Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:48 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

Digital Object Identifier (DOI®) System

Norman Paskin
Tertius Ltd., Oxford, U.K.

Abstract
The Digital Object Identifier (DOI®) System is a managed system for persistent identification of content on digital networks. It can be used to identify physical, digital, or abstract entities. The identifiers (DOI names) resolve to data specified by the registrant, and use an extensible metadata model to associate descriptive and other elements of data with the DOI name. The DOI system is implemented through a federation of registration agencies, under policies and common infrastructure provided by the International DOI Foundation which developed and controls the system. The DOI system has been developed and implemented in a range of publishing applications since 2000; by early 2009 over 40 million DOIs had been assigned. The DOI system provides identifiers which are persistent, unique, resolvable, and interoperable and so useful for management of content on digital networks in automated and controlled ways.

IDENTIFIER CONCEPTS

An identifier is a concise means of referencing something. The term "identifier" can mean several different things:

- A "string," typically a number or name, denoting a specific entity (the referent of the identifier string). For example, the identifier ISBN 978-0-00-721331-3 denotes the book "Francis Crick" by Matt Ridley.
- A "specification," which prescribes how such strings are constructed. For example, the ISO standard ISO 2108:2005[1] is the current specification of the ISBN numbering system; but having that standard alone will not enable someone to construct and register a new valid ISBN.
- A "scheme," which implements such a specification. For example, the ISBN International Agency[2] implements the ISBN standard in an implemented scheme, by assigning ISBN prefixes to publishers, registering specific ISBNs (strings); providing rules on use of the ISBN (such as the incorporation of the ISBN as a bar code on the cover of a book). Typically, such schemes provide a managed registry of the identifiers within their control, in order to offer a related service.

Some important concepts relating to identifiers are "uniqueness," "resolution," "interoperability," and "persistence."

Uniqueness is the requirement that one string denotes one and only one entity (the "referent"). Note that the converse is not a logical consequence: it is not necessary that an entity have only one identifier. For example, a book may have an ISBN and also an LCCN. An identifier scheme may even allow multiple identifiers for one entity, though usually these are deprecated.

Resolution is the process in which an identifier is the input to a service to receive in return a specific output of one or more pieces of current information related to the identified entity. For example, a bar code ISBN in a bookshop is scanned by a bar code reader and resolves to some point of sale information, such as title and price. Note that resolution depends on a particular application: while a bar code in a bookshop may resolve to price, the same bar code in a warehouse application might resolve to current stock number, or pallet position. Another familiar example of resolution is the Internet Domain Name System (DNS) which resolves a domain name address (URL) to a file residing on a specific host server machine.

Interoperability denotes the ability to use an identifier in services outside the direct control of the issuing assigner: identifiers assigned in one context may be encountered in another place or time without consulting the assigner. This requires that the assumptions made on assignment will be made known in some way. For example, a customer may order a book from a bookseller or a library system by quoting its ISBN, without consulting the publisher who assigned the number.

Persistence is the requirement that once assigned, an identifier denotes the same referent indefinitely. For example, ISBNs, once assigned, are managed so as to reference the same book always (and are not reassigned). Persistence can be considered to be "interoperability with the future."

The management of content on digital networks requires identifiers to be persistent, unique, resolvable, and interoperable. As an example, URLs do not identify content but a file location: using them as a substitute for such identifiers

Encyclopedia of Library and Information Sciences, Third Edition DOI: 10.1081/E-ELIS3-120044418
Copyright © 2010 by Taylor & Francis. All rights reserved.

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:50 Filepath://spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

is not sustainable for reliable automation. The content may be removed ("404 not found"), or changed (not being the same as the user anticipated, or the user being unaware of such change). There have been a number of efforts to address the need for such reliable identifiers, notable among them URN[3] and URI[4] specifications; however these do not of themselves provide an implemented managed scheme and registry for specific content sector applications. Such full schemes require more: a model for identifiers and their management; shared, standards-based, persistent identifier management infrastructure; support for adoption of persistent identifiers and services, and a plan for sustainable shared identifier infrastructure.[5,6] The Digital Object Identifier (DOI®) system is such a managed system for persistent identification of content on digital networks, using a federation of registries following a common specification.

