

Managing Intellectual Property Resources Using the DOI System

Eamonn Neylon

Abstract

The DOI System supports management of distributed intellectual property resources on networks by maintaining state data and providing a resolution service. It is used to enable linking between millions of journal articles within the CrossRef application. This presentation provides an overview of the DOI System and demonstrations of its capabilities.

1. Introduction

New business models, such as peer-to-peer super-distribution, could change the way we access and use content. The future role of current participants in the information supply chain will be affected by the breakdown of distributed islands of information into one primordial sea of content that can be accessed at various levels of granularity. We are witnessing this now in the music industry where the track is fast replacing the packaged item as the unit of information commerce transaction - at this time new paradigms for information management are needed to address the real opportunity to generate new revenue streams and abate copyright abuse. The commercial framework for content distribution will force us to address fundamental issues such as what it is to publish and who can publish.

In a distributed networked environment no single entity has control over the marketplace. Rather than just relying on an ability to protect access to content to enforce copyright respect, there must be the means to allow access to content on an as-required basis. Digital rights management is often promoted on the basis of how it can protect content from copyright infringement. But digital rights management can do more than simply stop certain activities from taking place - it can also be used to allow new uses of existing content. The creative use of this technology can reduce the need for protection and increase the reward to intellectual property holders by

allowing the incremental licensing of content depending on a user's requirement. However for this nirvana to be realized there must be a means for the content to be interoperable.

There are many ways in which interoperability can be achieved. This paper is restricted to content identification and metadata applications. We start with the premise that the publishing industry is built from a collection of disparate databases, many of which contain very similar data but few of which can directly interoperate with each other. Without (metadata) standards, in order to track, find, and sell or purchase electronic content, parties bear the cost of inefficient mechanisms of entering and communicating metadata. Metadata enables us more completely to understand content for commercial and other purposes. Most of the information on the title page of a book, for example, is metadata. Metadata can have many functions e.g., for discovery purposes and some metadata such as rights information may be private and unseen by the unauthorized. It is important that participants in the value chain have a common understanding of metadata to describe content efficiently and effectively.

The key to metadata descriptions is the use of identifiers. The Digital Object Identifier (DOI) is presented as a carefully designed system that addresses key issues in the use and interoperability of identifiers. Taking the Digital Object Identifier as the starting point, this paper shows how applications can be constructed with great flexibility that are bound together through the design principles of the Digital Object Identifier System.

2. Identification

Is the digital object identifier an identifier for digital objects or a digital identifier of objects? The answer is that the DOI is not restricted to the electronic networked environment and so is a digital identifier for objects. Those objects that are identified do not need to exist in a manifestation but can be abstract ideas or physical manifestations outside of an electronic environment. The DOI is not about distributed content management but the distributed management of the services relating to the identified content.

In the approach that is taken the Digital Object Identifier becomes very powerful,

providing the infrastructure to locate all services related to a particular identifier through its resolution system. The decision to define the Digital Object Identifier as a digital identifier for objects is not arbitrary but reflects the design of the Digital Object Architecture from which the DOI was derived [KW].

The DOI is a persistent identifier of intellectual property entities. The DOI can be used to identify any of the various physical objects that are manifestations of intellectual property: for example, printed books, CD recordings, videotapes, journal articles. A DOI can also be used to identify less tangible manifestations, the digital files that are the common form of intellectual property in the network environment. But the use of a DOI can go beyond the identification only of manifestations - it can also be used to identify performances of intellectual property or the abstractions that underlie the different manifestations.

The DOI system consists of three distinct components. The syntax is a flexible structure which accommodates existing identifiers. The system is a means of making the identifier actionable so that predictable access to resources related to the content being identified can be achieved. The administration is the means by which policies are enforced to ensure that the principles upon which the DOI system is based are maintained.

A DOI is made of a globally unique prefix combined with a prefix holders suffix. The DOI syntax [NISO] allows the inclusion as a DOI suffix of any string, which can therefore include support of existing identification systems. The DOI syntax as currently implemented specific Roman alphabet characters; however this may be extended if required to the full character set supported by the underlying technology, which is based on Unicode 2.0 with UTF-8 encoding.

