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

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This standard may be revised or withdrawn at any time.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

Contents

iii

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 Identifie*r, 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
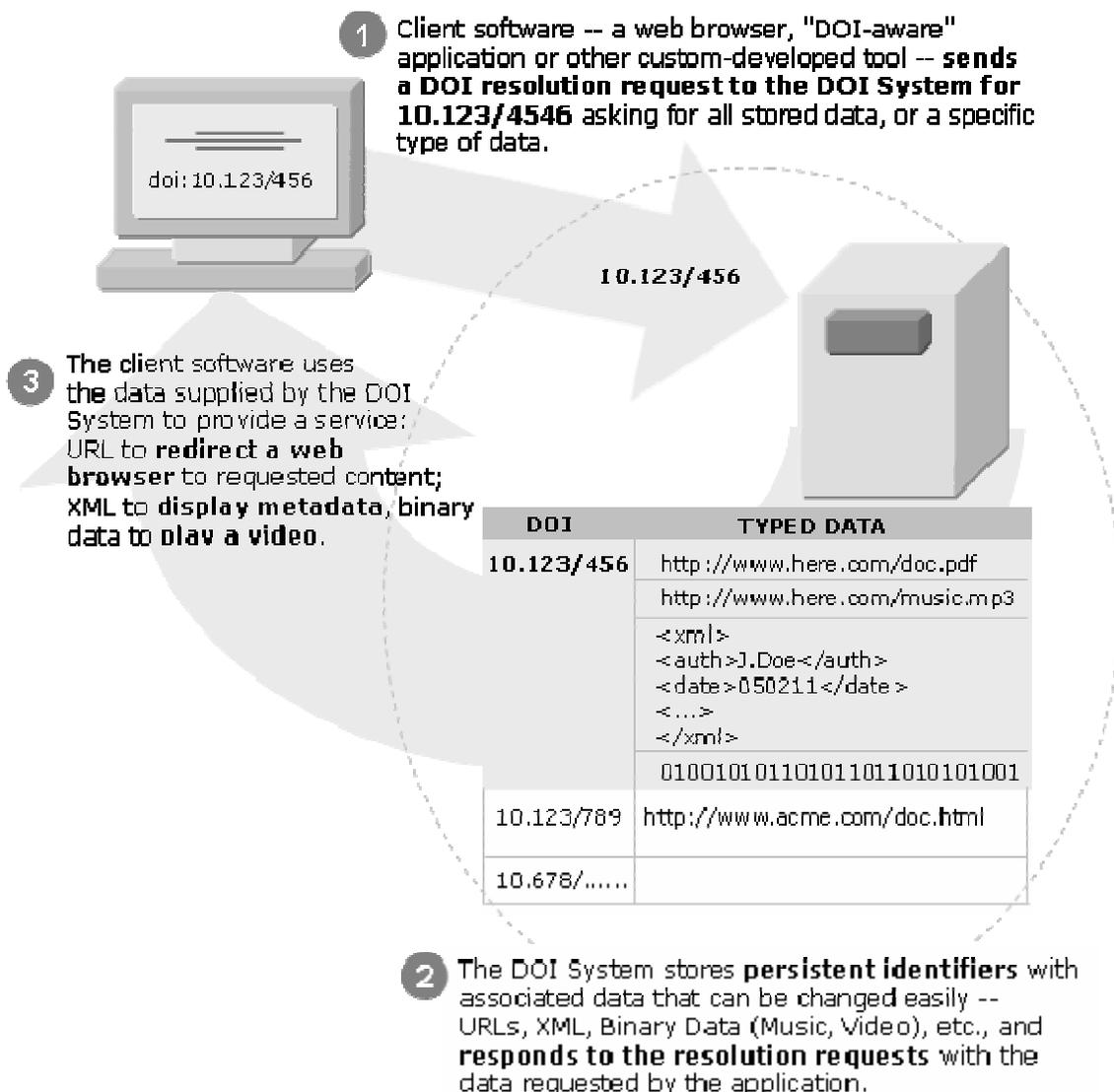
Figure 1: Resolution of a DOI through the DOI system

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

APs. Each AP can be linked to one or more specific instances of a Service. Each defined Service can be made available in multiple ways, referred to as instances. This makes it possible to add a Service to many DOIs by adding that Service to relatively few APs.