The uncapitalized term "digital object identifier" may be used nonspecifically to describe a number of varied technologies concerned with the identification of entities in a digital environment. The capitalized term "Digital Object Identifier" refers to one specific system defined and managed by the International DOI Foundation,[7] which provides an infrastructure for persistent unique identification of entities (here termed "objects") on digital networks deployed in a number of content-related applications.

DOI SYSTEM: OUTLINE

DOI is an acronym for Digital Object Identifier. The DOI system provides for unique identification, persistence, resolution, metadata, and semantic interoperability of content entities ("objects"). Information about an object can change over time, including where to find it, but its DOI name will not change.

The DOI system brings together

- A syntax specification, defining the construction of a string (a DOI name)
- A resolution component, providing the mechanism to resolve the DOI name to data specified by the registrant
- A metadata component, defining an extensible model for associating descriptive and other elements of data with the DOI name
- A social infrastructure, defining the full implementation through policies and shared technical infrastructure in a federation of registration agencies

More detail on each of these aspects is given later in this entry.

The DOI system operates through a tiered structure:

- The International DOI Foundation is the umbrella organization defining the rules and operation of the system. It is a non-profit member-funded organization.
- Registration agencies are all members of the International DOI Foundation, and have a contractual arrangement with the Foundation including a license to operate the DOI system. They provide defined services in specific sectors or applications. DOI registration is normally only a part of the service such an organization offers, since assignment of an identifier is usually done for the purpose of a specific initial service or application. An example is the CrossRef registration agency,[8] which provides services to publishers for linking reference citations in articles based on DOI-identified articles. Registration agencies may collaborate, or remain relatively autonomous.
- DOI names are registered by clients via a registration agency (e.g., in the case of the CrossRef agency, individual publishers are clients using the CrossRef service). Part of this process may be undertaken by the registration agency, as part of its service offering. If a suitable registration agency cannot be found for a certain sector, the International DOI Foundation will seek to appoint one.

DOI is a registered trademark of the International DOI Foundation, Inc. (abbreviated to IDF). The preferred usage, to avoid ambiguity, is with a qualifier to refer to either specific components of the DOI system (e.g., "DOI name": the string that specifies a unique referent within the DOI system); or the system as a whole ("DOI system": the functional deployment of DOI names as the application of identifiers in computer-sensible form through assignment, resolution, referent description, administration, etc.).

SCOPE

The term "Digital Object Identifier" is construed as "digital identifier of an object," rather than "identifier of a digital object": the objects identified by DOI names may be of any form—digital, physical, or abstract—as all these forms may be necessary parts of a content management system. The DOI system is an abstract framework which does not specify a particular context of its application, but is designed with the aim of working over the Internet.[9]

A DOI name is permanently assigned to an object, to provide a persistent link to current information about that object, including where it, or information about it, can be found. The principal focus of assignment is to content-related entities; that term is not precisely defined but is exemplified by text documents; data sets; sound carriers; books; photographs; serials; audio, video, and audiovisual recordings; software; abstract works; artwork, etc., and related entities in their management, for example, licenses or parties. A DOI name is not intended as a replacement for other identifier schemes, such as those of ISO TC46/SC9[10] ISBN, ISSN, ISAN, ISRC, etc., or

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:51 Filepath://spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

other commonly recognized identifiers: if an object is already identified with another identifier string, the character string of the other identifier may be integrated into the DOI name syntax, and/or carried in DOI metadata, for use in DOI applications.

A DOI name may be assigned to any object whenever there is a functional need to distinguish it as a separate entity. Registration agencies may specify more constrained rules for the assignment of DOI names to objects for DOI-related services (e.g., a given registration agency may restrict its activities to one type of content or one type of service).

