The DOI® Handbook

Version 1.0.0 February 2001

Please send any comments or questions to

Dr Norman Paskin
Director, International DOI Foundation
PO Box 233
Kidlington
Oxford OX5 1XU
United Kingdom

Email: n.paskin@doi.org
Tel: +44 (0) 1865 843545
Fax: +44 (0) 1865 843446

www.doi.org

10.1000.182

Contents

How to…

This section provides a short, practical guide on how users should approach various tasks related to the Digital Object Identifier.

How to obtain a DOI Prefix

To obtain a DOI Prefix, you need to apply to a DOI Registration Agency. In February 2001, the only DOI Registration Agencies operating are the IDF itself and CrossRef (www.crossref.org). Additional DOI Registration Agencies will be coming on stream during 2001, and up-to-date details can be found on the DOI web site.

If you wish to obtain a DOI prefix from the IDF, the information you need (and the necessary application form) can be found on the DOI web site at http://www.doi.org/started.html. The current charge for obtaining a DOI prefix direct from the IDF is a one-time fee of $1000. (An explanation of DOI prefix fees can be found under "How much does it cost to use a DOI?" in the Web site's FAQ list, http://www.doi.org/faq.html.)

How to select your Registration Agency

The number of RAs is expected to grow very quickly during 2001. Why should a potential registrant select one RA rather than another?

RAs are established to provide services on behalf of specific user communities. CrossRef, for example, is providing citation-linking services for the scientific publishing sector. Publishers will choose CrossRef as their Registration Agency because they wish to avail themselves of the specific service or services offered by CrossRef.

We anticipate the development of a growing number of RAs with sectoral specialisms of this kind, which may have global application. At the same time, we also anticipate a requirement for regionally based RAs, able to offer (for example) local language support. The smooth running of the DOI System will require close collaboration between different RAs so that registrants can avail themselves of the full range of services that are offered.

Any potential registrant seeking to register DOIs who cannot identify an appropriate RA able to meet their specific needs should approach the Director of the IDF. The IDF anticipates that it will itself remain a Registration Agency for the foreseeable future, to ensure that all market needs are fully met. It will not, however, compete with RAs that have an established market position.

How to register and maintain a DOI

The process and the rules for the registration and maintenance of DOIs and associated metadata will differ between different Registration Agencies.

Registering a DOI with the IDF

The IDF itself, as a Registration Agent, registers DOIs in the “Zero-AP” (see Appendix 2) with no associated mandatory declaration of metadata. There is therefore a very simple mechanism for the registration and maintenance of DOIs. CNRI provides the necessary technical services to support registration, under contract to the IDF.

Once your application for a DOI Prefix has been approved, and your payment processed, you will be notified by CNRI by email. The email message contains prefix information (prefix and

password), information about the location of the login forms (which can be found at https://admin.doi.org/DOI/login_form.html) and other instructions and general information. The administrative web forms provide the ability to create, delete, and update DOIs individually or in batch form.

At present, the batch loading process deposits each DOI and its corresponding URL into the DOI System. The file format is very simple; each line consists of a DOI separated from its corresponding URL by a space. In addition to the batch function from the web forms, there is a client-side batch application available from CNRI. This application is written in Java and can be automated and run from a Registrant's site.

It is anticipated that use of the Zero-AP will be restricted in future and that new DOIs will not be registered without metadata in other than exceptional circumstances. The IDF will soon offer registration of DOIs in the Base-AP as an alternative (see Appendix 2).

An example of a process including metadata declaration

The following procedure, adopted by the CrossRef project, is included here as an example of the process followed by an individual Registration Agency for the registration of a DOI with declared metadata.

This procedure allows for the batch registration of DOIs and associated metadata records into a DOI Central Metadata Directory run by the Registration Agency; this directory can subsequently be queried. The batch file format currently in use is XML as defined by a specific XML DTD, and submission is via HTTP POST. Security is HTTP basic authentication; PGP encryption will be added later. Batch receipt is confirmed to the sender via email.

Metadata Creation: The Registrant prepares XML batch files in accordance with the DTD; these are further constrained by a set of rules for the data, which define the expected content of each metadata element.

An XML batch may contain metadata for hundreds of DOIs, but is limited to 5 MB in total size, a limit established to avoid potential network errors that can result from sending large files over the Internet, and to accommodate the memory constraints of both the web server and the data validation process.

The development and implementation of quality control measures used to ensure the validity of the metadata content are the responsibility of the Registrant. An evaluation of the extent to which quality control and data checking can be managed by the metadata collection process has not yet been undertaken, but is viewed as an important next step.

Metadata Collection: The XML is validated upon receipt against the DTD. If the XML does not parse, the batch is refused; the Registrant must correct the XML and resubmit the batch. Again, it is anticipated that more extensive data checking could be accomplished during the validation process, for example by ensuring that date elements actually contain dates.

XML batches are submitted to a named HTTP server via HTTP POST to a Java "servlet", which parses and validates the XML file, and notifies the Registrant in real time whether or not the XML is valid and has been accepted. The submission process captures and verifies a DOI System prefix holder login and password prior to validating the XML. The XML files themselves contain timestamps used as identifiers of the batch; should the Registrant so wish, each DOI record may have its own timestamp.

DOI Deposit: The servlet then deposits each DOI and its corresponding URL into the DOI System as described above (in the process for registering a DOI without metadata). If the DOI is not new and therefore already exists in the DOI System, the timestamp is key to determining whether the DOI data being contributed is newer than the data that is already in the system; if so, the existing DOI data is updated. A log file also written in XML is created for each batch, indicating the total number of DOI records in the batch, the number of successful deposits into the DOI System, and the number of failures. For each failure, the DOI is provided, along with the reason for the failure. While DOI System failures may be the result

of system errors, they are most typically caused by an attempt to overwrite existing DOI data with older data.

Metadata Database Record Generation: The original XML batch files, along with the log files for the batches, are made available daily to the metadata database deposit process, where they are indexed and then made available for searching. A final XML log file is generated to indicate the success of the database deposit (again, failures are due primarily to network or system errors) combined with those from the DOI deposit process, and this combined XML batch diagnostic is emailed to the Registrant.

The entire metadata collection process is expected to be completed and reported to the Registrant in as close to real time as possible; 24 hours is currently seen as a reasonable target time. However, when Registrants initially make deposits, there are large amounts of legacy material and coordination is needed on when the legacy batches are deposited or system performance can be affected.

Data Querying The metadata database (MDDB) may be queried by submitting a batch file of known metadata fields in a specified format, currently pure ASCII text on separate lines, with fields delimited by vertical bars. The batch interface will query the database and return the corresponding DOIs (if known), or a diagnostic message. Batch query files are submitted by HTTP POST to a named HTTP server.

Future work flow developments

It is anticipated that individual Registration Agencies will develop their own workflow and procedures for the management of DOI registration, and metadata deposit and maintenance. This may prove a fruitful field of competition between Registration Agencies.

How to apply to become a Registration Agency

The following points of guidance are offered for any organization that is interested in pursuing a Registration Agency (RA) function with the IDF.