10.1045/6784-3456(12)6-845
{prefix}{<-----suffix----->}

Figure 1. Structure of a Digital Object Identifier

The DOI system is scalable in terms of numbers (there are no fixed field lengths and an infinite number of DOIs may be conceived) and performance. Persistence is a

function of policy, not technology, though technology may assist: DOI policies are designed to enforce persistence and to prevent deletion or renaming of identifiers. Extensibility is the fundamental design goal of the DOI system (being synonymous with flexibility).

The global uniqueness of each DOI is encouraged through syntax and administrative procedures and is absolutely enforced through the technology, which does not allow the entry of duplicate DOIs. The DOI string and the physical location of the Digital Item are completely unrelated. A DOI is not derived in any way from the entity which it names, but is assigned to it independently. While an existing name, or even a mnemonic, may be included in a DOI for convenience - that is, the DOI may incorporate "intelligent" or "significant" identifiers from a particular community - but DOIs are opaque strings intended for the purposes of the resolution and DOI system. The use of opaque strings is a design principle which encourages maximal flexibility and does not hard-wire metadata into an identifier.

The DOI is a persistent identifier: even if ownership of the entity or the rights in the entity change, the identification of that entity should not (and does not) change. The responsibility for managing the DOI changes, but not the DOI itself.

3. Resolution

The Digital Object Identifier has been implemented in a system. The system provides an extensible framework for managing intellectual content services in any form at any level of granularity, for linking customers with content suppliers, for facilitating electronic commerce, and enabling automated copyright management for all types of media.

Making the identifier actionable by providing information linked to the DOI, and the technology to deliver the services that this can provide to users. The DOI system uses the Handle System, which provides a globally distributed capability for assigning, managing, and resolving persistent identifiers, known as 'handles' to facilitate the access of digital objects and other resources on networks such as the Internet over long periods of time. Handle resolution enables an identifier (DOI) to resolve to multiple pieces of current state data such as type(s) and location(s) of instances of the identified entity, type(s) and location(s) of associated metadata,

public keys, accessibility, etc. The Handle System provides the underlying capabilities of state maintenance.

The Handle System is itself a component is a wider digital object infrastructure that makes information objects 'first-class citizens' in a digital networked environment. In this scheme, the digital object is the conceptual elemental unit in the information view; it is interpretable (in principle) by all participating information systems. The digital object is thus an abstraction that may be implemented in various ways by different systems. It is a critical building block for interoperable and heterogeneous information systems. Each digital object has a unique and, if desired, persistent identifier that will allow it to be managed over time. This approach is highly relevant to the development of third-party value added information services in the Internet environment.

DOI	type	index	handle data values
multiple typed values per DOI	URL	3	http://www.server.net/...
	URL	5	http://www.else.org/...
	TYP	11	Linking application
	PMD	16	http://metaquery.com/...
	right	27	rap:enquiry(10.1234/567
	hdl	45	0.type/linking

Figure 2. Resolution Structure of the Handle System

The handle system is accessible through public proxies or the handle client library a programmer's toolkit that allows the functionality of the DOI system to be embedded into applications.

Existing resolution systems tend to return a single value of a certain type and have limited application. The DOI being based on the handle system can return arbitrarily complex data depending on what has been deposited in the system. This ability to return more than one value is called multiple resolution and allows rich option sets to

be returned to a user.

Links from an identifier to any associated data may be made through the resolution mechanism. Such associated data may include [DOIs](#), or other identifiers, for versions, manifestations, aliases, components, and other related items, and the content of such links are decided by the implementers of the system, who determine the appropriate level of functional granularity required. Any related item may be defined within the context of a structured metadata framework.

4. Description

A core kernel of metadata is required for all application profiles developed using the system ensuring a known basis for building applications for resources described within the system. The primary requirement for this publicly declared metadata is to ensure that objects are disambiguated from each other - that is that they are uniquely identified.

Standardized metadata can be the foundation for interoperable metadata that can address some concerns about the dearth of digital rights management interoperability. This need for interoperable metadata was highlighted in the work of indecs [[INDECS](#)] which has made clear that "electronic trading depends to a far greater extent than traditional commerce on the way in which things are identified (whether they are people, stuff or deals) and the terms in which they are described, that is, metadata." indecs provides a framework for the creation of metadata schema supporting commerce transactions in information commodities.