SYNTAX

A DOI name is the string that specifies a unique object (the referent) within the DOI system. The DOI syntax (standardized as ANSI/NISO Z39.84-2005)[11] prescribes the form and sequence of characters comprising any DOI name. The DOI syntax is made up of a "prefix" element and a "suffix" element separated by a forward slash. There is no defined limit on the length of the DOI name, or of its prefix or its suffix elements. The DOI name is case-insensitive and may incorporate any printable characters from the Unicode Standard.

- Example: a DOI name with the prefix element "10.1000" and the suffix element "123456": 10.1000/123456

The combination of a unique prefix element (assigned to a particular DOI registrant) and a unique suffix element (provided by that registrant) is unique, and so allows the decentralized allocation of DOI numbers. The DOI name is an opaque string for the purposes of the DOI system: no definitive information should be inferred from the specific character string of a DOI name. In particular, the inclusion in a DOI name of any registrant code allocated to a specific organization does not provide evidence of the ownership of rights or current management responsibility of any intellectual property in the referent. Such information can be asserted in the associated DOI metadata.

The DOI prefix has two components: a "Directory" indicator followed by a "Registrant" code, separated by a full stop (period) (e.g., 10.1000). The directory indicator is always "10" and distinguishes the entire set of character strings (prefix and suffix) as DOIs within the wider Handle System® used for resolution. The registrant code is a unique alphanumeric string assigned to an organization that wishes to register DOI names (four digit numeric codes are currently used though this is not a compulsory syntax). The registrant code is assigned through a DOI registration agency, and a registrant may have multiple-registrant codes. Once a DOI name is assigned the string should not be changed, regardless of any changes in the ownership or management of the referent object; if an object is withdrawn from digital access, its DOI name should still resolve to some appropriate message to this effect.

The DOI suffix may be a sequential number, or it may incorporate an identifier generated from or based on another system used by the registrant (e.g., ISBN, ISSN, ISTC). In such cases, the existing system may specify its own preferred construction for such a suffix:

- Example: a DOI suffix using an ISSN: 10.1038/issn.0028-0836.

When displayed on screen or in print, a DOI name is normally preceded by a lowercase "doi": unless the context clearly indicates that a DOI name is implied.

- Example: the DOI name 10.1006/jmbi.1998.2354 is displayed as doi:10.1006/jmbi.1998.2354.

The use of lowercase string "doi" follows the specification for representation as a URI (as for e.g., "ftp:" and "http:").

DOI names may be represented in other forms in certain contexts. For example, when displayed in Web browsers the DOI name itself may be attached to the address for an appropriate proxy server (e.g., http://dx.doi.org/ resolves DOIs in the context of Web browsers using the Handle System resolution technology) to enable resolution of the DOI name via a standard Web hyperlink.

- Example: the DOI name 10.1006/jmbi.1998.2354 would be made an active link as http://dx.doi.org/10.1006/jmbi.1998.2354.

DOI names so represented in a URL and transported by the HTTP protocol are constrained to follow standard IETF guidelines for URI representations. The syntax for URIs is more restrictive than the syntax for DOIs; some characters are reserved and will need encoding (the NISO Z39.84 DOI syntax standard provides more detail). Certain client or server software may be able to handle DOIs using native handle resolution technology (where doi:10.1006/jmbi.1998.2354 would be understood by the browser and automatically resolved without the addition of the proxy server address). DOI names may also be represented in other schemes, for example, in the info URI schema[12,13] as info:doi/10.1006/jmbi.1998.2354.

RESOLUTION

A DOI name can, within the DOI system, be resolved to values of one or more types of data relating to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers, and descriptive metadata (or any additional types defined extensibly by the registration

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:52 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

agency). Resolution is the process of submitting a specific DOI name to the DOI system (e.g., by clicking on a DOI in a Web browser) and receiving in return the associated values held in the DOI resolution record for one or more of those types of data relating to the object identified by that DOI name. Since the referent objects referred to by DOI names may be of various types (including abstractions as "works," physical "manifestations," performances), they may or may not be directly accessible in the form of a digital file or other manifestation; hence the resolution may or may not return an instance of the object.