In the first instance a prospective Registration Agency should notify their interest to the Director of the International DOI Foundation. This will provide the opportunity for informal discussion to be instigated at an early stage and for advice to be given to the applicant about the requirements that the IDF has of a Registration Agency.

Initially, a Letter of Intent and an accompanying Terms document are available for consideration by prospective Registration Agencies. These allow the possibility of signature of the Letter of Intent prior to, or in parallel with, the further development of the Terms document. The Terms document is being further refined in collaboration with existing and prospective Registration Agencies. Ultimately, it is intended that the Letter of Intent will lead to a more detailed Agreement between IDF and each individual Registration Agency. Signature of the Letter of Intent by both the IDF and the candidate Registration Agency requires the organisation to become a member of the IDF and to enter into further negotiations.

The candidate Registration Agency will be asked to submit a detailed written proposal. Where a Registration Agency requires that its application and written proposal remains confidential, the parties will agree to the terms of an appropriate Non-Disclosure Agreement. In the event that the candidate Registration Agency chooses not to pursue its application, or the IDF Board rejects its application, the parties will agree to be bound by the terms of the Non-Disclosure Agreement.

Ideally a proposal to the IDF should address the following issues:

1. The ‘community of interest’ that is to be represented should be identified and delineated.
2. The DOI Application Profile or Profiles that the Registration Agency intends to implement should be described as fully as possible. As one element of this description, the applicant should submit such detail as is available of the metadata schema to be implemented. In

the event that elements of the metadata schema either do not exist or are under development, the applicant should set out the timetable for their development.

3. The candidate RA should provide a description of the business model it intends to adopt. The description should include an outline of the anticipated relationship with DOI Registrants, a revenue model and plans to establish the appropriate consultative framework within the community of interest.

4. The candidate should set out its technical and organizational plans for establishing a metadata repository. It is expected that some applicants will already have a mature systems environment in place for the storage of identifiers and metadata while others will be planning to implement their own repository for the purpose of becoming a Registration Agency.

5. In consultation with the Director with the IDF, the applicant will provide a preliminary assessment of the issues they will face in adopting Handle technology. Specifically some thought should be given to the potential scale of the implementation, reflecting the anticipated volume requirements of DOI Registrants. This is important in terms of anticipating the scalability of the Handle System and deciding whether or not the applicant may be required to host a Local Handle Server for this purpose.

A detailed document "IDF Terms of Governance and Policies with Registration Agencies" is available on a controlled basis for parties engaged in detailed discussions with the IDF on registration issues. This covers governance and policy issues as well as technical and operational issues that need to be addressed in formalizing the relationship between the IDF and a new Registration Agency.

How to develop a DOI Application Profile

If you believe that you can establish a “community of interest” to develop a DOI Application Profile (see Appendix 2) please contact Dr Norman Paskin, the Director of the International DOI Foundation: n.paskin@doi.org before commencing any development work. This will avoid any duplication of effort with other groups who may already be working on similar projects.

How to become a member of the International DOI Foundation

Information about joining the IDF can be found at:

http://www.doi.org/membership/brochure.html

An application for membership can be found at

http://www.doi.org/membership/interactive-form.html

Glossary

This short Glossary is not intended to be a complete guide to all the technical terms used in the Handbook. Its primary purpose is to avoid ambiguity and misunderstanding by pulling together definitions of certain terms that are defined (implicitly or explicitly) in the text. Where (DOI) appears in brackets, this is to indicate that it has been ignored from the point of view of sorting the entry alphabetically.

(DOI) Application Profile (DOI-AP): the functional specification of a specific application (or set of applications) of the DOI System to a class of intellectual property entities that share a common set of attributes.

(DOI) Base-AP: a DOI-AP that requires declaration only of the Kernel Metadata and follows the general rules of the IDF.

DOI: Digital Object Identifier, the opaque string used as an identifier by the DOI System.

DOI System: the integrated system – comprising enumeration, description, resolution and policymaking – managed by the International DOI Foundation (IDF).

Entity: something that is identified.

IDF: The International DOI Foundation, the organization established to manage the DOI System.

Intellectual property: rather than attempt our own definition of what "intellectual property" may be, we depend instead on definitions agreed by the World Intellectual Property Organization and related international treaties like the Berne Convention.

Intellectual property entity: the scope of things that may be identified with a DOI.

(DOI) Kernel Metadata: the minimum set of metadata that it is mandatory to declare alongside every DOI (and which is available for public inspection).

Metadata: data that describes something. In this document we are concerned only with **DOI Metadata**, metadata that describes something identified with a DOI.

(DOI) Registrant: an organization that obtains a DOI prefix and registers DOIs.

(DOI) Registration Agency (RA): an organization which offers services to **(DOI) Registrants**, including prefix allocation, registration of DOIs, and maintenance of sufficient infrastructure to allow Registrants to declare and maintain DOI metadata and state data.

Resolution: the process of submitting an identifier to a network service and receiving in return one or more pieces of current information related to the identifier

(DOI) User Community (DOI-UC): the organization of users (including at least one RA) established to define and manage a DOI Application Profile.

(DOI) State data: the data associated with a DOI in the Handle System (to which the DOI resolves).

(DOI) Zero-AP: All DOIs that were earlier registered without metadata declaration will eventually be migrated to either appropriate Application Profiles (adding the requisite metadata), or to a default special AP, the Zero-AP that has no associated metadata and therefore very limited functionality.

1 Introduction

The introduction sets out the background to the establishment of the IDF itself and the implementation of the Digital Object Identifier (DOI).

1.1 Identifiers and the Internet

The future of the "content industries" is now totally intertwined with the future of the Internet and its successor networks, in a fundamental shift from physical to electronic dissemination. All varieties of intellectual property – including books, recorded music, academic journals, video, software and games, as well as types yet to be invented – will be swept along in this migration.

The boundaries that currently exist between different types of content, especially at the level of the infrastructure that supports their production and distribution, will be broken down and ultimately eliminated. Instead of different physical formats requiring different content distribution infrastructures, all content will consist of streams of digital data moving over networks. Diverse content industries will increasingly find themselves sharing the same challenges and opportunities in delivering content to their customers, whether direct or through intermediaries.

One of the key challenges in the move from physical to electronic distribution of content is the rapid evolution of a set of common technologies and procedures to identify, or name, pieces of digital content ("digital objects"). A widely implemented and well understood approach to naming digital objects is essential if we are to see the development of services that will enable content providers to grow and prosper in an era of increasingly sophisticated computer networking.

The International DOI Foundation (IDF) was established in 1998 to address this challenge, 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.

The main activity of the IDF is to encourage the widespread implementation and use of a standard digital identifier: the *Digital Object Identifier* (DOI), an "actionable identifier" for intellectual property on the Internet.

1.2 The DOI – a period of rapid development

The DOI is currently undergoing a period of rapid development. That development is evident in all aspects of the DOI System – technology, procedure and policy (see Chapter 3).

However, despite the speed of change in the DOI System, it has become evident that it is essential to codify the main elements of the DOI in a form that can be easily assimilated by the non-technical reader, and yet at the same time to provide a central point of reference for more complex technical content.

