Namespace Registration for Digital Object Identifier (DOI)

Namespace ID: DOI Requested of IANA.

Version: 1

Date: 2020-09-30

Registrant:

Name: Paul Jessop
E-mail: paul@countyanalytics.com
Affiliation: Technology Adviser, The DOI Foundation
Address:
Web URL: https://doi.org/

Background:

A Digital Object Identifier (DOI) name is a "digital identifier of an object" rather than an "identifier of a digital object". DOI names may be used for persistent and unique identification of objects of any type (digital, physical or abstract) and current applications of the DOI system include articles published in scholarly journals, scientific data sets and movie/tv production objects.

The DOI system was initiated by The DOI Foundation in 1998. First applications were launched at 2000. As of this writing, about 190 million DOI names have been assigned to objects managed by more than 5000 organizations. These identifiers are resolved more than 5 billion times per year (see https://www.doi.org/factsheets/DOIKeyFacts.html).

DOI is an international standard under ISO 26324:2012 and ISO has appointed The DOI Foundation as the registration authority. The DOI Foundation publishes the DOI Handbook and revises it from time to time - the current version was published in 2019-12-19. The Handbook is available at https://doi.org/10.1000/182 and gives information on implementation details that are not contained in the international standard, including the resolution system.

The Handle System, as described in RFC 3651 (https://tools.ietf.org/html/rfc3651), has been designated as the DOI resolution system so that DOI names are Handles with additional characteristics specified in ISO 26324, in the DOI Handbook and by the registrant (the DOI Federation-affiliated agency assigning the DOI name). Handles assigned outside this framework are not required to have these characteristics.

It is possible to resolve DOIs using the native Handle resolution system, but this is typically accessed through an https proxy provided by the DOI Foundation with a global, scalable, load balancing caching architecture. This proxy is accessed by prepending "https://doi.org/" and optionally appending a query parameter. Details are in section 3.8 of the DOI Handbook.

NOTE Handle-related standards are now developed in the DONA Foundation (https://www.dona.net/). The DONA Foundation is currently updating the Handle Identifier and Resolution Protocol (IRP) which is specified in RFCs 3650-3652.

Purpose:

Enable presentation of DOI names as URNs.

Benefits:

1. When represented as URNs, DOI names can be easily located within full text.

2. When represented as URNs, DOI names can in the future become technology independent hyperlinks.

3. The F- and Q-components (as specified in the RFC 8141 URN syntax) may in future be specified to work with DOI names represented as URNs. Such specification would be documented in the DOI Handbook and implemented in the DOI https proxy.

4. Similarly, the R-component could be specified to send requests to the resolver when it is presented with a DOI name represented as a URN.

Character set:

DOI names may incorporate any printable characters from ISO/IEC 10646, the ISO equivalent of the Unicode character set.

Handle syntax imposes two constraints on the prefix — both slash and dot are "reserved characters", with the slash separating the prefix from the suffix and the dot used to extend sub prefixes.

When DOI names are presented as HTTP URIs, they are prefixed by the DOI proxy resolver address https://doi.org (formerly also http://dx.doi.org). However, the resolver address is not part of the identifier string.

URI percent encoding (see RFC 3986, section 2.1) is required for characters in a DOI that are:

(a) outside the ASCII printable character set, or
(b) reserved in the URI syntax (see RFC 3986, section 2.2).

The same rule is applied in creating a URN.

Thus, # becomes %23 and the DOI 10.1000/456#789 becomes URI https://doi.org/10.1000/456%23789 and becomes URN urn:doi:10.1000/456%23789.

Similarly, Unicode characters other than ASCII printable characters SHALL be percent encoded when a URN is created from a DOI.

Details of this process are contained in the DOI handbook.

NOTE In practice, most DOIs use only printable ASCII characters.

Syntax:

The DOI Namespace-Specific String (NSS) consists of three parts:

- DOI prefix;
- a forward slash (/) as the delimiting character; and,
- DOI suffix.

In the current edition of ISO 26324, the DOI prefix is composed of a directory indicator “10” followed by a registrant code, which is a unique string assigned to a registrant. These two components are separated by a full stop (period).

The DOI Foundation has launched a revision of ISO 26324 to amend the syntax of the prefix to remove the requirement for the directory indicator to be "10" and to allow DOI names without a registrant code.

Both the existing and proposed syntax are compatible with this registration.

EXAMPLE 1 10.1000 DOI prefix comprising a directory indicator "10" followed by registrant code "1000".

The DOI suffix consists of a character string of any length chosen by the registrant. Each suffix is required to be unique within the space defined by the prefix element that precedes it. In some circumstances it incorporates an identifier generated from or based on another system used by the registrant (e.g. ISAN, ISBN, ISRC, ISSN, ISTC, ISNI). Information on incorporating other identifiers is contained in section 2.7 of the DOI Handbook.

EXAMPLE 10.1000/182 DOI name with the DOI prefix 10.1000 and suffix 182.