The initial implementation of DOI system was that of persistent naming: a single redirection from a DOI name to a digital location (URL) of the entity (Fig. 1).

A significant DOI function is the capability for multiple resolution, that is, delivering more than one value back from a resolution request. The values are grouped into defined types, which can form the basis of services (Fig. 2). An example of current usage of this facility is resolution to a specific local copy of an article, determined by combining the resolution result (several URLs) and local information about the user's location (from the user's browser application).

Objects (identified by DOI names) which have common behavior (defined by metadata) can be grouped, using DOI application profiles; these application profiles can in turn be associated with one or more services applicable to that group of DOI names (see Fig. 3).

The Handle System,[14] the resolution component used in the DOI system, is a general-purpose distributed information system designed to provide an efficient, extensible, and secure 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 DOI system is one implementation of the Handle System; hence a DOI name is a "Handle." DOI names are distinguished from other handles by additional "metadata" and "policy." The Handle System enables entities to be assigned first-class names, independent of domain names and other location-specific information, which can then be resolved (redirected) to appropriate locations: since the resolution destination is managed and can be changed, this provides a tool for persistence, avoiding "404 not found" and similar problems with URLs. The Handle System is used in a variety of applications such as the Content Object Repository Discovery and Resolution Architecture (CORDRA) of the U.S. Department of Defense (DoD) Advanced Distributed Learning initiative; The Library of Congress National Digital Library Program; and applications in grid computing and advanced future Internet architectures. The Handle System also includes several features not currently used in the DOI system, such as trusted resolution using public key infrastructure.

The Handle System is part of a wider Digital Object Architecture[15]; that architecture specifically deals only with digital objects with identifiers (Handles). There is no conflict in these two views, since any non-digital entity may be reified (or represented) as a corresponding digital object for the purposes of digital object management

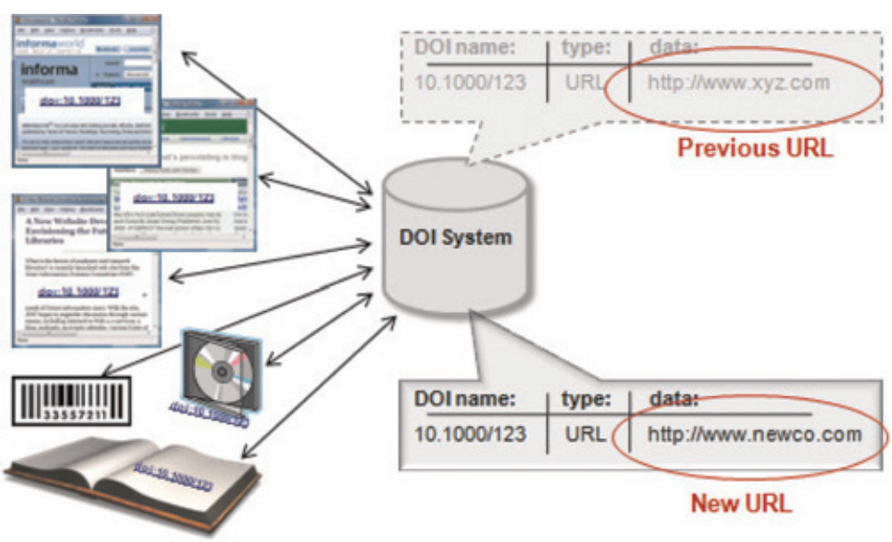

Fig. 1 The role of the DOI system as a persistent identifier. A DOI name (10.1000/123) has been assigned to a content entity; the DOI system provides resolution from that name to a current URL. When the content, previously at URL xyz.com, is moved to a new URL newco.com, a single change in the DOI directory is made: all instances of the DOI name identifying that content (even if already recorded in print, as bookmarks, etc.) will resolve to the new URL, without the user having to take any action or be aware of the change. Note that the DOI name is persistent, i.e., remains unchanged.
Source: From International DOI Foundation.