The main text of the Handbook is designed to be accessible to any reader: it starts in Chapter 2 with a high-level introduction to the DOI for those coming to the subject for the first time. The following chapters are more detailed, but use technical terminology only to the extent that it is necessary. A glossary is included for some of the terms used in these chapters (see page 11).

The Appendices provide a core of more complex technical documentation, while the list of references and further readings seeks to provide the necessary links for the most demanding of users.

1.3 The DOI Handbook – a period of rapid development

Because the DOI itself is a work in progress, so this Handbook is in a state of constant development and updating. Much that may be expressed in the future tense in this release of the Handbook will be in the present or the past tense in a matter of months.

The primary publication medium of the Handbook is the World Wide Web; the most recent release is always be available online. However, small quantities of the Handbook will also be printed on demand; users are advised to ensure that they are using the most up-to-date version by checking the version on the DOI Web site (http://dx.doi.org/10.1000/182).

Whenever a chapter section (for example this Section 1.3) is added, deleted, or substantially modified, a new version number will be allocated to the Handbook. The numbering system follows the convention of **edition.release.update** (the most significant digit on the left). Minor changes such as typographical corrections with no substantive effect will be numbered as updates; more substantive changes as releases; major changes as editions. Criteria for numbering are pragmatic: the IDF's aim is to clearly distinguish new versions for users, especially when use of an earlier version may result in error

The first pre-release version of this Handbook (V0.1) was published in June 2000. Following a process of considerable update during its first 6 months of public availability, a number of significant changes have been made, including the inclusion of an additional Appendix (3). The addition of this Appendix, and some significant changes in terminology (particularly introducing the DOI User Community and the DOI Application Profile to replace the DOI Genre) have persuaded us that the Handbook is now (in February 2001) sufficiently mature to introduce this first formal release, Version 1.

We are confident that all those who use the DOI, or who are thinking of doing so, will find this document of assistance.

2 What is a Digital Object Identifier?

This chapter provides a high level overview of the Digital Object Identifier (DOI) for those who have no previous knowledge of it. Those who already have some understanding of the basic issues of identification and of the DOI itself should go straight to Chapter 3.

2.1 Background

As commerce has become increasingly less dependent on the physical presence of both buyer and seller, means of identifying things uniquely and describing them unambiguously have become more and more important. The use of computers in mediating some aspects of the trading relationship has further accentuated this requirement. The near-universal adoption of "unique identifiers" such as the ISBN or the EAN barcode has been a direct consequence of (and a precondition for) the development of EDI (electronic data interchange) and electronic trading of all kinds.

The Internet, as it becomes a medium for trading in intellectual property, drives us several steps further. The digital network linking trading partners has for the first time to embrace consumers rather than simply supporting business-to-business transactions. The identity of the things that can be traded becomes much less clearly delineated when they may be computer files rather than physical objects. Users no longer have to access "content" only in pre-packaged products – it becomes possible to provide them with the precise customized package of content that they want (and which, theoretically at least, no one else may want).

The management of the myriad transactions implicit in such a complex network environment will only be possible if the transactions are mediated by computer systems. This puts additional pressure on the requirement for unambiguous identification and description of the content – the "metadata" that has become the buzzword of e-commerce in intellectual property. Persistent identification and description is a prerequisite for the management of intellectual property rights in the digital environment.

2.2 Origins

It became increasingly apparent during the 1990s that existing approaches to identification would prove inadequate to meet the need. Publishers, for example, could see the deficiencies of the ISBN as an identifier for electronic publishing, because of its limitation to identifying physical objects, and the difficulties with applying it to items smaller than "a book". At the same time, the only content identifier commonly in use on the Internet – the Uniform Resource Locator (URL) used to find particular pages on the World Wide Web – was clearly deficient, not least because it was not used to identify *content* but rather *location.* The location is transient, whereas what was necessary was a means of identifying content itself, persistently and without ambiguity.

What has now developed, from a research project begun by the American Association of Publishers in 1996, is the complete **DOI System** described in this Handbook. Since 1998, the DOI has been managed by the International DOI Foundation (see Chapter 12).

2.3 The DOI – a "persistent" identifier

The DOI can be described as "persistent identifier of intellectual property entities". This requires some further explanation.

Firstly, a definition: "entity" is a term that we will use throughout this handbook, and it is important to understand what **we** mean by it: by our definition, it is simply something that is identified. (The underlying idea, borrowed from the <indecs> project, is that nothing exists *in any useful sense* until it is identified.)

There are many synonyms and near synonyms for this term “entity” as we use it. The equivalent term often used by the World Wide Web community is “resource”.

So, what is “intellectual property”? We all know instinctively what we mean by “intellectual property” – but do we always mean the same thing? One definition that has been advanced is “works of human intellect or imagination” – which may take us a part of the way towards common understanding, but perhaps not all of the way.

Rather than attempt our own definition of what “intellectual property” may be, we depend instead on definitions agreed by the World Intellectual Property Organization and related international treaties like the Berne Convention.

So, 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 (see Chapter 9).

Critically, 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.

2.4 The DOI – an “actionable identifier”

The DOI goes beyond simply providing a scheme for the unique and persistent naming of intellectual property entities in the network environment. The identifier itself is simply one element of a complex *system*; the system is described in Chapter 3 and the chapters that follow it. The purpose of the **DOI System** is to make the DOI an *actionable* identifier. A user can use a DOI to *do* something.

The simplest action that a user can perform using a DOI is to *locate* the entity that it identifies. In this respect, a DOI may look superficially like a URL. However, the technology which underlies the DOI facilitates much more complex applications than simple location; and the DOI identifies the intellectual property entity itself rather than its location.

2.5 The DOI – an “interoperable” identifier

The DOI System has been designed to interoperate with past, present and future technologies.

- So-called “legacy” identifiers, those we have used in the past and continue to use today – like the ISBN – can form an integral part of the DOI naming system. Businesses can continue to use familiar – and proven – naming or numbering systems in this new environment.
- The DOI, even in its simplest implementation, provides an actionable identifier on the Internet which is fully compatible with URLs and the World Wide Web, providing users with a persistent identifier that can overcome the problem of entities that are relocated from one place to another on the Web (because of change of ownership or simply for administrative reasons).
- The more sophisticated and complex applications that are being developed as parts of the DOI System will be fully compatible with the standard environment of the Internet as it develops. The core Handle System technology that the DOI uses will always be “open standards” based.

2.6 Development of the DOI System

The DOI was first demonstrated (in a relatively simple form) in 1997. Since then, the International DOI Foundation has initiated a process of continuous development and

improvement, in terms of technology, processes and policy. Some aspects of the DOI System have now progressed far enough to be formally standardised (the “syntax” of the number itself – see Chapter 4 and Appendix 1). However, other aspects of the DOI are still subject to rapid change and development.

2.7 Who is using the DOI System?

Two and a half million DOIs have already been registered, by 150 different registrant organizations. Some examples of how DOIs are being used can be found at http://www.doi.org/gallery/tour.html. These examples come mostly from traditional print-publishing companies that have already established major online publishing programs. This reflects the origins of the DOI.

