

## Digital Object Identifier

The Digital Object Identifier (DOI) is an Internet-based global naming and resolution system that provides for the precise identification, retrieval, and trading of digital items in the form of articles, books, images, bibliographies, supporting data, videos, charts, tables, audio, and other electronic files. Development of the DOI system began in 1996 when content creators and technologists jointly recognized that information and entertainment objects could not be commercially distributed on the Internet unless there was a common system of unique identification for those objects. These early stakeholders envisioned an unambiguous, machine-readable identifier that could be used for all electronic communications and transactions involving content throughout its life cycle, including its creation, editing, publication, distribution, and archiving. Such an identifier would be especially critical for commercial transactions, from initial licensing through sales tracking, royalty computation, and financial reporting.

The dominant practice used today by publishers for naming Internet-distributed objects involves referring to their location, using an addressing system known as the uniform resource identifier (URL). URL-only naming fails whenever the resources are moved or reorganized. By contrast, the DOI system introduces a level of indirection that ensures persistent or permanent links to objects by way of a global directory. This level of indirection enables object administrators to update a single, centralized database record for each object, such that requests for the object are automatically and reliably redirected to the object itself, wherever it might be located on the network. The DOI also has a multiple-resolution feature that enables a single DOI to be resolved to locations for related services, transactions, or other information that the object's owner wishes to make available.

**Unique, persistent identification.** In order to discover, retrieve, manage, and trade the vast array of creative works that are becoming available in the digital domain, a way to refer to them unambiguously, by means of unique identifiers, is required. While uniqueness ensures that the identifier will refer to only one object, persistence ensures that if that object is moved or if ownership of the object changes, the identification of that object does not need to change.

Uniqueness and persistence are facilitated if an identifier is designed as an opaque string or dumb number, meaning that no meaning should be inferred from the assigned value of the name or number. This characteristic distinguishes the DOI from many identifier systems that build “intelligence” into the number itself, a feature that might enable the user to deduce something about the entity that is being identified, or perhaps about the person or organization that registered the identifier. In the case of a dumb number like the DOI, the only reliable way to know anything about an identified object is to view

the metadata (description of the object) declared at the time of registration. This ensures that even when the ownership of a particular item changes, its identifier remains the same, with the same descriptive information.

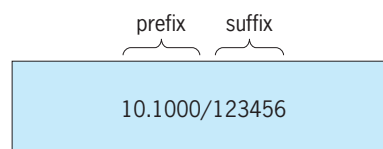
**Uses.** The DOI is a persistent identifier of intellectual property entities, where an “entity” is any object that can be usefully identified, including resources made available on the World Wide Web (WWW). In general, “intellectual property” is defined by the World Intellectual Property Organization (WIPO) and related international treaties such as the Berne Convention.

Typically, the DOI may be used by publishers to identify the various physical objects that are manifestations of intellectual property, including printed books, CD recordings, videotapes, and journal articles. A DOI may also be used to identify less tangible manifestations, especially the digital files that are the common form of intellectual property expression on the Internet. But the power of the DOI goes beyond its ability to identify manifestations—it may also be used to identify performances of intellectual property or the abstractions, including works such as musical scores or lyrics, that are the basis for those manifestations.

Finally, the DOI can be used to route users and applications to multiple services related to the named object, such as bibliographic or transaction services.

**Prefix and suffix.** A DOI is made up of two components (see *illus.*). The first element, the prefix, is assigned to an appropriate grouping of content such as the publisher, label, or imprint by a registration agency. All prefixes begin with 10, followed by a number designating the organization, publisher, or any rights holder or controller who has obtained that prefix and is responsible for depositing the individual DOIs. Organizations such as publishers might choose to request a prefix for each of their imprints or product lines, or they might use a single prefix.

The second element of a DOI is the suffix, a unique character string assigned by the prefix holder to the specific object being identified. Existing identifier or product numbering systems used within a particular industry, as well as private naming systems used within the corporation, are typically incorporated within the suffix of the DOI. The suffix may be assigned to entities of any size or granularity (such as a book, article, abstract, chart, album, song, or melody) or any file type (such as text, audio, video, image, or software). The prefix holder decides the level or granularity of identification based on the nature of the objects to be distributed or sold electronically.



Digital Object Identifier components.

*Resolution.* The power of the DOI system is its function as a global routing or resolution system. Since digital content frequently changes ownership and location over the course of its useful life, the DOI system uses a distributed global directory. Whenever a request is made to resolve a DOI (for example, when a user clicks on a DOI link in a Web page), a message is sent to the central directory on a server where the current address associated with that DOI appears. This location is sent back to the user, enabling redirection within the user's browser to this particular Internet address. The user would simply see the content itself, or further information about the object with information on how to obtain it.

