

Internet-Draft
Document: [draft-paskin-doi-uri-04.txt](#)
Expires: December 2003

Norman Paskin
International DOI
Foundation
Eamonn Neylon
Manifest Solutions
Tony Hammond
Elsevier
Sam Sun
CNRI
June 2003

The "doi" URI Scheme for the Digital Object Identifier (DOI)

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC 2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at

<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at

<http://www.ietf.org/shadow.html>.

Abstract

This document defines the "doi" Uniform Resource Identifier (URI) scheme for the Digital Object Identifier (DOI). DOIs are identifiers for entities of significance to the content industries. The "doi" URI scheme allows a resource associated with an entity identified by a DOI to be referenced by a URI for Internet applications. A "doi" URI is dereferenced to a set of service descriptions through discoverable resolution mechanisms.

Table of Contents

1	Introduction.....	2
2	Terminology.....	3
3	The "doi" URI Scheme.....	3

The "doi" URI Scheme

June 2003

4	Normalization and Comparison of "doi" URIs.....	5
5	DOI Administration.....	6
6	DOI Resolution.....	6
7	Rationale.....	7
8	Security Considerations.....	8
9	Acknowledgements.....	8
10	References.....	8
11	Authors' Addresses.....	9
12	Full Copyright Statement.....	9

[1](#) Introduction

This document defines the "doi" Uniform Resource Identifier (URI) scheme for the Digital Object Identifier (DOI). DOIs are identifiers for entities of significance to the content industries. The "doi" URI scheme allows a resource associated with an entity identified by a DOI to be referenced by a URI for Internet applications. A "doi" URI is dereferenced to a set of service descriptions through discoverable resolution mechanisms.

The term "Digital Object Identifier" should be construed as meaning an identifier ("Identifier") of an entity ("Object") for use in networked environments ("Digital"). In this sense an "Object" can be any entity - any digital or physical manifestation or performance, or any abstract work or concept - that is identified by a DOI.

Some concepts relevant to DOI follow:

International DOI Foundation (IDF) û The International DOI Foundation, Inc. is a non-stock membership corporation organized in 1997 and existing under and by virtue of the General Corporation Law of the State of Delaware, USA. The Foundation is controlled by a Board elected by the members of the Foundation. The Corporation is a "not-for-profit" organization, i.e. prohibited from activities not permitted to be carried on by a corporation exempt from US federal income tax under [Section 501\(c\)\(6\)](#) of the Internal Revenue Code of 1986 et seq.

The activities of the Foundation are controlled by its members, operating under a legal Charter and formal By-laws. Membership

is open to all organizations with an interest in electronic publishing, content distribution, rights management, and related enabling technologies.

The Foundation was founded to develop a framework of infrastructure, policies and procedures to support the identification needs of the content industries.

DOI Prefix Holder û Any network user who has been assigned the use of a DOI naming authority under which DOIs may be created.

DOI Registration Agency - An IDF-appointed body that provides administration facilities to DOI Prefix Holders.

DOI Resolution û A process of service indirection whereby a service is selected from a set of service descriptions returned on dereference of a "doi" URI and this service subsequently activated.

DOI Service û One or more network services accessible on resolution of a DOI.

DOI Metadata û A set of data associated with a DOI which is deposited into a repository at time of creation by a DOI Registration Agency and thereafter maintained.

2 Terminology

In this document the key words "must", "must not", "required", "shall", "shall not", "should", "should not", "recommended", "may", and "optional" are to be interpreted as described in [RFC 2119](#) [1] and indicate requirement levels for compliant implementations.

3 The "doi" URI Scheme

3.1 Definition of "doi" URI Syntax

The "doi" URI syntax defined in this document conforms to the generic URI syntax. This specification uses the Augmented Backus-Naur Form (ABNF) notation of [RFC 2234](#) [2] to define the URI. The following core ABNF productions are used by this specification as

defined by [Section 6.1 of RFC 2234](#): ALPHA, DIGIT, HEXDIG. The complete "doi" URI syntax is as follows:

```
doi-uri      = scheme ":" encoded-doi [ "?" query ]
              [ "#" fragment ]

scheme       = "doi"

encoded-doi  = prefix "/" suffix

prefix       = segment

suffix       = segment *( "/" segment )
```

```
segment      = *pchar

query        = *( pchar / "/" / "?" )

fragment     = *( pchar / "/" / "?" )

pchar        = unreserved / escaped / ";" /
              ":" / "@" / "&" / "=" / "+" / "$" / ","

unreserved   = ALPHA / DIGIT / mark

escaped      = "%" HEXDIG HEXDIG

mark         = "-" / "_" / "." / "!" / "~" / "*" / "'" /
              "(" / ")"
```

A "doi" URI has an (encoded) DOI as its scheme-specific part followed by an optional query component followed by an optional fragment identifier. A DOI is constructed by appending a unique suffix string to an assigned prefix string separated by a slash "/" character. The prefix is always assigned to a DOI Prefix Holder by a DOI Registration Agency. The DOI Prefix Holder is responsible for the creation of a valid suffix. The prefix in a DOI corresponds to the naming authority. The administration of any particular DOI may be transferred to another party at any time. The prefix does not denote the owner of a DOI.