This framework is implemented in the handle system, using DOIs for both APs and Services and linking them together through typed handle values. A Service is simply a defined result from a defined action, i.e., do X and the result will be Y.

One of the Application Profiles, possibly the only one for some DOIs, is the zero AP which affirms that there is no structured metadata or services associated with that DOI.

One of the services, possibly the only one for some DOIs, is the provision of kernel metadata for each DOI. Other sets of metadata may also be available for some DOIs and this, as with other services, would be known through the inclusion of a given DOI in an AP and the association of that AP with the given service.

APIs (application programming interfaces) have been created that abstract out the details and make it possible to administer the AP and service structures and to use them in applications while ignoring the details of the Handle implementation. They provide "hooks" down into the handle system without manipulating the handle records directly.

Role of the International DOI Foundation (IDF)

The ongoing management of the DOI System is by the International DOI Foundation, an international, not-for-profit, membership-based organization with offices in the United States and United Kingdom. The Foundation is responsible for licensing a Directory Manager, Registration Agencies and technology providers, for setting policy for the system, and for encouraging development of the related enabling technologies to build the infrastructure for electronic transaction systems such as copyright management. Though the DOI System was originally developed by the publishing industry, it was recognized that the DOI system would have a broader scope and that it should work with established standards bodies as much as possible. NISO participation was requested in early 1998 to develop a standard for the syntax of the DOI identifier string, in order to maximize the broad potential use of a digital object identifier.

DOI is implemented through a federation of Registration Agencies which use policies and tools developed through a parent body, the International DOI Foundation (IDF). The IDF is the governance body of the DOI system, which safeguards (owns or licenses on behalf of registrants) all intellectual property rights relating to the DOI System. It works with RAs and with the underlying technical standards of the DOI components to ensure that any improvements made to the DOI system (including creation, maintenance, registration, resolution, and policymaking of DOIs) are available to any DOI registrant, and that no third party licenses might reasonably be required to practice the DOI standard. DOI resolution is freely available to any user encountering a DOI.

The DOI System has the flexibility to deliver identification and resolution services that fulfill the requirements of any application domain. The rules about what is identified, and whether two things being identified are (or are not) "the same thing" in a given application are made at a lower level: in a specific application of the DOI. This is a role of DOI Registration Agencies. This provides an identification system of enormous flexibility and power while increasing the importance of an explicit structured metadata layer, because without this the identifier essentially can have no meaning at all outside a specific application.

The IDF provides implementation governance. It also provides a technical infrastructure (resolution mechanism, proxy servers, mirrors, back-up, central dictionary) and a social infrastructure (persistence commitments, fall-back procedures, cost-recovery (on a self-sustaining model), and shared use of the system. The IDF is not a standards body, but a central authority and maintenance agency. The IDF is already the appointed registration authority for the ISO/IEC MPEG 21 Rights Data Dictionary, and is to be proposed as the registration authority for the DOI System within ISO TC46/SC9. IDF delegates and licenses

authority to use the system through Registration Agencies, each of which can develop its own applications and use DOI in "own brand" ways appropriate for their community.

Suggestions for improving this standard are welcome. They should be sent to the National Information Standards Organization, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814 USA, telephone (301) 654-2512.

This Standard was processed and approved for submittal to ANSI by the National Information Standards Organization. NISO approval of this Standard does not necessarily imply that all Voting Members voted for its approval. At the time it approved this Standard, NISO had the following members:

Trademarks, Service Marks

Wherever used in this standard, all terms that are trademarks or service marks are and remain the property of their respective owners.

DOI and DOI.ORG are registered service marks of the International DOI Foundation, Inc.