The IDF is working closely with many businesses in other sectors of the “content industries” to extend the application of the DOI to many other types of intellectual property. We will see live application of the DOI in several other sectors during 2001.

3 The components of the DOI System

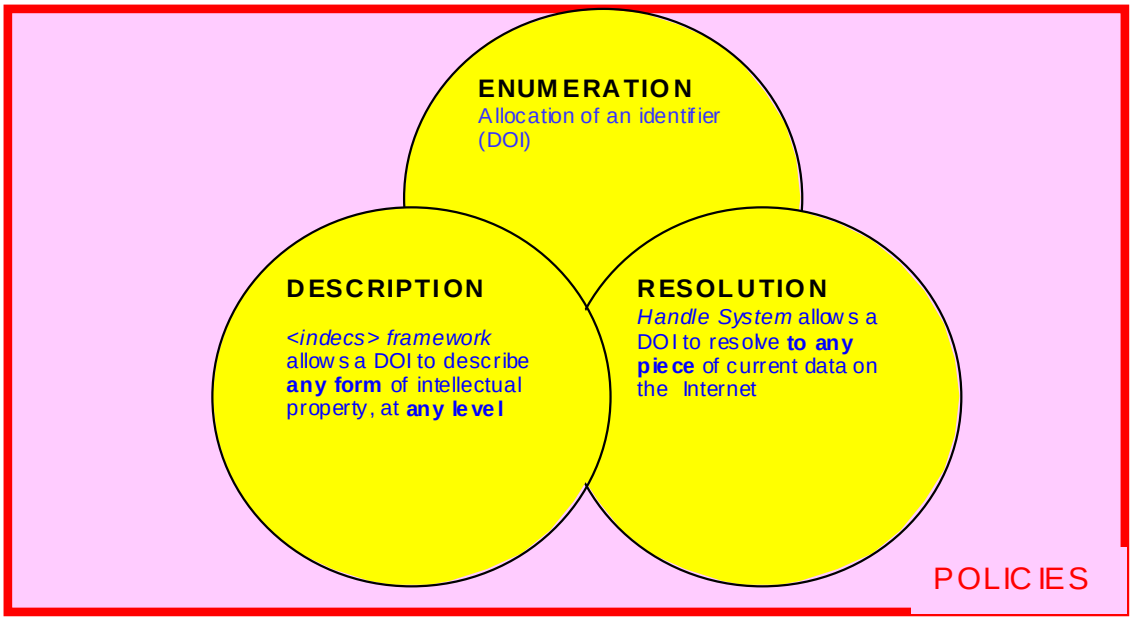
This chapter is a brief, non-technical description of the different components of the DOI System, and describes how those components fit together. It is intended to be read by anyone who has an understanding of the basic issues relating to identification of intellectual property that are covered in Chapter 2; it provides an introduction and framework for the more detailed chapters that follow.

3.1 Overview of the DOI System

The DOI is more than just a way of naming things – it is an integrated system. The **DOI System** is made up of a number of interacting components that depend on one another for their value. The whole is much greater than simply the sum of the parts.

However, to understand the value of the DOI System, we must first deconstruct it. In this Handbook, we will view the DOI System as being made up of four primary components:

- **Enumeration**: assigning a number (or name) to the intellectual property entity that the DOI identifies (an "intellectual property entity" is defined in Chapter 2). It is more correct to talk about the DOI as an alphanumeric string, since a DOI may contain characters as well as numbers. However, we will use the term "number" to apply to this string, to avoid unnecessary complexity.
- **Description**: creating a description ("metadata") of the entity that has been identified with a DOI.
- **Resolution**: making the identifier "actionable" by providing information about what the DOI should resolve to, and the technology to deliver the services that this can provide to users.
- **Policies**: the rules that govern the operation of the system.



3.2 Enumeration

Each DOI is a unique “number”, assigned to identify only one entity. Although the DOI system will assure that the same DOI is not issued twice, it is a primary responsibility of the Registrant (the company or individual assigning the DOI) to name each object within a DOI prefix uniquely. That uniqueness is enforced by the DOI System.

It is important for the integrity of the system that the same number is not used twice to identify different things; it is equally desirable that two DOIs should not be assigned to the same thing (although the same thing may have other, different identifiers applied to it for other applications – a book may have both an ISBN and a DOI).

The DOI is designed in such a way as to make it as simple as possible for anyone to name uniquely any item of intellectual property – tangible or intangible, in physical or digital form. Existing identifiers – like the ISBN – can be used as part of the DOI, which should make it much easier for registrants to issue DOIs to all their existing “content assets”. The structure of a DOI is explained in detail in Chapter 4.

However, the DOI goes much further than most existing identifiers, in being able to identify much smaller “fragments” of content – and types of intellectual property for which no existing identification scheme (or “legacy identifier”) exists.

In use, the DOI is an “opaque string” or “dumb number” – nothing at all can or should be inferred from the number. Some identifiers include some “intelligence” in the number itself, which will tell the user something about the entity that is being identified (or about the person or organization that registered the identifier). However, the only secure way of knowing anything about the entity that a particular DOI identifies is by looking at the metadata that the Registrant of the DOI declares at the time of registration. This means, for example, that even when the ownership of a particular item changes, its identifier remains the same – in perpetuity. This is why the DOI is called a “persistent identifier”.

3.3 Description

If an identifier is to have value, it is essential that its user can identify unambiguously what it identifies. Because there is no intelligence in the DOI itself, “metadata” (a term which we use here in the sense of “descriptive data”) is an *essential* component of the DOI System. The DOI itself is, of course, a rather specialized piece of metadata

The idea of what may constitute metadata has expanded enormously over the last decade; almost any piece of data can be seen as describing some other piece of data. We need to set appropriate limits on the amount of metadata which *must* be declared at the time when a DOI is registered. For this reason, there will be only a very small “kernel” of metadata that it will be mandatory to declare with every DOI:

- **Identifier**: an identifier associated with the entity from a legacy identification scheme (where such an identifier exists).
- **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, audiovisual).
- **Primary agent**: the identity of the “primary agent”, normally the first-named creator of the entity.
- **Agent role**: the role that the primary agent played in the creation of the entity.

This metadata will be available to any user of the DOI System, to enable them to find a basic description of the entity that any particular DOI identifies. This basic description will allow the user to understand some basic things about the entity.

However, we recognize that the structure of the metadata required to describe different types or "classes" of intellectual property entity will be different. The description of a journal article is different from the description of a musical recording or a photograph. What is more, the elements of metadata that are needed for different purposes or different *applications* are different (see Chapter 4 and Appendix 2). The **DOI Application Profile** within which a particular entity is registered also forms part of the kernel metadata.

Every entity that is identified with a DOI will be assigned to at least one and possibly more DOI Application Profiles (DOI-AP). The precise rules for the metadata to be declared for an entity in each DOI-AP are managed separately, but the metadata for all DOI-APs *includes* the kernel elements that are publicly available.

