

Internet Draft
Document: [draft-paskin-doi-uri-02.txt](#)
Expires: April 2003

Norman Paskin
IDF
Eamonn Neylon
Manifest Solutions
Tony Hammond
Elsevier Science
Sam Sun
CNRI
October 2002

The "doi" URI Scheme for 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 RFC2026](#).

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/1id-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 Digital Object Identifiers (DOIs). The DOI system was developed by the International DOI Foundation (<http://www.doi.org>), an open membership-based organization founded to develop a framework of infrastructure, policies and procedures to support the identification needs of providers of intellectual property. DOI identifiers are persistent across time and unique across network space. The "doi" URI scheme allows a DOI to be referenced by a URI for Internet applications.

Table of Contents

1.	Introduction.....	2
2.	The "doi" URI Scheme.....	3
2.1	"doi" URI Syntax Definition.....	3

2.2	Reserved and Excluded Characters under "doi" scheme.....	3
2.3	Examples of "doi" URIs.....	4
3.	Security Considerations.....	4
4.	Further Information.....	4
5.	Acknowledgements.....	4
	References.....	5
	Author's Addresses.....	5

[1.](#) Introduction

DOI stands for Digital Object Identifier [[DOI](#)], which is a managed identifier of an intellectual property entity across a common business sector. The DOI identifier enables the network retrieval of a set of related services. The DOI identifier is not constrained to a network application context. DOI identifiers have been widely deployed by the publishing industry. This specification defines the "doi" URI scheme for DOI identifiers referenced within Internet applications.

DOI identifiers are globally unique across the URI namespace and persistent over time. A DOI identifier can "be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name" [[RFC1737](#)]. A "doi" URI has associated data related to the entity that the DOI identifies.

The "doi" URI scheme defines a standard way to represent a DOI identifier under URI namespace. A "doi" URI may serve as a pure name or may be de-referenced by a network service. When used as a name, a "doi"-based URI is independent of any service protocol and accordingly, is not network de-referenceable. When used within a network reference (e.g. within a hyperlink), a DOI identifier does not have a native resolution system. It is instead transported using a network protocol to a specific service (e.g. the Handle System [[HS](#)], or a HTTP request to a proxy). Such service requests may also include supplemental query components specific to that service.

DOIs must be registered through an appointed registration agency. The International DOI Foundation, which is the maintenance agency for the DOI, is responsible for the appointment of registration agencies.

The "doi" URI scheme defined in this document conforms to the generic URI syntax as specified in [RFC2396](#) [[RFC2396](#)]. UTF-8 [[UTF-8](#)] encoding is mandated for any DOI transmitted between "doi" user agent and any DOI service. Syntax for DOI identifier within the "doi" scheme is defined in accordance with ANSI/NISO Z39.84

[[NISO39.84](#)] standard for Digital Object Identifier Syntax.

2. The "doi" URI Scheme

2.1 "doi" URI Syntax Definition

```
doi           = scheme ":" doi-identifier
scheme        = "doi"
doi-identifier = prefix "/" suffix
prefix        = chars-no-slash
suffix        = chars
chars-no-slash = 1*(%x00-2E / %x30-FF)
                ; any character of the UCS [ISO10646] of
                ; U+00A0 and beyond, except the '/'
                ; character.

chars         = 1*(%x00-FF)
                ; any character of the UCS [ISO10646] of
                ; U+00A0 and beyond.
```

The prefix is always assigned to a registrant by a registration agency. The registrant is responsible for the creation of a valid suffix. The prefix corresponds to the creator naming authority at the time of construction only. The administration of any particular DOI may be transferred to another party at any time, so the prefix does not denote the administrative ownership of a particular DOI.

NISO Z39.84 is the authoritative reference that specifies the rules for constructing a DOI. Once constructed, a DOI is to be interpreted as an opaque identifier. The minimum constraints for validation of a DOI string are that the prefix and suffix components be non-empty.

2.2 Reserved and Excluded Characters under "doi" scheme

The "doi" syntax abide by the same set of excluded US-ASCII characters as specified in [RFC2396](#). It further reserves the following characters that are used in common service requests that may be used to append information to a DOI in certain circumstances (e.g. adding parameters resolution instructions to a HTTP URL encoded service request):

```
reserved = "?" | "&" | "=" | "#"
```

If the data for a "doi-identifier" component would conflict with the reserved purpose, then the conflicting data must be escaped before forming the URI. Details of the escape encoding can be found in [RFC2396, section 2.4](#).

2.3 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) doi:1.23/2002/january/21/4690

where "1.23" is the prefix and "january/21/4690" is the suffix.

(d) <element xmlns="doi:1.23/2002/january/21/4690">

The acquisition of DOI services can be achieved through the use of other protocols as a proxy to transfer to dedicated networked service components.

Examples of such use are given below:

(e) <http://my.resolver.inc/resolve?id=doi%3Aalpha-beta%2Fmsws>

is an OpenURL [[OPENURL](#)] service request for "doi:alpha-beta/msws".

(f) rtsp://service.net/query?doi%3A10.abc%2Fab%2Fcd%2Fef

is a service request for "doi:10.abc/ab/cd/ef".

3. Security Considerations

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

When DOI values are used in resolution services, retrieval of DOI data will be subject to the security considerations of the underlying protocol used to access the DOI service.

4. Further Information

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

5. Acknowledgements

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

References

- [DOI] The DOI System <http://www.doi.org/>
- [HS] The Handle System <http://www.handle.net/>
- [RFC2396] Berners-Lee, T., R. Fielding and L. Manister, "Uniform Resource Identifiers (URI): Generic Syntax", August 1998.
- [HTTP] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, T. Berners-Lee, "Hypertext Transfer Protocol - HTTP/1.1", January, 1997.
- [NISO39.84] ANSI/NISO Z39.84-2000 Syntax for Digital Object Identifier.
- [OPENURL] OpenURL specification. <http://www.sfxit.com/OpenURL>
- [ISO10646] Information Technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane", ISO/IEC 10646-1:2000.
- [UTF-8] Yergeau, Francois, "UTF-8, A Transformation Format for Unicode and ISO10646", October 1996.
- [RFC1737] K. Sollins and L. Masinter "Functional Requirements for Uniform Resource Names", December 1994.

Author's Addresses

Norman Paskin
The International DOI Foundation
PO Box 233, Kidlington
Oxford, OX5 1XU, UK
n.paskin@doi.org

Eamonn Neylon
Manifest Solutions
John Eccles House, Oxford Science Park
Oxford, OX4 4GP, UK
eneylon@manifestsolutions.com

Tony Hammond
Elsevier Science Ltd
84 Theobald's Road

London WC1X 8RR, 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

