glyphs can be ambiguous. As a result, two DOI names rendering to the same sequence of glyphs can identify referents, including, for example, two software executables with wildly different side-effects. Presenting a DOI name in its URI form, which consists of a limited subset of characters, can lessen this risk.

The DOI name resolution process is conducted using the Hypertext Transfer Protocol Secure, which ensures confidentiality and integrity of the transaction, and the security considerations at [RFC9110] apply.

The result of the DOI name resolution process is a JSON object and the security considerations at [RFC8259] apply.

Chapter 7

IANA CONSIDERATIONS

The following is the permanent URI Scheme Registration request, as defined in [RFC7595]:

Scheme name

doi

Status

Permanent

Contact

Pierre-Anthony Lemieux <pal@sandflow.com>

Change controller

DOI Foundation

Web: <https://www.doi.org>

Email: <info@doi.org>

References

This document

Chapter 8

REFERENCES

8.1 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>.

[RFC9110] Fielding, R., Ed., Nottingham, M., Ed., and J. Reschke, Ed., "HTTP Semantics", STD 97, RFC 9110, DOI 10.17487/RFC9110, June 2022, <https://www.rfc-editor.org/info/rfc9110>.

8.2 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>.

[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, <https://www.rfc-editor.org/info/rfc3652>.

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

[DOI-RP] DONA Foundation, "Digital Object Identifier Resolution Protocol Specification", <https://www.dona.net/sites/default/files/2022-06/DO-IRPV3.0--2022-06-30.pdf>.

[UNICODE-TR36] Unicode Consortium, "Unicode Security Considerations", <https://www.unicode.org/reports/tr36/>.