In some DOI-APs, the exploitation of the extended metadata is an integral part of the business model of the registration process (see Chapter 11). A more detailed technical description of the requirements for DOI metadata can be found in Chapter 5.

3.4 Resolution

The DOI System is different from most other identification systems in being *actionable*: a DOI on the Internet can be "resolved", leading the user of a DOI to any piece of data that is Internet-accessible.

This does not imply that the DOI will necessarily resolve to the entity that it identifies – although that will sometimes be the case. The DOI, though, can be used to identify classes of intellectual property –abstract "works", physical "manifestations", performances – that cannot be directly accessed in a digital file. Even when the DOI does identify a digital file, this will not always be the most appropriate or useful data for the DOI to resolve to.

It is very important to distinguish what the DOI *identifies* from what the DOI *resolves to.* They may be the same thing but they will often be very different.

The technology that is used to manage the resolution of the DOI is called the Handle System; a high-level technical description can be found in Appendix 5.

The Handle System is unlike most other resolution technologies in supporting *multiple resolution*. A DOI may have multiple data values of different types associated with it (email addresses and URLs, for example), and multiple data values of the same type (several URLs). The same DOI can resolve to different data, depending on the way in which the Handle System is queried. This enables the DOI, and the metadata with which it is associated, to form the foundation for many different *services* relating to the management of intellectual property in the network environment, to the benefit of intellectual property owners and users alike.

In order for the DOI to be resolved, the Registrant needs to maintain the data associated with that DOI in the Handle System; this data is referred to as "state data". The simplest form of state data is a single URL. However, a DOI can resolve to many other forms of data, including for example another DOI.

By using the flexibility and power of the Handle System, much of the potential for delivery of intellectual property on the Internet can finally be unlocked. More detailed information on resolution of DOIs will be found in Chapter 6.

3.5 Policies

Any system requires rules for its operation, and the DOI is no different. Indeed, the flexibility and power of the DOI System make policies more rather than less important if the DOI is to achieve its potential. The DOI can be distinguished from other identification schemes, in particular from other implementations of the Handle System, by its policies. These policies ensure that the DOI System provides reliable and predictable results to the user. Reliability and predictability can be regarded as the metrics of success of the *quality assurance regime* provided by the DOI's policy framework.

The IDF aims for simplicity and consistency in its policy. Such an approach does not guarantee success; but its absence surely guarantees failure.

An overview of the major policies adopted by the IDF can be found in Chapter 7. There are many rules implicit in the description of the DOI System that is contained in this Handbook; additional policies are being developed by the IDF all the time. For example, the administrative processes through which DOIs are managed are all the outcome of DOI policy decisions. In many ways, the establishment of such policy rules often proves more complex and difficult than managing the technology.

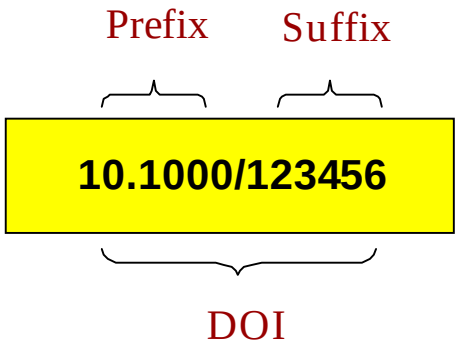
The IDF serves two distinct but related purposes: it provides governance for the DOI System as a whole (see Chapter 12) and acts as the “maintenance agency” for the technical standards that are essential to the operation of the DOI system, whether those are formal, externally-recognized standards or simply matters of internal procedure (see Chapter 10).

4 Enumeration

This chapter explains how a DOI is constructed and assigned. It discusses the use of the prefix and outlines the potential for the user of existing ("legacy") identifiers within the DOI System.

4.1 The structure of a DOI

The DOI has two components, known as the prefix and the suffix. These are separated by a forward slash. The two components together form the DOI.



DOIs may incorporate any printable characters from the Universal Character Set (UCS-2), of ISO/IEC 10646, which is the character set defined by Unicode v2.0. The UCS-2 character set encompasses most characters used in every major language written today.

However, because of specific uses made of certain characters by some Internet technologies (the use of pointed brackets <> in xml for example), there may some effective restrictions in day-to-day use (see Appendix 1 for guidance).

There is no technical limitation on the length of either the prefix or the suffix; in theory, at least, there is an infinite number of DOIs available.

4.2 The DOI prefix

The prefix itself has two components.

All DOIs start with "10." This distinguishes a DOI from any other implementation of the Handle System (see Chapter 4). The next element of the prefix is the number (string) that is assigned to an organization that wishes to register DOIs.

There is no limitation placed on the number of DOI prefixes that any organization may choose to apply for. For example, a publishing company might have a single DOI prefix, or might have a different one for each of its journals, or one for each of its imprints.

This use of different prefixes within one organization may prove administratively convenient. It can help with ensuring that unique numbers are allocated (it is not always easy within a large organization to maintain uniqueness of suffixes unless numbers are centrally allocated). It may also help if some part of an organization (such as a journal) is transferred to the control of another organization. If all of the entities that make up that part of the organization share the same DOI prefix, it can make transferring responsibility for the relevant DOIs rather more straightforward. (See also Section 11.2)

Blocks of DOI prefixes are allocated to DOI Registration Agencies (see Chapter 11) for them to allocate to individual user organizations. All DOI prefixes so far issued have been simple numeric strings, but there is nothing to prevent alphabetical characters being used.

For the time being, only this relatively simple form of DOI prefix is in use. In future, it is anticipated that the prefix may be further divided into sub-prefixes:

10.1000.10/123456

Remember, though, that the DOI is an opaque string (a dumb number). No definitive information can or should be interpreted from the number in use. In particular, the fact that the DOI has a prefix issued by a particular organization should not be used to identify the owner of any given intellectual property – the DOI remains persistent through ownership changes, and the prefix is unaltered.

4.3 The DOI suffix

Following the prefix (separated by a forward slash) is a unique suffix (unique to a given prefix) to identify the entity. The combination of a prefix for the Registrant and unique suffix provided by the Registrant avoids any necessity for the centralized allocation of DOI numbers.

The DOI suffix can be any alphanumeric string that the Registrant chooses. This can simply be a sequential number, or it can make use of an existing (legacy) identifier. The latter may often be administratively convenient for the Registrant.

Both of the following would be valid as DOIs:

10.1000/123456

10.1000/ISBN1-900512-44-0

When a legacy identifier is incorporated into the DOI in this way it is not intended to be interpretable (the “opaque string” rule).

4.4 The DOI and uniqueness

It is critical that the combination of prefix and suffix is unique, in order to support the integrity of the DOI System. The issuing of unique prefixes to Registrant organizations places the onus on those organizations to ensure that the DOIs that they are registering are indeed unique. However, the DOI System will make internal checks for uniqueness at the time of registration.

It is good practice never to reissue any unique identifier that has been once issued in error.