NISO Voting Members

At the time this standard was approved, NISO had the following voting members:

3M
Susan Boettcher
Roger D. Larson (Alt)

American Association of Law Libraries
Robert L. Oakley
Mary Alice Baish (Alt)

American Chemical Society
Matthew Toussant

American Library Association
Betty Landesman

American Society for Information Science and Technology (ASIS&T)
Gail Thornburg

American Society of Indexers
Judith Gibbs

American Theological Library Association
Myron Chace

ARMA International
Diane Carlisle

Armed Forces Medical Library
Diane Zehnpfennig, Emily Court (Alt)

Art Libraries Society of North America (ARLIS/NA)
Mark Bresnan

Association for Information and Image Management (AIIM)
Betsy A. Fanning

Association of Information and Dissemination Centers (ASIDIC)
Margie Hlava

Association of Jewish Libraries
Caroline R. Miller, Elizabeth Vernon (Alt)

Association of Research Libraries
Duane E. Webster, Julia Blixrud (Alt)

Auto-Graphics, Inc.
Paul Cope

Barnes & Noble, Inc.
Douglas Cheney

Book Industry Communication
Brian Green

California Digital Library
Daniel Greenstein, John Kunze (Alt)

Cambridge Information Group
Michael Cairns, Matthew Dunie (Alt)

Checkpoint Systems, Inc.
Douglas Karp, Frank Palazzo (Alt)

The Cherry Hill Company
Cary Gordon

College Center for Library Automation (CCLA)
Richard Madaus, Ann Armbrister (Alt)

Colorado State Library
Brenda Bailey-Hainer, Steve Wrede (Alt)

CrossRef
Edward Pentz, Amy Brand (Alt)

NISO Voting Members (continued)

Davandy, L.L.C.
Michael J. Mellinger

DYNIX Corporation
Ed Riding, Gail Wanner (Alt)

EBSCO Information Services
Gary Coker, Oliver Pesch (Alt)

Elsevier
Paul Mostert

Endeavor Information Systems, Inc.
Sara Randall, Shelley Hostetler (Alt)

Entopia, Inc.
Igor Perisic

ExLibris, Inc
James Steenbergen

Fretwell-Downing Informatics
Robin Murray

Gale Group
Katherine Gruber, Justine Carson (Alt)

Geac Library Solutions
Eric Conderaerts, Eloise Sullivan (Alt)

H.W. Wilson Company
Ann Case, Patricia Kuhr (Alt)

Helsinki University Library
Juha Hakala

Index Data
Sebastian Hammer, David Dorman (Alt)

INFLIBNET Centre
T A V Murthy, Rajesh Chandrakar (Alt)

Infotrieve
Jan Peterson

Innovative Interfaces, Inc.
Gerald M. Kline, Betsy Graham (Alt)

The International DOI Foundation
Norman Paskin

Ithaka/JSTOR/ARTstor
David Yakimischak, Bruce Heterick (Alt)

John Wiley & Sons, Inc.
Eric Swanson

Library Binding Institute
Debra Nolan

The Library Corporation
Mark Wilson, Ted Koppel (Alt)

Library of Congress
Sally H. McCallum

Los Alamos National Laboratory
Richard E. Luce

Lucent Technologies
M.E. Brennan

Medical Library Association
Nadine P. Ellero, Carla J. Funk (Alt)

MINITEX
Cecelia Boone, William DeJohn (Alt)

Modern Language Association
Daniel Bokser, B. Chen (Alt)

Motion Picture Association of America (MPAA)
Michael Owens

MuseGlobal, Inc.
Kate Noerr, Clifford Hammond (Alt)

Music Library Association
Mark McKnight, David Sommerfield (Alt)

National Agricultural Library
Eleanor G. Frierson, Gary K. McCone (Alt)

National Archives and Records Administration
Nancy Allard

National Library of Medicine
Barbara Rapp

National Security Agency
Kathleen Dolan

NFAIS
Marjorie Hlava

OCLC, Inc.
Thomas Hickey

Openly Informatics, Inc.
Eric Hellman

OpenRFP, Inc.
Ralph Shoffner, Donald Chvatal (Alt)

Paratext
Robert Asleson, Eric Calaluca (Alt)

Polaris Library Systems, Inc.
Candy Zemon, Paul Huf (Alt)

ProQuest Information and Learning
Thomas Hamilton, Carol Brent (Alt)

Random House, Inc.
Laurie Stark

Recording Industry Association of America
Bruce Block, Carlos Garza (Alt)

RLG
Lennie Stovel

Sage Publications
Carol Richman, Richard Fidczuk (Alt)

Serials Solutions, Inc.
Mike McCracken

SIRSI Corporation
Greg Hathorn, Slavko Manojlovich (Alt)

Society for Technical Communication (STC)
Frederick O'Hara, Annette D. Reilly (Alt)

Society of American Archivists
Lisa Weber

NISO Board of Directors

At the time NISO approved this standard, the following individuals served on its Board of Directors:

Syntax for the Digital Object Identifier

1   Introduction

1.1   Purpose

This standard defines the syntax for a character string called the Digital Object Identifier (DOI).