The indecs analysis of metadata is fundamental to [DOI](#)'s role in providing description of items identified, and ensures extensibility and interoperability of data associated with DOIs. Indecs was designed as a fast track, infrastructure project focused on the practical interoperability of digital content identification systems and related rights metadata within multi-media e-commerce. The work of indecs led to a logical framework analysis, and also to practical which are a basis for the metadata component of the [DOI](#) system.

The indecs work drew three principal conclusions:

- All metadata is just a view - it is important that things are not mistakenly identified and that views are interoperable
- Almost all terms need identifiers - each of the values in a view must be defined and identified if other views are to recognize them leading to the requirement of controlled vocabularies
- Events are the key to interoperability - interoperability is achieved by relating things and people through events

These same conclusions are being reached in the workings of the RDF and Topic Maps communities and are becoming accepted statements for metadata theory

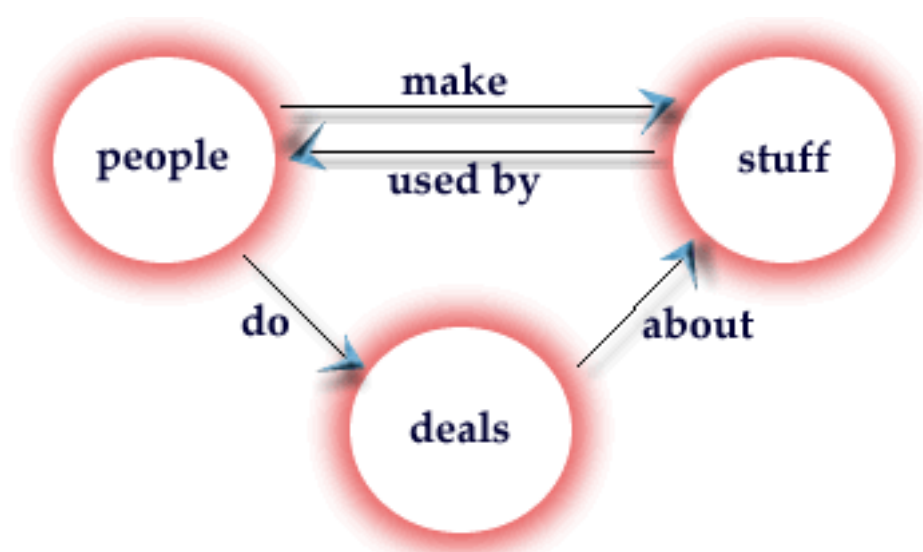


Figure 3. Illustration of the index structure

The interoperable metadata associated with each DOI allows mappings from any structured metadata implementations in specific sectors to provide the practical basis for the development of DOI Application Profiles. Application Profiles include a structured common set of attributes appropriate to the class of intellectual property concerned.

The DOI system requires that users operate within a defined application profile. Application profiles determine the schema requirements of mandated and optional metadata along with rules that govern the use of the information.

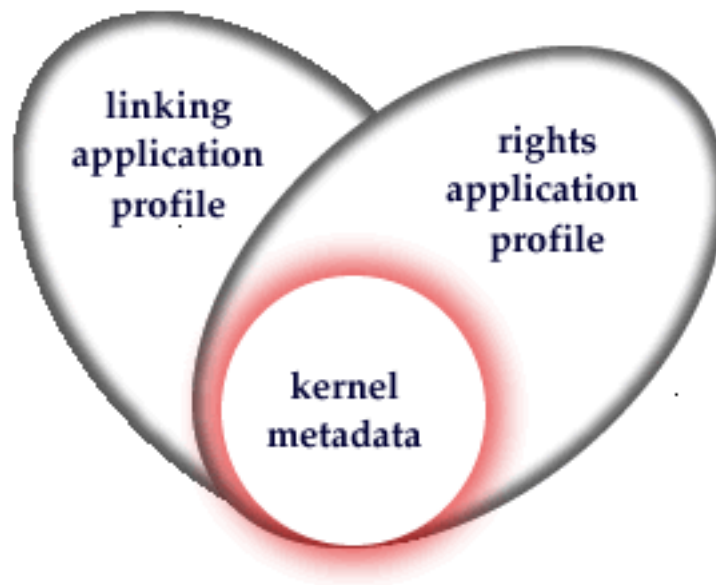


Figure 4. How kernel metadata relates to application profiles

By combining identifiers with metadata in a systematic manner it is possible to describe any media object in a common way that facilitates the building of interoperable information commerce systems.