When the object moves to a new server or the copyright holder transfers the product line to another company, only one change needs to be recorded in the directory; from this time onward, all users will instantly be sent to the new location. The assigned DOI remains reliable and accurate because the link to the associated information or source of the content is easily and efficiently changed. The underlying technology of the secured global name service, called the Handle System<sup>®</sup>, has been optimized for speed, efficiency, and persistence.

*Metadata.* Since there is no intelligence built into the DOI itself, the ability to retrieve descriptive metadata about DOI-named objects and related services is an essential component of the DOI system. Information about an identified object must be created and maintained by the holder of the DOI prefix. The DOI system mandates a minimum level of publicly available structured metadata, known as kernel metadata, which serves a role similar to a telephone directory entry; the data enable users to ensure that the identified entity found is the one sought. Kernel metadata elements include:

*Identifier:* an identifier associated with the entity from a legacy identification scheme, such as an ISBN.

*Title:* a name by which the entity is known.

*Type:* the primary type of intellectual property entity that is being identified (an abstract work, a tangible or intangible manifestation, a performance).

*Mode:* the sensory mode through which the intellectual property entity is intended to be perceived (visual, audio, and audiovisual).

*Primary agent:* the identity of the primary agent, normally the first-named creator of the object.

*Agent role:* the role that the primary agent played in creating the object.

DOI application developers and their user communities may also define specialized metadata structures that further describe the object and enable its use in proprietary ways, such as transaction systems. These structures are known as application profiles, and their data may or may not be publicly available.

*Technology.* The underlying technology for resolving DOIs on the Internet, the Handle System<sup>®</sup>, is a distributed, scalable system based on open protocols developed by the Center for National Research Initiatives (CNRI) in the United States. The DOI's metadata system is based on the <indecs> metadata framework, which has also been adopted as the basis of the data dictionary of the MPEG-21 Multimedia Framework. Since both the Handle System<sup>®</sup> resolution and the DOI metadata components are structured, consistent, and manageable, it is possible to apply DOIs to any form of content and to develop additional tools for content management. Ultimately, the system allows the development of automated agents that can use DOIs to manage entities throughout the transaction life cycle of the object.

**Reference linking.** Reference linking is a mechanism that easily and accurately takes readers of electronic journals from one document into another. References are the traditional means for authors to establish links between their work and scholarship that preceded it. Reference linking based upon the DOI is a means for making those links immediately actionable across publications and publishers.

In January 2000, a collaborative venture, called CrossRef, was incorporated among 12 of the world's largest scientific and scholarly publishers, both commercial and not-for-profit, to enable cross-publisher reference linking throughout the digital journal literature. With over 150 members, 6700 journals, and 5.1 million articles linked, CrossRef is the most significant and successful application of the DOI.

A researcher clicking on a CrossRef link within a journal article will be connected to a page on the publisher's Website showing a full bibliographical citation of the article, including in most cases the article abstract. The reader then has the option to immediately access the target article; subscribers to the target can typically go straight to the full text, while nonsubscribed users are presented with options for access.

**Governance.** Since 1998, the controlling authority for the DOI has been the not-for-profit International DOI Foundation (IDF), based in Washington, DC, and Geneva. Members of the IDF may include larger companies from the publishing, music, still-image, broadcast, online news, software, and other content industries; technology providers to these industries; associations representing these industries; Internet technology companies; associations representing stakeholders such as authors, artists, libraries, secondary publishers, and users; government agencies with a mission in a related area; others for whom Internet commerce is a critical business, social, cultural, political, or personal interest.

**Future developments.** The DOI is currently undergoing a period of rapid development, following the early acceptance of the principles of the system and its adoption in initial applications. As of July 2002, several million DOIs had been issued, with over 200 organizations allocating DOIs; five DOI Registration Agencies had been appointed (with more planned

to come); and the DOI had been well integrated into several related standards activities, with many applications actively under development. That development is evident in all aspects of the DOI System—technology, procedure, and policy.

For background information *see* DATABASE MANAGEMENT SYSTEM; INFORMATION MANAGEMENT; INTERNET; MULTIMEDIA TECHNOLOGY; WORLD WIDE

WEB in the McGraw-Hill Encyclopedia of Science & Technology. John S. Erickson

Bibliography. N. Paskin, Toward unique identifiers, *Proc. IEEE*, 87(7):1208-1227, 1999; *Syntax for the Digital Object Identifier*, Z39.84-2000 Document Number: ANSI/NISO Z39.84-2000, National Information Standards Organization Staff and National Information Standards Organization, 2000.

Reprinted from the McGraw-Hill Yearbook of Science & Technology 2003. Copyright © 2003 by The McGraw-Hill Companies, Inc. All rights reserved.