1.2   Scope

This standard is limited to defining the syntax of the DOI character string. Policies governing the assignment and use of DOIs are determined by the International DOI Foundation (IDF) and are outside the scope of this document.

2   Standards and References

Referenced standards are those that need to be used to construct a DOI. Secondary standards and references include citations to documents that can be of use in conjunction with the DOI. See Appendix D for related standards and references.

2.1   Referenced Standard

The Unicode Consortium. *The Unicode Standard*, Version 4.0.1, defined by: The Unicode Standard, Version 4.0 (Reading, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1), as amended by Unicode 4.0.1 (http://www.unicode.org/versions/Unicode4.0.1/).

3   Definitions

Term	Definition
Deposit	The act of entering into the Directory a DOI and associated information necessary for the DOI to be used.
Digital Object Identifier (DOI)	A character string used in a System conforming to the rules of, and deposited in the Directory administered by, the IDF.
Directory	A repository in which DOIs are deposited and attendant locations are maintained.
Directory Manager	The organization that manages the Directory on behalf of the IDF.

Term	Definition
DOI prefix	The Directory and the Registrant codes issued by a Registration Agency to a Registrant for use in the DOIs allocated by that Registrant.
DOI suffix	The character string assigned by a Registrant. The suffix shall be unique within the set of DOIs specified by the DOI prefix held by the Registrant.
International DOI Foundation (IDF)	The body set up to support the needs of the intellectual property community in the digital environment by establishing and governing the DOI System, setting policies for the System, appointing service providers for the System, and overseeing the successful operation of the System.
Registrant	An organization or entity that has requested and been allocated one or more DOI prefixes by a Registration Agency.
Registration	The act of allocating the DOI prefix to a Registrant by the Registration Agency
Registration Agency [DOI Registration Agency]	An organization appointed by the International DOI Foundation to register and allocate DOI prefixes to Registrants, and which subsequently accepts DOIs being deposited by Registrants.

4   Format and Characteristics of the DOI

The DOI is composed of the prefix and the suffix. Within the prefix are the Directory Code <DIR> and the Registrant Code <REG>. The suffix is made up of the DOI Suffix String <DSS>.

The syntax of the DOI string is: <DIR>.<REG>/<DSS>

There is no practical limit on the length of a DOI string, or any of its components (the Handle system allows strings of up to 4 GB; under UTF-8 encoding each ASCII character takes one byte, hence in ASCII encoding a DOI may be approx 4 billion characters).

Characters 'a' - 'z' and 'A' - 'Z' in the DOI string are case insensitive (e.g. 10.123/ABC is identical to 10.123/AbC). These characters in the DOI string are converted to upper case upon registration and resolution. If a DOI were registered as 10.123/ABC, then 10.123/abc would resolve it and a later attempt to register 10.123/AbC would be rejected with an error message stating that the DOI was already in existence. Comparison of two DOIs (to decide if they match or not) should be done by first converting all characters 'a' - 'z' in DOI strings to upper case, followed by octet-by-octet comparison of the entire DOI string..

4.1   DOI Character Set

Legal characters are the legal graphic characters of Unicode. This specifically excludes the control character ranges 0x00-0x1F and 0x80-0x9F, which are therefore not valid characters for DOI strings, and will never be present in DOI conformant systems. Reserved characters, if any, are listed in the following descriptions of the prefix and suffix.

4.2   Prefix

<DIR> Directory Code (required)

See Appendix A for all valid values for the Directory Code. The Maintenance Agency is responsible for updating the list of valid values. The Directory Code is numeric; currently the only valid value is <DIR>=10.

<REG> Registrant's Code (required)

Separated from <DIR> by ".". This is assigned to the Registrant by the International DOI Foundation.

DOI Prefix Character Set

Any character within the DOI Character Set as defined above.

