

Syntax for the Digital Object Identifier

Abstract: Defines the composition and order of the unambiguous alphanumeric identifier string in the Digital Object Identifier® (DOI) system used to reference an intellectual property entity in the digital environment.

An American National Standard
developed by the
National Information Standards Organization

Approved: September 30, 2005
by the
American National Standards Institute

Published by the National Information Standards Organization
Bethesda, Maryland



NISO Press, Bethesda, Maryland, U.S.A.

About NISO Standards

NISO standards are developed by the Standards Committees of the National Information Standards Organization. The development process is a strenuous one that includes a rigorous peer review of proposed standards open to each NISO Voting Member and any other interested party. Final approval of the standard involves verification by the American National Standards Institute that its requirements for due process, consensus, and other approval criteria have been met by NISO. Once verified and approved, NISO Standards also become American National Standards.

This standard may be revised or withdrawn at any time. For current information on the status of this standard contact the NISO office or visit the NISO website at:

<http://www.niso.org>

Published by:

NISO Press
4733 Bethesda Avenue, Suite 300
Bethesda, MD 20814
www.niso.org

Copyright © 2005 by the National Information Standard Organization

All rights reserved under International and Pan-American Copyright Conventions. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without prior permission in writing from the publisher. All inquiries should be addressed to NISO Press, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814.

ISSN: 1041-5635

ISBN: 1-880124-68-8

Contents

	Page
1 Introduction	1
1.1 Purpose	1
1.2 Scope.....	1
2 Standards and References.....	1
2.1 Referenced Standard	1
3 Definitions.....	1
4 Format and Characteristics of the DOI	2
4.1 DOI Character Set	2
4.2 Prefix.....	3
4.3 Suffix.....	3
5 Maintenance Agency	3
Appendix A: DOI Specifications.....	4
Appendix B: Designation of Maintenance Agency	5
Appendix C: Examples of Digital Object Identifiers.....	6
Appendix D: Related Standards and References.....	7
Appendix E: Application Issues.....	8

Foreword

(This foreword is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only.)

This standard defines the composition and order of components of the DOI® (Digital Object Identifier) alphanumeric character string used to identify entities of interest in intellectual property management in the digital environment. The syntax of the scheme is such that any existing identifier string can be expressed in a form suitable for use with the DOI system. The DOI system was developed by the International DOI Foundation (<http://www.doi.org>) to provide a framework for managing intellectual content including activities such as linking users with content owners, facilitating electronic commerce, and enabling automated copyright management.

This standard is a revision of ANSI/NISO Z39.84-2000, *Syntax for the Digital Object Identifier*, and is compatible with the syntax described in that earlier document and with any DOIs implemented under that earlier standard.

Overview of the DOI System

A Digital Object Identifier (DOI) is a name (not a location) for an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks.

Unique identifiers are essential for the management of information in any digital environment. Identifiers assigned in one context may be encountered, and may be re-used, in another place (or time) without consulting the assigner, who cannot guarantee that his assumptions will be known to someone else. To enable such *interoperability* requires the design of identifiers to enable their use in services outside the direct control of the issuing assigner. The necessity of allowing interoperability adds the requirement of *persistence* to an identifier: it requires interoperability with the future (i.e., an identifier assigned now may be retrieved and meaningfully used at a future point). Further, since the services outside the direct control of the issuing assigner are by definition arbitrary, interoperability implies the requirement of *extensibility*. Hence DOI is designed as a generic framework applicable to any digital object, providing a structured, extensible means of identification, description, and resolution. The entity assigned a DOI can be a representation of any logical entity.

The DOI system is built using several existing standards-based components which have been brought together and further developed to provide a consistent system. The DOI was developed as a cross-industry, cross-sector, not-for-profit effort managed by an open membership collaborative development body, the International DOI Foundation (IDF) founded in 1998. The DOI is in widespread use, e.g. for scientific articles. DOIs need not be explicitly declared (though this may be useful): e.g. in a Web context a DOI may be used in an http form as a URL (through a proxy server), while retaining the advantages of managed persistence. DOIs may be used to offer an interoperable common system for identification of entities of many forms.

DOI system components

The DOI system comprises several components:

- a specified standard numbering syntax;
- a resolution service (based on the existing Handle System);
- an optional data model incorporating a data dictionary (based on the indecs Data Dictionary; indecs = interoperability of data in e-commerce systems); and

- an implementation mechanism through policies and procedures for the governance and application of DOIs.

This present NISO standard defines only the specified standard numbering syntax. Information on the other components is provided in this forward solely for information and to place the standard syntax in context.

Resolution

The DOI System is an implementation of the Handle System[®]. The Handle System[®], developed by the Corporation for National Research Initiatives (CNRI) is a general-purpose distributed information system designed to provide an efficient, extensible, and secured global name service for use on networks such as the Internet. The Handle System includes an open set of protocols, a namespace, and a reference implementation of the protocols. The protocols enable a distributed computer system to store names, or handles, of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. These associated values can be changed as needed to reflect the current state of the identified resource without changing the handle, thus allowing the name of the item to persist over changes of location and other current state information. Each handle may have its own administrator(s) and administration can be done in a distributed environment. The name-to-value bindings may also be secured, allowing handles to be used in trust management applications.