4.5 Standardization of the DOI syntax

The DOI syntax is a NISO standard (see Appendix 1)

4.6 Prefix, suffix, and character set issues

DOI syntax is covered in depth in Appendix 1. This section adds a few explanatory comments about DOI syntax (prefix and suffix) in relation to character sets, standards and implementation.

When thinking about prefixes, suffixes and character sets, it is important to distinguish DOI from the underlying technology, the Handle System (see Appendix 5). The DOI is a Handle

System implementation. *Current* usage (though not the only possible or potential usage) takes place almost entirely within the context of the World Wide Web (which is not the same as the Internet) and is governed by an evolving set of IDF policies.

Prefix/suffix. Neither the Handle System nor DOI policies, nor any web use currently imaginable, impose any constraints on the *suffix*, outside of encoding (see below). Handle syntax imposes two constraints on the *prefix* – both slash and dot are “reserved characters”, with the slash separating the prefix from the suffix and the dot used to extend sub prefixes. The root administrator for the Handle System has reserved all prefixes starting with “10.” (for example 10.1000, 10.1000.1, 10.23) for the IDF to use for DOIs.

Encoding. The Handle System at its core uses UTF-8, which is a Unicode implementation (further details are given in Appendix 1 of this Handbook, Appendix E of the *NISO standard*) and so in its pure form has no character set constraints at all: any character can be sent to, stored in, and retrieved from a handle server. The IDF imposes no additional character set constraints. In practice, though, there are many character set constraints enforced by the current web environment, depending on the individual user's context – for example, what kind of browser is being used. (This is something of a moving target – does your current browser display kanji characters, for example? Do you know?)

Implementation. It is essential to consider standards and the practical realities of implementation together. So, for example, it is imperative to “hex encode” the character “# “ in a URL, since this character is used to indicates the beginning of a URL fragment. The character means nothing special to the Handle System or in DOI syntax: nonetheless, a handle contained within a URL must have the # character encoded, otherwise a browser will abbreviate the handle at the # sign. This is true across all *web* implementations. The need to “hex encode” other characters, for example “< “or “ >”, varies with a particular browser implementation. Such required encoding in the DOI syntax is considered within the NISO standard (see Appendix 1). In a more general sense, any implementation of identifiers in a digital context needs to consider likely encoding issues that may be encountered, and should address character set constraints and the need to move those characters through environments such as the web in such a way that they pass through unaltered.

5 Description

This chapter explains why interoperable metadata is an essential component of the DOI System and lays out the structure and content of DOI Kernel metadata. It also explains the concept of the DOI Application Profile. Here as elsewhere, we use the word "metadata" in its most commonly accepted meaning – data that describes something.

5.1 Why does the DOI require metadata?

Identifiers are simply names – names that follow a strict convention and are unique if properly applied, but names just the same. Unique identifiers are particularly valuable in machine-mediated commercial environments, where unambiguous identification is crucial.

Some identifiers tell you something about the thing that they identify – the identifier "ISBN 1-900512-44-0" can reasonably safely be assumed to identify a book (always assuming that ISBN rules have been correctly followed, which is not universally the case).

However, to find out *which* book it identifies, it is necessary to consult metadata – the identifier links the metadata with the entity it identifies and with other metadata about the same entity. Metadata is an integral part of making the identifier useful. Some of this metadata may be held in private systems (the publisher's warehouse system, for example) but some of it is more widely available (e.g. *Books in Print*).

If the DOI were simply a system providing persistent single point location on the Internet, then metadata would not be essential to its function (see Chapter 2). However, the DOI is conceived as much more than that. In order for the DOI to be able to fulfill its wider potential in providing the basis for a full range of services relating to intellectual property in the network environment, metadata becomes an essential component of the DOI System as a whole.

The DOI can identify any kind of intellectual property entity, and because it is an "opaque string", the user can tell nothing about what it identifies from just looking at the DOI. This makes it all the more important that the user can access and inspect metadata related to the DOI, since the entity it identifies *may not itself be open to direct inspection* – it may be an abstract "work" or a performance.

Metadata is an essential component of the DOI System, and declaration of a limited "kernel" of metadata is becoming mandatory for *all* DOIs that are registered. This "kernel" of metadata has been designed to be as limited in scope as possible and equally applicable to any entity that can be identified with a DOI.

5.2 The significance of "well-formed" metadata

Metadata permits both recognition of the entity that is identified by a DOI and its unambiguous specification; it also allows for the interaction between the entity and other entities in the network (and with metadata about those entities).

The future of a significant proportion of intellectual property dissemination lies in the network environment, and it is inconceivable that commerce in IP entities will not be mediated by machines. This implies that the metadata that supports the management of that intellectual property must be machine interpretable.

Computers do not manage ambiguous information successfully in the way that human beings often can. It is essential for information presented to computers to be uniform and unambiguous.

The development of data models to manage information in computer systems is a well-established discipline. What is unusual for all who are now engaged in the management of intellectual property is the scope of the computer system within which we have to manage our data – the whole of the Internet.

This provides considerable challenges to anyone designing metadata to be used in the network environment.

The DOI accepts the analysis of the <in*d*ecs> project of the requirement for "well formed" metadata (see Appendix 6). This does not propose that all metadata for intellectual property has to be managed in a single metadata scheme. It does though propose that all such metadata needs to be "well formed"; this will allow metadata developed in conformance to different schemes to interact or "interoperate" unambiguously. Without that interaction, different metadata schemes risk becoming the "trade barriers" of the future.

What does it mean for metadata to be "well formed"? There are only two types of metadata that can be regarded as well-formed.

The first of these are labels: the names by which things are called (of which "titles" are a subset). These are by their nature uncontrolled and broadly uncontrollable. Identifiers are a specialized type of label, created according to rules, but names nevertheless. The fact that they are created in accordance with a prescribed syntax makes them less prone to ambiguity than other types of label and therefore more readily machine-interpretable than completely free-form labels.

All other metadata (if it is well formed) needs to be drawn from a controlled vocabulary of values, which are supported by a data dictionary in which those values are concisely defined. This means that the values in one metadata scheme (or in one "namespace") can be mapped to those in another scheme; this mapping may not be exact – where two definitions in one scheme both overlap with (but are not wholly contained within) a single definition in another, for example. However, the use of a data dictionary avoids the sort of ambiguity that is inherent in natural language, where the same word may have very different meanings dependent on its context. Where precision of meaning is essential, human beings can clarify definition through a process of dialogue. This is not generally the case with computers.

5.3 Adopting the <in*d*ecs> metadata analysis

The mapping between different metadata schemes may be more or less exact. It may also involve considerable loss of information, or no loss of information at all. It is obviously advantageous to achieve as close a mapping as is possible; this is most easily achieved between schemes that share a common high-level data model. The <in*d*ecs> data model underlies all DOI metadata (both the kernel and all Application Profile extensions). The same analysis underlies the EPICS data dictionary (which, through ONIX International, is rapidly becoming widely accepted as the metadata dictionary for the publishing industry internationally); similar developments are anticipated in other media sectors.

The <indecs> data model was devised to cover the same field of endeavor as the DOI – all types of intellectual property ("creations" in <indecs> terminology). It is an open model, which is designed to be extensible to fit the precise needs of specific communities of interest.