<DIR> and <REG> are assigned by the International DOI Foundation.

4.3   Suffix

<DSS> DOI Suffix String (required)

This is assigned by the Registrant.

DOI Suffix Character Set

Any character within the DOI Character Set as defined above, with the exception that the Suffix cannot start with */ where * is any single character. This is reserved for future use. The DSS is case insensitive.

5   Maintenance Agency

The Maintenance Agency designated in Appendix B shall review suggestions for new data elements, interpret the rules prescribed by this standard, and maintain a listing of inquiries and responses that may be used for potential future enhancement of this standard. Questions concerning the implementation of this standard and requests for information should be sent to the Maintenance Agency.

Appendix A:
DOI Specifications

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

This appendix provides information on aspects of the DOI system syntax implementation which are determined by the International DOI Foundation and which will not change the DOI syntax defined in this standard.

Valid values for Directory Code

<DIR> <REG> is assigned by the International DOI Foundation. The prefix is numeric.

Valid value for <DIR> = 10

DOIs are persistent, as defined in IETF RFC 1737. Functional Requirements for Uniform Resource Names. (http://www.ietf.org/rfc/rfc1737.txt): "It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well 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."

UTF-8 encoding is mandated by the Handle System. Therefore, all Unicode characters must be encoded using UTF-8.

The Handle System used as the basis for the DOI system allows an unlimited length for the DOI string. However it is recommended that the suffix (<DSS>) be kept as short as possible to allow for human readability and ease of use in systems where size may be a consideration (e.g., watermarking).

This information is maintained by the DOI Maintenance Agency (see Appendix B).

Appendix B:
Designation of Maintenance Agency

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

The functions assigned to the Maintenance Agency as specified in Section 5 will be administered by The International DOI Foundation (http://www.doi.org/).

Questions concerning the implementation of this standard and requests for information should be sent to:

E-mail: n.paskin@doi.org

Dr Norman Paskin
Director
International DOI Foundation
5, Linkside Avenue
Oxford
OX2 8HY
UK
Tel: (+44) 1865 559070

Appendix C:
Examples of Digital Object Identifiers

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

DOI registrants can use a variety of strings for the DSS including private identifiers and existing standards such as SICI (Serial Item and Contribution Identifier). The syntax of the identifier numbering scheme is such that any existing identifier syntax string can be expressed in a form suitable for use with the DOI system.

The following are examples of Digital Object Identifiers:

DOI for the Authors' Licensing and Collecting Society's Byline service:
10.054/1418EC1N2LE

DOI (incorporating a SICI) from an article in the Journal of the American Society for Information Science, published by John Wiley & Sons:
10.1002/(SICI)1097-4571(199806)49:8<693::AID-ASI4>3.0.CO;2-O

DOI for an article from JAMA, the Journal of the American Medical Association:
10.1001/PUBS.JAMA(278)3,JOC7055-ABST:

DOI for the article "ABO Blood Group System" from Encyclopedia of Immunology Online, 2nd edition, published by Academic Press:
10.1006/rwei.1999.0001

Appendix D:
Related Standards and References

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

The standard cited in Section 2 is required for the construction of the DOI syntax. This appendix includes references to other standards and citations that may be useful with DOIs or which provide additional information on the DOI.

When American National Standards cited below are superseded by a revision, the revision shall apply.

ANSI X3.4:1986 American National Standard for Information Systems – Coded Character Sets – 7-bit American National Standard Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986.

DOI Handbook: DOI 10.1000/182, http://www.doi.org/hb.html

DOI factsheets (DOI and Handle; DOI and Numbering Schemes; DOI and Data Dictionaries; DOI and Internet Identifier Specifications; DOI Applications; Value added by the DOI System: http://www.doi.org/factsheets.html

Handle System: http://www.handle.net/

Sun, Sam; Lannom, Larry; Boesch, Brian. "Handle System Overview". RFC 3650, November 2003. http://www.handle.net/rfc/rfc3650.html

Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle System Namespace and Service Definition". RFC 3651, November 2003. http://www.handle.net/rfc/rfc3651html

Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". RFC 3652, November 2003. http://www.handle.net/rfc/rfc3652html