The following formal definition uses ABNF [RFC5234] and has been adapted from the one used in the EIDR URN namespace registration [RFC 7972].

```
URN-DOI = "urn:doi:" DOI-NSS
DOI-NSS = DOI-PREFIX "/" DOI-SUFFIX
DOI-PREFIX = 1*DOI-CHARS
DOI-SUFFIX = 1*DOI-CHARS
DOI-CHARS = %x20-7E
```

NOTE The composition of the DOI prefix has as a matter of practice always been restricted to DIGIT / "." though there is no formal restriction in ISO 26324 that the registrant code must be numeric. The Handle System similarly is largely restricted to numeric codes in the prefix (though exceptionally the Library of Congress Handle prefix is not numeric). Following revision of the DOI specification the same restriction will in practice apply.

Rules for lexical equivalence:

Since the DOI-NSS is case insensitive with respect to printable ASCII characters, parsers SHALL treat URN:DOIs accordingly and ignore such case when URN:DOI identifiers are compared.

F-, Q- and R-components SHALL be ignored in the comparison since they are not part of the NSS.

ISO 26324 requires that when displayed on screen or in print, a DOI name should be preceded by a lowercase "doi:" unless the context clearly indicates that a DOI name is implied. The "doi:" label is not part of the DOI name and SHALL be ignored in comparison. Although this display syntax is compatible with URI, no scheme registration for DOI has yet been made by the DOI Foundation.

Where comparison is made between different representations of a DOI name, the scheme and namespace of a URN representation, the "doi:" prefix of a display DOI name, the http/https proxy resolver address SHALL be deleted and any percent encoding SHALL be removed before equivalence of two such DOI name representations is analyzed.

EXAMPLE The following representations of a DOI name SHALL be regarded as equivalent:

urn:doi:10.1000/456%23789
https://doi.org/10.1000/456%23789

doi:10.1000/456#789
10.1000/456#789

The EIDR URN namespace is also made up of DOI names, assigned in this case by the Entertainment ID Registry. When comparing a URN:EIDR and a URN:DOI for equivalence of the respective underlying DOI names, it is necessary to note that the syntax in RFC 7972 replaces the "/" between the DOI prefix and the DOI suffix with a ":" because the URN definition at the time RFC 7972 was published did not allow a "/" in a URN Namespace Specific String and the subsequent RFC 8141 does not retrospectively change this.

DOI assignment:

DOI assignment is managed by DOI Registration Agencies which act as registrants. Their roles and responsibilities are set out in chapter 8 of the DOI Handbook (see https://www.doi.org/doi_handbook/8_Registration_Agencies.html).

Rules for assignment of DOI names have been set in chapter 2.3 of the DOI Handbook (see https://www.doi.org/doi_handbook/2_Numbering.html#2.3).

The assignment of a DOI name requires that the registrant provide kernel metadata describing the object to which the DOI name is being assigned. The metadata must describe the object to the degree that is necessary to distinguish it as a separate entity within the DOI system.

The (extensible) DOI kernel metadata schema has been specified in Annex B of ISO 26324.

Security and Privacy:

DOI names do not have any unmanaged privacy issues. Where the referent of a DOI name is a living person, the registration agency assigning the name is responsible for ensuring compliance with privacy and data protection requirements such as the General Data Protection Regulation of the EU and the California Consumer Privacy Act.

Due to the assignment by the DOI Foundation of prefixes and the assignment by DOI registration agencies of suffixes – all of which are aware of the need to manage the potential problem - it is unlikely to be possible to utilize Unicode confusables to create DOI names which may be confused by the users.

Interoperability:

DOI names do not have any known interoperability related problems.

Because the DOI Foundation has designated the Handle System for the resolution of DOI names, they may be resolved by the Handle System proxy resolver at https://hdl.handle.net as well as by the DOI Foundation's own resolver network at https://doi.org.

Where, following the requested revision to ISO 26324, an existing Handle-based identification system becomes part of the DOI system, both the above proxy resolvers will continue to work (as well as native resolution by the Handle System in accordance with RFC 3651).

URNs in the URN:DOI namespace are interoperable with URNs in the URN:EIDR namespace (as defined in RFC 7972), in that the above resolution servers work with each and they share the same kernel metadata schema.

Resolution:

URNs in the URN:DOI namespace SHALL be resolvable. Each resolution of a URN in the URN:DOI namespace SHALL return information about the referent of the DOI name represented.

DOI resolution is described in chapter 3 of the DOI Handbook (see https://www.doi.org/doi_handbook/3_Resolution.html). To resolve a URN in the URN:DOI, the string to the right of "urn:doi:" may be

- resolved in the Handle System in accordance with the protocol defined in RFC 3651 or
- be prepended with "https://doi.org/" and an https GET performed.

Where the retrieved record includes a URL, the resolver will normally provide a redirect to that URL, which may be the referent itself, information about the referent or a "landing page" explaining how the referent may be accessed.

The DOI Handbook contains information on how to use the DOI proxy resolver with various query parameters to, for instance, prevent redirection or to bypass the cache. It also contains information about the use of multiple URLs and multiple resolution, allowing access to specific information stored in the Handle System about a referent.

Persistence:

DOI is a well-managed identifier system.

Persistence of DOI information is a key aim of the DOI system, and is guaranteed by the DOI social infrastructure, policies and agreements. In the event of any Registration Agency ceasing to maintain DOI information, for any reason, the records will be transferred to another Registration Agency. This forms part of the legal agreement which each Registration Agency enters into with the DOI Foundation.

Additional documentation / information:

See ISO 26324 and the DOI Handbook.

Revision Information:

None