The Handle System is a protocol specification. An implementation of a Handle System is made up of local handle services (LHS). A local handle service is made up of one or more sites. A site is made up of one or more handle servers. Handle servers store handles.

Within the DOI system, an implementation of the Handle System is used for DOI resolution. Resolution is the process of submitting a DOI to a network service and receiving in return one or more pieces of current information related to the identified object. In the case of the Domain Name System (DNS), as an example, the resolution is from domain name, e.g., www.doi.org, to a single IP address, e.g., 132.151.1.146, which is then used to communicate with that Internet host. In the case of the Handle System, the resolution is from a handle, e.g., 10.1000/140, to one or more pieces of typed data, e.g., three URLs representing three copies of the object.

The Handle System provides a general purpose global system for the reliable management of information on networks such as the Internet over long periods of time and is currently in use in a number of projects. The International DOI Foundation, Library of Congress, the Defense Technical Information Center, D-Space, the Dept of Defense Advanced Distributed Learning Initiative, and the Globus Alliance are some implementers of the Handle System. The DOI System is managed by the International DOI Foundation, which sets policies, appoints service providers, and ensures the successful operation of the System. The IDF maintains a Handbook describing current rules and practices for DOI implementation (<http://dx.doi.org/10.1000/182>).

The resolution of a DOI results in one or more pieces of typed data ("associated values"). Examples of typed data are a URL, an e-mail address, another DOI, and metadata. A DOI can be resolved to an arbitrary number of different associated values. Resolution requests may return all associated values of current information, individual values, or all values of one data type. These associated values can be displayed in a menu in a client application, from which the user may select the desired value. Moreover, a client application may deliver messages to a user wherein the messages are constructed from a syntax that allows the user to select the desired value. In addition these values may be processed automatically to achieve the objective of the resolution system.

Current Web browser technology requires additional functionality to allow the browser to make full use of DOIs: additional browser features are necessary. It is anticipated that features supporting resolution will commonly be built into browsers in the future. A freely

available “resolver plug in” can be downloaded from <http://www.handle.net/resolver/>. The plug-in extends the browser's functionality so that it understands the Handle protocol.

Alternatively, without the need to extend the Web browsers' capability, DOIs may be structured to use the default public DOI proxy server (<http://dx.doi.org>). The resolution of the DOI in this case depends on the use of URL syntax. For example, “doi:10.123/456” would be written as <http://dx.doi.org/10.123/456>