"UTF-8, A Transform Format for Unicode and ISO10646", RFC2044, October 1996, Yergeau, Francois - http://www.normos.org/ietf/rfc/rfc2044.txt

Appendix E:
Application Issues

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

Except for the specific requirements imposed by this standard (such as use of Unicode and reserved characters), no restrictions are imposed or assumptions made about the characters used in DOIs. Appendix E discusses some encoding issues that arise when using DOIs in specific application contexts like URLs and with the HTTP protocol. Other application contexts in which DOIs are used may have similar types of requirements or restrictions. However, such requirements for encoding or restrictions on the use of particular characters only apply when DOIs are used within those particular application contexts. They are not part of the DOI syntax itself as defined by this document.

UTF-8 Encoding

The Handle System specifies UTF-8 as the encoding for DOI strings. ASCII characters are preserved under UTF-8 encoding. No changes need to be made to ASCII characters to comply with UTF-8 encoding. The default encoding of Unicode is that each character consists of 16 bits (2 octets). UTF-8 is a variation of the Unicode encoding that allows characters to be encoded in terms of one to six octets. UTF-8 encoding plays a role when non-ASCII characters are used. For example, the Japanese word "nihongo" is written as:

日 本 語

The Unicode sequence representing the Han characters for "nihongo" is: 65E5 672C 8A9E

These may be encoded in UTF-8 as follows: E6 97 A5 E6 9C AC E8 AA 9E

For further information on UTF-8 see "UTF-8, A Transform Format for Unicode and ISO10646", RFC2044, October 1996.

Encoding Recommendations When Used in URLs

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.

There is a freely available "resolver plug in" that can be downloaded from http://www.handle.net/resolver/. For both Netscape and Microsoft IE browsers, 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.

DOIs are also primarily used in HTML pages. The DOI 10.1006/rwei.1999".0001as a link in an HTML page would be: <A HREF="http://dx.doi.org/10.1006/rwei.1999%22.0001">10.1006/

rwei.1999%22.0001</A>

Note that " has been encoded (see next section) to distinguish the DOI in the URL from the surrounding text. The DOI is displayed in its encoded form since users may type the DOI directly into their browsers.

Encoding Issues

There are special encoding requirements when a DOI is used with HTML, URLs, and HTTP. The syntax for Uniform Resource Identifiers (URIs) is much more restrictive than the syntax for the DOI. A URI can be a Uniform Resource Locator (URL) or a Uniform Resource Name (URN).

Hexadecimal encoding must be used for characters in a DOI that are not allowed, or have other meanings, in URLs or URNs. Hex encoding consists of substituting for the given character its hexadecimal value preceded by percent. Thus, # becomes %23 and http://dx.doi.org/10.1000/456#789 is encoded as http://dx.doi.org/10.1000/456%23789. The browser does not now encounter the bare #, which it would normally treat as the end of the URL and the start of a fragment, and so sends the entire string off to the DOI network of servers for resolution, instead of stopping at the #. Note: The DOI itself does not change with encoding, merely its representation in a URL. A DOI that has been encoded is decoded before being sent to the DOI Registry. At the moment the decoding is handled by the proxy server http://dx.doi.org/. Only unencoded DOIs are stored in the DOI Registry database. For example, the number above is in the DOI Registry as "10.1000/456#789" and not"10.1000/456%23789". The percent character (%) must always be hex encoded (%25) in any URLs.

There are few character restrictions for DOI number strings per se. When DOIs are embedded in URLs, they must follow the URL syntax conventions. The same DOI need not follow those conventions in other contexts.

Mandatory and Recommended Encoding for DOI Deposit and URLs

Tables 1 and 2 summarize the encoding guidelines for DOI. URLs have the most restricted set of characters. Table 1 lists the characters that should always be hex encoded. Table 2 lists additional characters where it is recommended that characters be replaced by hex-encoding. The distinction between the lists is between practical experience with current Web browsers and the more formal specification of URL syntax. In the DOI Directory all characters represent themselves.

Table 1: Mandatory Encoding

Character	Encoding
%	(%25)
"	(%22)
#	(%23)
SPACE	(%20)

Table 2: Recommended Encoding

Character	Encoding
<	(%3c)
>	(%3e)
{	(%7b)