ANSI/NISO Z39.84-2000 [3] is the authoritative reference that specifies the rules for constructing a DOI. Once constructed, a DOI may be regarded as an opaque identifier with no internal structure. The minimum constraints for validation of a DOI string are that the prefix and suffix components be non-empty.

[3.2](#) Allowed Characters Under the "doi" URI Scheme

The syntax for a DOI is defined in accordance with the ANSI/NISO Z39.84-2000 standard "Syntax for the Digital Object Identifier Syntax". A DOI is represented using the Unicode [4] character set and is encoded in UTF-8 [5].

The "doi" URI syntax uses the same set of allowed US-ASCII characters as specified in [RFC 2396](#) [6] for a generic URI. Reserved characters as well as excluded US-ASCII characters and non-US-ASCII characters must be escaped before forming the URI. Details of the escape encoding can be found in [RFC 2396, section 2.4](#).

[3.3](#) Examples of "doi" URIs

Some examples of syntactically valid "doi" URIs are given below:

(a) doi:alpha-beta/182.342-24

where "alpha-beta" is the prefix and "182.342-24" is the suffix.

(b) doi:10.abc/ab-cd-ef

where "10.abc" is the prefix and "ab-cd-ef" is the suffix.

(c) <rdf:Description about="doi:10.23/2002/january/21/4690"/>

where "10.23" is the prefix and "2002/january/21/4690" is the suffix.

(d) doi:11.a.7/0363-0277(19950315)120%3A5%3C%3E1.0.TX%3B2-V

where "11.a.7" is the prefix and "0363-0277(19950315)120%3A5%3C%3E1.0.TX%3B2-V" is the prefix. Note that

in unescaped form this DOI is represented in UTF-8 as "11.a.7/0363-0277(19950315)120:5<>1.0.TX;2-V".

(e) doi:dk/P%C3%A6dagogi%2037(2),%20562

where "dk" is the prefix and "P%C3%A6dagogi%2037(2),%20562" is the suffix. Note that in unescaped form this DOI is represented in UTF-8 as "dk/P¾dagogi 37(2), 562" and in ISO-Latin-1 as "dk/P¼dagogi 37(2), 562".

[4](#) Normalization and Comparison of "doi" URIs

In order to facilitate comparison of "doi" URIs and to reduce the risk of false negatives, normalization to the canonical form should be applied to minimize the amount of software processing for such comparisons.

The following normalization steps should be applied:

1. Normalize the case of the leading "doi:" token to be lowercase
2. Unescape all unreserved %-escaped characters
3. Normalize the case of the scheme-specific part including any %-escaped characters to be uppercase

The following forms of a "doi" URI

1. DOI:dk/P%C3%A6dagogi%2037(2),%20562
2. doi:DK/P%C3%A6dagogi%2037(2),%20562
3. doi:dk/P%c3%a6dagogi%2037(2),%20562
4. doi:dk/p%c3%a6dagogi%2037(2),%20562
5. doi:dk%2FP%C3%A6dagogi%2037%282%29%2C%20562

are normalized to the canonical form

doi:DK/P%C3%A6DAGOGI%2037(2),%20562

[5](#) DOI Administration

The International DOI Foundation (IDF) is a not-for-profit membership-based organization founded to develop a framework of

infrastructure, policies and procedures to support the identification needs of the content industries. The IDF is the maintenance agency for DOI and appoints DOI Registration Agencies.

DOIs are created by DOI Prefix Holders and must be registered via a DOI Registration Agency. Any network user can become a DOI Prefix Holder by agreement with a DOI Registration Agency.

DOI Registration Agencies perform the following functions: allocating DOI prefixes, registering DOIs, and providing the necessary infrastructure to allow DOI Prefix Holders to declare and maintain the metadata associated with a particular DOI. DOI Registration Agencies also maintain knowledge of the current owner of each individual DOI to ensure administrative updates.

The IDF maintains the DOI system (to allow registration and ensure resolution of DOIs) and provides governance to ensure appropriate use. DOI assignment requires a fee to ensure that the system costs are met. This allows the system to be managed and supports persistence as a function of organization rather than technology. The fee is for the registering of DOIs (and may optionally be passed on to registrants, waived or subsidized by a DOI Registration Agency), but not for the resolution of a DOI.

The DOI system relies on copyright and trademark law to protect the DOI brand and reputation. DOI is not a patented system; the IDF has not developed any patent claims on the DOI system and does not rely on patent law for remedy.

[6](#) DOI Resolution

A "doi" URI references a set of service descriptions which is returned on dereference of the URI. Following such a dereference a service description is typically selected and the corresponding

service activated. This process of service indirection is commonly referred to as "resolution" a DOI. Examples of services that can be accessed by the resolution of a DOI include redirection to another network resource, return of a metadata record describing the entity identified by the DOI, etc. A discussion of such services is beyond the scope of this document.