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:53 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

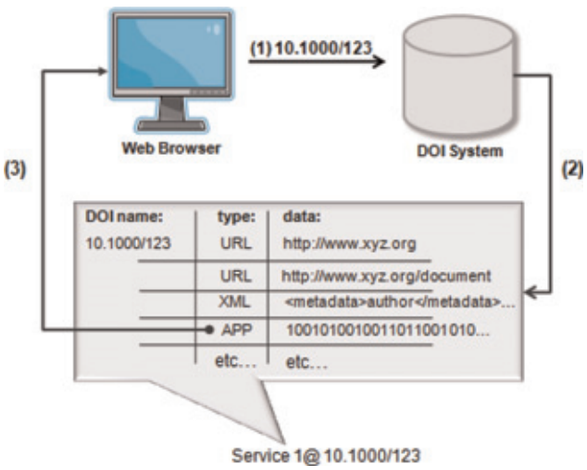

Fig. 2 Multiple resolution. A Web browser is running an application "Service 1." That service resolves DOI name 10.1000/123 to the DOI system (1) where it finds four values within the relevant DOI record (2): here, two are of the type URL, one is XML, and one is a conjectural application. Service 1 selects one of these results (in this case, the APP value) on the basis of combining information provided in the resolution result and the local application.
Source: From International DOI Foundation.

(though some care is needed in the definition of such objects and how they relate to non-digital entities).

METADATA

The object associated with a DOI name is described unambiguously by DOI metadata, based on an extensible data model to support interoperability between DOI applications. Assignment of a DOI name requires the registrant to record metadata describing the object to which the DOI name is being assigned. The metadata describes the object to the degree that is necessary to distinguish it as a separate entity within the DOI system.

A minimum set of such metadata, the DOI kernel, is specified by the IDF. This includes elements such as "other identifier(s) commonly referencing the same referent (e.g., ISBN, ISRC)," and the name by which the referent is usually known (e.g., title). This minimum kernel may be enhanced by registration agencies through the development of specific application profiles with metadata elements appropriate to a particular application or set of applications. The IDF also specifies the template for the exchange of metadata between DOI registration agencies to support their service requirements, and specifies a Data Dictionary as the repository for all data elements and allowed values used in DOI metadata specifications.

The basis of the metadata scheme and extensions used in the DOI system is the indecs (interoperability of data in