5. Application profiles

Every entity that has a [DOI](#) must belong to a [DOI](#) Application Profile (DOI-AP). Whilst the schema of the DOI-AP is public, the individual content of each item's description is not public unless the [DOI](#) User Community concerned decrees this. The metadata rules (including business rules, such as access) for each DOI-AP are necessarily different, and may require the declaration of a more extensive metadata set as determined by the particular user community. In every case, though, the metadata does as a minimum incorporate the [DOI](#) kernel. The [DOI](#) kernel metadata schema is given in DTD form as:

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE doi:metadata [<!ELEMENT doi:metadata (doi:entity)+
>
<!ELEMENT doi:entity (doi:doi, (doi:profile)+ ) >
```

```

<!ATTLIST doi:entity doi:type (abstraction | tangible
manifestation |
    intangible manifestation | expression) #REQUIRED >
<!ELEMENT doi:doi (doi:prefix, doi:suffix) >
<!ELEMENT doi:prefix (#PCDATA) >
<!ELEMENT doi:suffix (#PCDATA) >
<!ELEMENT doi:profile (doi:identifier*, doi:title+,
doi:agent+) >
<!ATTLIST doi:profile doi:mode (visual | audio | audiovisual
| abstract |
    unknown) #REQUIRED
    doi:name CDATA #REQUIRED >
<!ELEMENT doi:identifier (#PCDATA) >
<!ATTLIST doi:identifier doi:scheme CDATA #REQUIRED >
<!ELEMENT doi:title (#PCDATA) >
<!ELEMENT doi:agent (#PCDATA) >
<!ATTLIST doi:agent doi:role CDATA #REQUIRED
doi:id CDATA #IMPLIED > ]>

```

An example of how this is used is:

```

<doi:metadata xmlns:doi="http://www.doi.org/doins/">
<doi:entity doi:type='intangible manifestation'&gt
<doi:doi> <doi:prefix>10.1045</doi:prefix>
<doi:suffix>january2001-mooney</doi:suffix> </doi:doi>
<doi:profile doi:mode='visual' doi:name='referencelinking'>
<doi:identifier doi:scheme='PII'>S
1031-5806(95)00403-9</doi:identifier>
<doi:title>Interoperability: Digital Rights Management and
the Emerging EBook
    Environment</doi:title>
<doi:agent doi:role='author'>Stephen Mooney</doi:agent>
</doi:profile> </doi:entity>
</doi:metadata>

```

As can be seen from this DTD it is possible for an identified entity to have more than one application profile so that various sets of services can be managed from the same identifier for intellectual property.

6. Administration

The implementation of the [DOI](#) infrastructure requires registration agencies to work with user communities to define the metadata requirements and enforce both global and application level policy. In doing so the registration agencies are creating the administrative environment that supports the maintenance of [DOIs](#) and the creation

of applications that rely upon the resolution services.

The creation of administrative interfaces allows for different parties to make assertions about a DOI. This provides a method of breaking free of the handle system data structure to provide more flexible means of administering the information about a DOI by all recognized participants in the supply chain. It is therefore important that registration agencies are established with the support of the communities that they represent and that they reflect the needs of those communities.

Options such as conditional access to DOIs and associated resolution information can take several forms. At a basic level, the Handle System allows read/write permission settings at the level of the individual DOI/value pair, and there may be many associated values per DOI. Thus a single DOI could resolve to 10 pieces of typed data, but only 5 of those would be publicly readable with the remaining 5 visible only to specified administrators.

Another issue that can be facilitated through appropriate administrative controls is contextualization: the same resolution information is not equally valid for the same identifier across all situations. Consider a library holding a local copy of an entity identified by a DOI: the associated information in the global DOI resolution system does not, and should not, account for the specifics of that local copy, but that is precisely the information required by a patron of that library. Contextualisation is a general issues for all identifiers in non-document contexts -- e.g., IP-telephony (enterprise dialing schemes taking precedence over, but linked to, global numbering).

7. Governance

The International DOI Foundation (IDF), a not-for-profit open membership organization set up for the purposes of developing and governing the DOI System, sets out the rules that govern the implementation and operation of the system. The IDF manages development, policy and licensing of the DOI to registration agencies and technology providers and advises on usage and development of related services and technologies [DOI].

The IDF was established in 1998 to provide governance to and raise awareness of the DOI, assuming a leadership role in the development of a framework of

infrastructure, policies and procedures to support the identification needs of providers of intellectual property in the multinational, multi-community environment of the network.

Major components of the IDF mission involve stimulating interest in and understanding of this framework, encouraging alliances and collaborative activities to explore in depth the complex issues to be addressed, and influencing the development of standards that will ensure the appropriate level of value-added and quality control across the spectrum of participation.

8. Developments

DOIs are currently in use, in prototype development, or under consideration for a wide range of applications. At the same time much work is going on to build on the initial success of the DOI and bring it to new content communities.

The DOI's initial focus, using the principles set out by indecs, is on identification of creations (intellectual property entities) but the solution is extensible to transactions. The IDF has an interest in further developing identification of other entities: the top level model of intellectual property commerce adopted in indecs assumes three high level entities: creations (identified by DOI, described using indecs-complaint terms); parties (users, creators, intermediaries); and transactions. The principles of DOI may be extended to any one of these entities, so that one might have "DOIs for transactions". The current policies relate only to creations and different rules would apply for other entities. The IDF is engaged in further work in the development of identifiers for transactions and has responded to an MPEG call for proposal in this area [MPEG]. The Handle system used by DOI enables the association with an identifier of any piece of state data; such state data could include appropriately expressed usage rules. The articulation of such rules requires further work in the Rights Data Dictionary space.

The largest deployment to date is to enable linking between millions of scientific journal articles within the CrossRef application. The Publishers International Linking Association runs CrossRef as a not for profit organization to allow the open linking across journal articles. Publishers can become members of PILA and can then deposit and lookup DOIs using supplemented metadata to locate the DOIs for

particular bibliographic references. These DOIs are then embedded into new scientific articles as they are published.

EBook publishers have recommended the use of DOIs to address their future requirements in a recent report developed in conjunction with Andersen Consulting. A project is now underway to define and prototype applications of DOIs for ebooks. Interest has also been expressed from communities as diverse as news, music, patents, software and accounting.

9. Conclusions

The DOI identifier exists in a well-defined environment – the DOI system. The DOI System is the means by which the DOI distinguishes itself from other identification systems. The DOI is a system that allows the registration of intellectual property resources, the maintenance of state information about those resources and a resolution system to access that state data.

A fundamental design principle of DOI is that each component is extensible: the enumeration syntax allows the maximum flexibility of assignment using newly created or existing identifier sequences; the descriptive metadata follows an extensible framework; the underlying resolution system can be extended across multiple dimensions, including the number and types of data items associated with DOIs, the protocols used to address the resolution system, the authentication methods which can be used both in resolution and administration, and the scaling and distribution of the core resolution service.

The full implementation offers DOI as a fully working identification system; however there is considerable additional development occurring (in collaboration with existing and potential users, and with many other organizations) in terms of both policies and additional functionality. In particular, the use of DOIs in rights management environments and other uses, offers significant potential benefits but requires further development of underlying technologies such as rights data dictionary and rights expression languages.

Bibliography

[KW] A Framework for Distributed Digital Object Services, Robert Kahn and Robert Wilensky, May 13, 1995, cnri.dlib/tn95-01, <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>

[NISO] ANSI/NISO Z39.84-2000 Syntax for Digital Object Identifier Syntax, <http://www.techstreet.com/cgi-bin/pdf/free/247384/z39.84.pdf>

[DOI] The DOI Handbook, Version 1.0.0, February 2001, <http://www.doi.org/>

[INDECS] Godfrey Rust and Mark Bide, The <indec> Metadata Framework, Principles, model and data dictionary, June 2000, <http://www.indec.org/pdf/framework.pdf>

[MPEG] ISO/IEC JTC1/SC29/WG11 N3942, January 2001, Revised Call for Proposals for Digital Item Identification and Description.

Glossary

DOI Digital Object Identifier

Biography

Eamonn **Neylon**

Consultant

International DOI Foundation

Oxford

United Kingdom

Eamonn Neylon - Eamonn has worked in both the publishing and software development industries. During eight years with the Thomson Corporation he developed several innovative systems for publishing on the Internet. Eamonn then joined software developer RCP Consultants where he oversaw several software maintenance releases and created the Lynkbase system. More recently he was employed by Yankee Rights Management to work on identifiers for rights management.