It was also designed to be readily extensible into the field of rights management metadata, the data that is essential for the management of all e-commerce in intellectual property. The <in*d*ecs> analysis asserts that it is essential for the *dynamic* data necessary for the management of rights to be built on a foundation of the rather more *static* data that identifies and describes the intellectual property, and that these two layers of metadata can easily interoperate with one another.

The adoption of the <in*d*ecs> metadata model gives DOI metadata a firm basis in an intellectual analysis of the requirements for metadata in a network environment that has been tested in real world applications. It should provide easy interoperability with other metadata schemes constructed using the same analysis, and a basis for interoperability with metadata schemes based on alternative analyses. It will also allow for interoperation with rights metadata based on the common <in*d*ecs> analysis, as this emerges over the next year or so.

The <in*d*ecs> metadata model does not of itself answer the challenges of managing metadata for intellectual property in the network environment, but it does provide a platform on which answers can be based.

<indecs> was a time-limited project, which finished its work early in 2000. Its output is highly regarded and its analysis has been adopted in a number of different implementations. However, it does not greatly matter to the DOI whether the <indecs> analysis and developments based on its framework come to be widely used for the management of intellectual property on the Internet (although we believe it will be very helpful if they do.) What matters to the DOI at this stage is whether DOI metadata itself provides a good basis for the management of intellectual property entities in the network environment. We are convinced that good data models, based on rigorous analysis, will be essential for this purpose. For example, we see libraries as likely be looking to IFLA's FRBR work as a basis; FRBR maps excellently to <indecs>. Data dictionaries and transfer protocols based on the <indecs> analysis are already being implemented in commercial contexts.

However, only time will tell whether these implementations are useful (and therefore successful). All a DOI needs is the few kernel elements and a map to a consistent data model. We use an underlying model as a way of guaranteeing that those few elements are useful when people want to extend on them. The reason for using the <indecs> model is that it is alone in having demonstrated its extensibility to rights management.

5.4 The DOI kernel metadata set

The following table sets out the DOI kernel metadata that must be declared whenever a DOI is registered. Every entity that has a DOI must belong to a DOI Application Profile (DOI-AP: see Section 5.5 and Appendix 2). The metadata rules for each DOI AP are different, and *may* require the declaration of a more extensive metadata set. In every case, though, the metadata will as a minimum incorporate the DOI kernel.

Element	Definition	Status	Number	Allowed values	Possible AP qualifications
DOI	A DOI	Mandatory	1 only	DOI	
DOI-AP	A class of entities with common attributes	Mandatory	1 minimum	From DOI-AP tables	
Identifier	A unique identifier (e.g. from a legacy scheme) applied to the entity	Qualified by AP	1 minimum	Any alphanumeric string but when present must include an identifier type, e.g. ISBN	Define in application: it is normal to include a legacy identifier if one exists.
Title	A name by which the entity is known	Mandatory	1 minimum	Any alphanumeric string	Define in application; a value of "untitled" may be allowable in certain APs.
Type	The primary structural type of the entity	Mandatory	1 only	From: *Abstraction* *Tangible Manifestation* *Intangible Manifestation*[1] *Performance*	
Mode	The primary sensory mode by which the entity is intended to be perceived	Mandatory	1 minimum	From: *Visual* *Audio* *Audio+Visual* *Abstract*	Define in application; a value of "unknown" may be allowable in certain APs.

[1] The kernel makes a distinction between tangible and intangible manifestations quite deliberately. Intangible manifestations are commonly, but not invariably, digital files. This distinction is made in the kernel arbitrarily, on pragmatic grounds, since these two different classes of manifestation have such different characteristics and behaviour.

Element	Definition	Status	Number	Allowed values	Possible AP qualifications
Primary agent	The name or identifier of the primary agent(s) (normally but not necessarily the creator).	Mandatory	All primary agents. (1 minimum, but all entities fulfilling the same agent role must be included.)	Identifier or Name from an agreed namespace	The specification of the Primary Agent for any AP is determined by the DOI-AP rules.
Agent role	The role(s) played by the primary agent(s)	Mandatory	1 minimum	Role code from an agreed namespace	

In addition to the minimal kernel metadata elements (and any extensions for a specific DOI AP), administrative data such as Registrant, date of registration, record version number, is also mandatory.

The DOI kernel metadata is publicly accessible, in the sense that any user with a DOI can freely access the kernel metadata that relates to that DOI. The opposite is not necessarily true – the "reverse look up" from metadata to DOI may be a commercial service, which will be offered in some circumstances to support the cost of registration (see also Section 5.6 and Chapter 11).

5.5 The DOI Application Profile

What is a DOI Application Profile (DOI-AP)? It is defined as "*the functional specification of an application (or set of applications) of the DOI System to a class of intellectual property entities that share a common set of attributes*".

A DOI-AP is created to serve the requirements of a particular "community of interest" within the broader DOI Registrant community. When the need for a particular AP is identified, the IDF is pleased to assist in the establishment of a DOI User Community (DOI-UC), the organization that defines and implements the DOI-AP. The definition of a DOI-AP involves specifying not only the metadata to be declared alongside a DOI for an entity within that DOI-AP, but also the procedural and commercial rules relating to the exploitation of that metadata and (by inference at least) the business model for the Registration Agencies involved.

The establishment of a DOI-AP normally requires the support of at least one Registration Agency, (see Chapter 11) although multiple Registration Agencies may often be involved.

All entities identified with a DOI must be assigned by their Registrant to at least one DOI-AP. They may be assigned to more than one AP, in which case the registrant must follow the rules laid down by each DOI-AP to which the entity is assigned. The question of which DOI-AP or APs any particular entity should be assigned to is a matter for the discretion of the Registrant.

More information about the establishment and management of DOI –APs can be found in Appendixes 2 and 3.

5.6 DOI kernel metadata and rights metadata

DOI kernel metadata does not include any metadata specifically to describe or ascribe any intellectual property rights in the entity that is identified. This is deliberate. The data in the kernel is intended to be persistent and static, whereas rights data is transient and dynamic.

Nevertheless, DOI metadata provides a firm foundation for interoperation with well-formed rights management data. We are certain that future DOI-AP metadata schemes will include

rights metadata, once the development of specific applications provide a reason for its inclusion.

5.7 DOI metadata usage and policy

The DOI system requires the declaration of metadata solely in order to permit unambiguous identification of the entity the DOI is assigned to. A minimum kernel of metadata will be declared for all DOIs registered and must be publicly available, so that a basic description of the entity that DOI identifies can be accessed (that is to permit look-up from a DOI to the declared metadata) by any user. Extensions to metadata beyond the kernel will not necessarily be publicly accessible, dependent on the application profile, commercial considerations, and registration agency agreements.