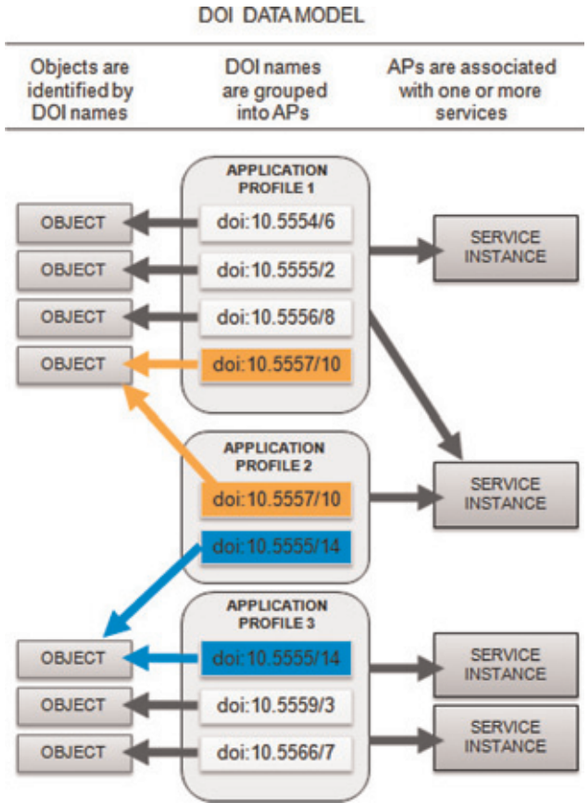

Fig. 3 DOI data model: the Application Profile Framework. DOI names (identifying the entities on the left) are grouped into application profiles. Any single DOI name can be a member of multiple application profiles (e.g., DOI 10.5557/10 is shown here in two). Each application profile can similarly be associated with one or more services: each service can be made available in multiple ways. This makes it possible to make a new service applicable to many DOI names, simply by adding that service to the relevant application profile(s).
Source: From International DOI Foundation.

e-commerce systems) project.[16] This contextual ontology approach to interoperability is shared by a number of significant content sector activities.[17,18] This allows the use of a variety of existing metadata schemes with DOI names in a common framework.

The use of these tools for DOI metadata has been limited in initial applications, but more applications are emerging as the sophistication of content management on digital networks and the need for interoperability increases.

SOCIAL INFRASTRUCTURE

DOI names are intended to be persistent identifiers: no time limit for the existence of a DOI name is assumed in any assignment, service, or DOI application. A DOI name and its referent are unaffected by changes in the rights associated with the referent, or changes in the

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:55 Filepath:// spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

management responsibility of the referent object. Since such persistence requires a social infrastructure, policies as well as technical infrastructure need to be defined and implemented. The IDF develops and implements policies such as rules for transfer of management responsibility between registration agencies, requirements placed on registration agencies for maintenance of records, default resolution services, and technical infrastructure resilience. These are codified in a formal agreement between the IDF and each of the registration agencies.

The DOI system is not a means of archival preservation of identified entities; it does not store the identified objects themselves; nor does the central DOI Directory store comprehensive metadata (only pointers to the registration agency or other source of such data). The system provides a means to continue interoperability through exchange of meaningful information about identified entities through at minimum persistence of the DOI name and a description of the referent.

HISTORY

The DOI system was the result of a publishing industry initiative in the late 1990s, which recognized the need to uniquely and unambiguously identify content entities, rather than refer to them by locations, and commissioned a study to specify an appropriate technical solution, selected if possible from one or more existing technologies rather than developing a new system. The International DOI Foundation was incorporated in 1998 to develop the system; where possible, existing technologies and standards were adopted for the implementation of the DOI system. The first DOI registration agency began in 2000; by early 2009 around 40 million DOI names had been assigned through eight registration agencies. The most widely known application of the DOI system is the Crossref cross-publisher citation linking service which allows a researcher link from a reference citation directly to the cited content on another publisher's platform, subject to the target publisher's access control practices. Other applications include government documentation, books, and data; further applications are under development.

The development of the DOI system has proceeded through three parallel tracks:

- An initial implementation of persistent naming: a single redirection from a DOI name to a digital location (URL) of the entity or information about it.
- The development of more sophisicated means of management such as contextual resolution, where the result of a redirection is also a function of some additional information such as local holdings information.
- Collaboration with other standards activities in the further development of tools for managing entities in a digital environment.

The DOI System is a Draft International Standard of ISO, it is expected that the final standard will be published in late 2009 or 2010.

RELATED ACTIVITIES

The DOI system is associated with two independent technical activities which it has used as components of DOI implementations: the Handle System and the series of contextual ontology initiatives derived from the indecs project. Each is used in other non-DOI applications (an aim of the International DOI Foundation was to use existing solutions where available and proven to be useful). Either of these components could be replaced in the DOI system by other technologies offering similar features in the future if necessary.

The International DOI Foundation, particularly through its registration agency CrossRef, has also been closely involved in the development of the OpenURL, a mechanism for transporting metadata and identifiers describing a publication for the purpose of context-sensitive linking. The DOI system is now widely implemented using OpenURL by many libraries: further information on this topic is available from the Crossref Web site. The use of open URL was the first widespread example of more sophisticated means of content management through contextual resolution.