Resolution of a DOI can be accomplished using a variety of network protocols. The combination of a network protocol, an access method defined by that protocol and a service endpoint provides the means of access to a resolution mechanism. As the maintenance agency for DOI, the IDF will publish the means of access for known resolution mechanisms of DOI. For the use of other resolution mechanisms prior knowledge of the means of access is required.

As such a "doi" URI can be classified both as a name and a locator. The locator references a set of service descriptions. Note that this locator must not be confused with the locator used to retrieve the ultimate representation that may be returned as a result of activating a service. The "doi" URI is thus an instance of an application-level URI and requires a methodology for mapping from the "doi" URI to a proxy locator URI in order to realize its locator role. These mapping methodologies provide the resolution mechanisms that enable a "doi" URI to function as a locator of a set of services.

7 Rationale

7.1 Why Create a New URI Scheme for DOI?

Under [RFC 2718](#), "Guidelines for new URL Schemes" [7], it is stated that a URI scheme should have a "demonstrated utility", and in particular should be applied to "things that cannot be referred to in any other way". DOI meets both of these criteria in that it is a well established identifier (see <http://www.doi.org/>) for entities of significance to the content industries, with some 10 million examples in current use on the Internet, and is being widely embraced by the content industries. DOI is not bound to any Internet protocol and so requires its own dedicated URI scheme.

The administration granularity of existing URI schemes typically operates at the authority component level. By contrast DOIs are managed at the individual identifier level. It is for this reason that the DOI prefix is not to be interpreted as an "owner" authority but rather as the "creator" authority. Once created the "doi" URI may be regarded as an opaque identifier with no internal structure.

7.2 Why Not Use a URN Namespace ID for DOI?

[RFC 2396](#) states that a "URN differs from a URL in that it's [sic] primary purpose is persistent labeling of a resource with an identifier". A "doi" URI on the other hand has a dual purpose: both to allow a resource associated with an entity identified by a DOI to be referenced by a URI for Internet applications, as well as to enable access to a set of service descriptions. In this regard a "doi" URI scheme should be considered as being similar to the "tel", "fax" and "modem" URI schemes documented in [RFC 2806](#) [8].

Further the syntactic requirements of the "doi" URI scheme are incompatible with the URN syntax. Specifically the use of optional query component and/or fragment identifier cannot be accommodated by the URN syntax (cf. Sect. 2.3.2, [RFC 2141](#) [9]).

8 Security Considerations

The "doi" URI scheme is subject to the same security considerations as the general URI scheme described in [RFC 2396](#).

Dereference of a "doi" URI to access a set of service descriptions will be subject to the security considerations of the underlying protocol used to access the resource referenced by the "doi" URI.

9 Acknowledgements

The authors acknowledge the contributions of Larry Lannom and Jason Petrone, of the Corporation for National Research Initiatives, to this specification.

The authors are also grateful to Larry Masinter and Martin Duerst for their constructive comments on this specification.

10 References

1. Bradner, S., "Key Words for Use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
2. Crocker, D.H. and Overell, P., "Augmented BNF for Syntax Specifications: ABNF", [RFC 2234](#), November 1997.
3. ANSI/NISO Z39.84-2000, "Syntax for the Digital Object Identifier", ISBN 1-880124-47-5.
4. The Unicode Consortium, "The Unicode Standard", Version 3, ISBN 0-201-61633-5, as updated from time to time by the publication of

The "doi" URI Scheme

June 2003

new versions. (See <http://www.unicode.org/unicode/standard/versions> for the latest version and additional information on versions of the standard and of the Unicode Character Database).

5. Yergeau, F., "UTF-8, A Transformation Format for Unicode and ISO10646", [RFC 2279](#), October 1996.

6. Berners-Lee, T., R. Fielding and L. Manister, "Uniform Resource Identifiers (URI): Generic Syntax", [RFC 2396](#), August 1998.

7. Masinter, L., H. Alvestrand, D. Zigmond and P. Petke, "Guidelines for new URL Schemes", [RFC 2718](#), November 1999.

8. Vaha-Sipila, A., "URLs for Telephone Calls", [RFC 2806](#), April 2000.

9. Moats, R., "URN Syntax", [RFC 2141](#), May 1997.

11 Authors' Addresses

Norman Paskin
The International DOI Foundation
Linacre House, Jordan Hill
Oxford, OX2 8DP, UK
n.paskin@doi.org

Eamonn Neylon
Manifest Solutions
Bicester
Oxfordfordshire, OX26 2HX, UK
eneylon@manifestsolutions.com

Tony Hammond
Elsevier Ltd
32 Jamestown Road
London, NW1 7BY, UK
t.hammond@elsevier.com

Sam Sun
Corporation for National Research Initiatives
1805 Preston White Dr., Suite 100

Reston, VA 20191, USA
ssun@cnri.reston.va.us

[12](#) Full Copyright Statement

Copyright (C) The Internet Society (2003). All Rights Reserved.

Paskin

Expires - December 2003

[Page 9]

The "doi" URI Scheme

June 2003

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Paskin

Expires - December 2003

[Page 10]