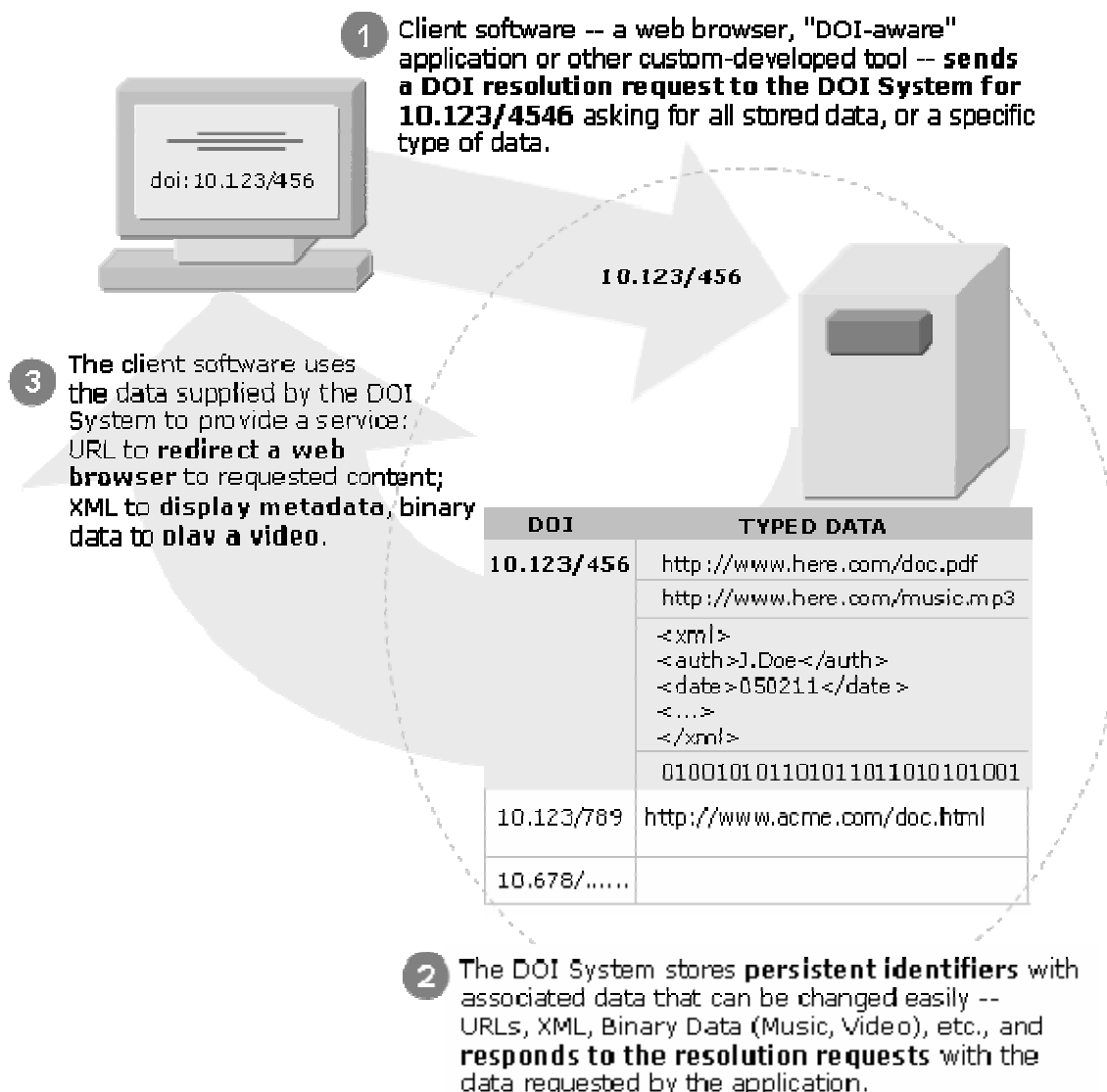


Figure 1: Resolution of a DOI through the DOI system

DOI data model: metadata

The DOI data model consists of a data dictionary and a framework for applying it. Together these provide tools for defining what a DOI specifies (through use of a data dictionary), and how DOIs relate to each other, (through a grouping mechanism, Application Profiles, which associate DOIs with defined common properties). This provides semantic interoperability,

enabling information that originates in one context to be used in another in ways that are as highly automated as possible.

The DOI system uses an interoperable data dictionary built from an underlying ontology. The data dictionary component is designed to ensure maximum interoperability with existing metadata element sets; the framework allows the terms to be grouped in meaningful ways (DOI Application Profiles) so that certain types of DOIs all behave predictably in an application through association with specified Services. This provides a means of integrating the features of Handle resolution with a structured data approach. DOIs need not make use of this data model, but it is envisaged that many will: any DOI intended to allow interoperability (i.e. which has the possibility of use in services outside of the direct control of the issuing Registration Agency) is subject to DOI Metadata policy, which is based on the registration of terms in the iDD.

The identifier syntax string of the DOI contains a set of components with minimal meaning. Each DOI string begins with the string "10" which within the Handle System of the Corporation for National Research Initiatives (CNRI) determines that the string will be further resolved by the Local Handle Systems reserved for the DOI. The next component of the DOI string notes the number assigned to the registrant who originally created the DOI: this number provides no information about current ownership of the object that the DOI string references. The DOI suffix string contains an unspecified identifier. The elements within the DOI string do not include information about the object identified; such information is to be obtained through associated metadata. The form of such metadata may be specified in a particular DOI application.

Metadata associated with a DOI may include names, identifiers, descriptions, types, classifications, locations, times, measurements, relationships and any other kind of information related to an identified entity. In order to enable their use in services outside the direct control of the issuing assigner such metadata must be expressed in a form which is unambiguously defined.

DOI policy places no restrictions on the form and content of a DOI's metadata declarations. DOI Registration Agencies (RAs) may specify their own metadata schemes and messages, or use any existing schemes in whole or part for their input and service metadata declarations. Where metadata is associated with a DOI in a form which is to be made available to others, a Kernel Metadata Declaration (a.k.a. DOI Kernel) should be made available for each DOI issued. Metadata exchanged between organizations supporting DOI services should be exchanged using an agreed DOI Resource Metadata Declaration (RMD). Proprietary terms (data elements and values) used in Kernel and Resource Metadata Declarations should be registered in the IDF's indecs Data Dictionary (iDD).

The DOI Kernel, which is formally specified in an XML schema, provides some basic information about the identified Resource to aid in disambiguation and identification. A DOI Resource Metadata Declaration (RMD) is a message designed specifically for metadata exchange in the form of an XML document which conforms to an XML Schema (xsd). All its elements and allowed values are mapped into the iDD.

The IDF's indecs Data Dictionary (iDD) is the repository for all data elements and allowed values used in Kernel Metadata Declarations and Resource Metadata Declarations. The iDD enables the definition and ontology of all metadata elements to be available, and provides the basis for mappings to support metadata integration and transformations required for data interchange between RAs. iDD is a structured ontology compliant with logical axioms and constructors common to ontology languages such as W3C's OWL (Web Ontology Language). It can, for example, support the production of legal OWL ontologies. All allowed values used by an RA in its Kernel Metadata, and all data elements used by an RA when mapping to an RMD, are registered in the iDD.

Each DOI is associated with one or more Application Profiles (AP), and each AP is associated with one or more defined Services. Any single DOI can be a member of multiple