The expertise of the International DOI Foundation in issues such as resolution and semantic interoperability has also led to some IDF members being active participants in discussions of further identifier scheme development such as the International Standard Text Code (ISTC) numbering system for the identification of textual works, and identifiers for parties (persons and organizations), and licenses.

REFERENCES

1. ISO 2108:2005 Information and documentation—International standard book number (ISBN), http://www.iso.org/iso/iso catalogue/catalogue tc/catalogue detail.htm?csnumber=36563 (accessed April 2009).
2. The International ISBN Agency web site: International Standard Book Number System for Books, Software, Mixed Media etc. in Publishing, Distribution and Libraries. http://www.isbn-international.org/ (accessed April 2009).
3. Sollins, K.; Masinter, L. 1994. Functional Requirements for Uniform Resource Names. Internet Engineering Task Force (IETF) Request for Comments (RFC) 1737, December 1994. http://tools.ietf.org/html/rfc1737 (accessed April 2009).
4. Berners-Lee, T.; Fielding, R.; Masinter, L. Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force (IETF) Request for Comments (RFC) 3986, January 2005. http://www.ietf.org/rfc/rfc3986.txt (accessed April 2009).

Comp. by: WOMAT Stage : Revises2 ChapterID: 0000892800 Date:27/6/09 Time:07:01:57 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000892800.3D

5. Dyson, E. Online Registries: The DNS and Beyond. Release 1.0, Volume 21, Number 8, 16 September 2003. http://doi.contentdirections.com/reprints/dyson_excerpt.pdf (accessed April 2009).
6. PILIN team: Persistent Identifier Linking Infrastructure Project Report Dec 2007. https://www.pilin.net.au/Closure_Report.pdf (accessed April 2009).
7. The DOI system web site: http://www.doi.org (accessed April 2009).
8. CrossRef web site, http://www.crossref.org (accessed April 2009).
9. Kahn, R.E.; Cerf, V.G. *What is the Internet (And What Makes It Work)*, Internet Policy Institute, December 1999. http://www.cnri.reston.va.us/what_is_internet.html (accessed April 2009).
10. ISO (International Organization for Standardization) TC 46 (Technical Committee for information and documentation standards SC9 (Subcommittee on the identification and description of information resources). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=48836&published on (accessed April 2009).
11. ANSI/NISO Z39.84 (2005): Syntax for the Digital Object Identifier. www.niso.org (accessed April 2009).
12. Van der Sompel, H.; Hammond, T.; Neylon, E.; Weibel, S. The info URI Scheme for Information Assets with Identifiers in Public Namespaces. Internet Engineering Task Force (IETF) Request for Comments (RFC) 4452, April 2006. http://www.ietf.org/rfc/rfc4452.txt.
13. About INFO URIs: Frequently Asked Questions, http://info-uri.info/registry/docs/misc/faq.html (accessed April 2009).
14. The Handle System, http://www.handle.net/ (accessed April 2009).
15. Kahn, R.; Wilensky, R. A framework for distributed digital object services. Int. J. Digital Libr. **April 2006**, *6* (2). [doi:10.1007/s00799–005–0128-x] (First published by the authors in May 1995.) Reproduced at http://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf with permission of the publisher (accessed April 2009).
16. Rust, G.; Bide, M. The <indecs> Metadata Framework: Principles, model and data dictionary. 2000. http://www.doi.org/topics/indecs/indecs_framework_2000.pdf (accessed April 2009).
17. Paskin, N. Identifier interoperability: A report on two recent ISO activities. D-Lib Mag. **April 2006**, *12* (4). http://www.dlib.org/dlib/april06/paskin/04paskin.html (accessed April 2009).
18. Dunsire, G. Distinguishing content from carrier: The RDA/ONIX framework for resource categorization. D-Lib Mag. **January 2007**, *13* (1). http://www.dlib.org/dlib/january07/dunsire/ 01dunsire.html (accessed April 2009).