Reverse look-up (from metadata to a DOI) is *not* a function of the DOI system itself. Reverse look-up may be offered by other services as a value-added feature. Individual applications or registration agency services will offer this service by agreement with their registrants and suppliers on commercial terms; this will not be determined by IDF. In many areas of intellectual property, extended metadata and reverse look-up via sophisticated searching techniques is an important business activity. As a matter of policy, the IDF will not consolidate DOI state data or kernel metadata for resale or re-use. This data is held by IDF solely for the purposes of permitting look-up from a DOI to the declared metadata by any user.

6 Resolution

This chapter describes the function of Handle System technology in making DOIs "actionable" – through "resolution". It explains what resolution is and discusses the reasons for the choice of Handle technology.

6.1 What is resolution?

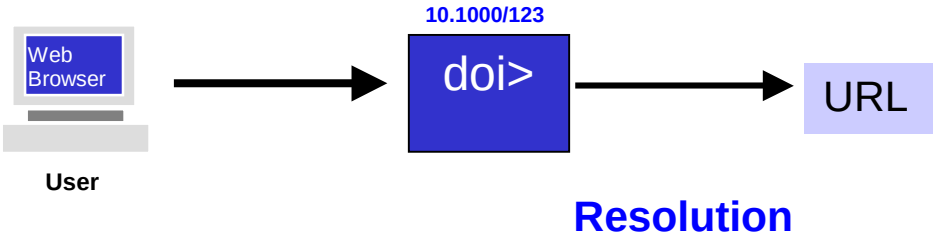
Resolution in this context means the process of submitting an identifier to a network service and receiving in return one or more pieces of current information related to the identifier. In the case of the Domain Name System (DNS), as an example, the resolution from domain name, e.g. www.doi.org, is to a single IP address, e.g. 132.151.1.146, which is then used to communicate with that Internet host.

6.2 Simple resolution

The lack of persistence in identification of entities on the Internet is a commonplace. Even the most inexperienced of users of the World Wide Web rapidly becomes familiar with the "Error 404" message that means that a specified Web address cannot be found – the URL for that web page cannot be resolved.

A DOI persistently identifies a specific intellectual property entity, which may or may not be an Internet-accessible file. The URL identifies a specific address on the Internet. These applications of identification are completely different. One identifies an entity; the other identifies a location (where a specific entity may or may not be found). The analogy is with the ISBN (which identifies the book) and the shelf-mark (which identifies the place where the book is to be found). When the location changes, the shelf mark changes – but the ISBN does not.

The earliest application of the DOI was for simple, single point resolution. Each DOI had a single URL to which is could resolve. This allows the *location* of an entity to be changed while maintaining the *name* of the entity as an actionable identifier.



The DOI is not alone in providing a solution to this problem. Certainly, other applications, for example PURLs (or Persistent URLs), can provide this simple level of resolution. It has been argued that URLs can (in theory) themselves be used as a persistent identifier – that their use as a transient identifier is a social, not a technological, problem.

However, this lack of persistence of the URL is only the first – and the simplest – challenge that the DOI System was designed to manage.

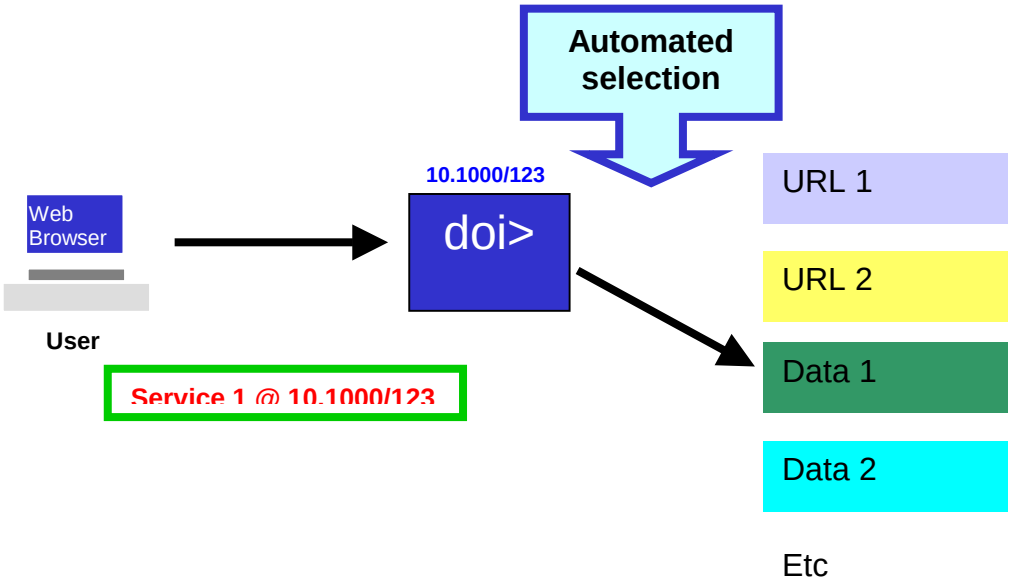
6.3 Multiple resolution

A DOI is a *name* for an entity; in the network environment, there may be many identical copies ("instances") of the same piece of content. How does a single DOI manage the

existence of multiple “instances”, where an individual user may be authorized (or may prefer) to access one instance but not another of the same content? This is sometimes known as the “appropriate copy problem” and the DOI System has the potential to solve it.

The DOI System is designed to manage much more complex services than are exemplified in the problem of discriminating between multiple instances of the same piece of content. One simple example has already been mentioned – how does a user move from a DOI to access metadata about the entity that the DOI identifies (see Chapter 5)? The DOI System in its entirety – enumeration, description, resolution, policy – provides a robust platform that is ideally suited for the many complex services that will be required for the management of intellectual property in the network environment, including digital rights management applications.

The solution to these challenges lies in automated “multiple resolution”. A DOI can be resolved to an arbitrary number of different points on the Internet: multiple URLs, other DOIs, other data types. If the DOI can point to many different possible “resolutions”, how is the choice made between different options? At its simplest, the user may be provided with a list



from which to make a manual choice. However, this is not a scalable solution for an increasingly complex and automated environment. The DOI will increasingly depend on automation of “service requests”, through which users (and, more importantly, users’ application software) can be passed seamlessly from a DOI to the specific service that they require.

The multiple resolution capability of the DOI, using Handle System technology, provides a platform on which applications of great complexity and sophistication can be built.

6.4 Handle System technology

Handle System technology, developed by CNRI, was selected for the resolution task within the DOI System because it offered a number of real advantages over other available technologies:

- Multiple resolution capability
- Scalability
- Reliability
- Resolution speed
- Proven usage in several digital library projects

- Already implemented and supported in several practical systems
- A commitment by its developers to open standards, and
- A commitment to further development

A detailed specification of Handle System technology can be found Appendix 5 and discussion of recent developments can be found in Chapter 8.

The DOI is one implementation of the Handle System; DOIs are a subtype of “handle”, but not the only one. DOIs are distinguished from other handles by the totality of the DOI System described in this Handbook.

6.5 The resolution interface with Handle System technology

Current Web browser technology requires additional functionality to allow the browser to deal with names of objects, rather than simple locations (a fact common to any approach to naming on the Web). Hence, in order to make full use of DOI resolution functionality, additional browser features are necessary. It is anticipated that features supporting resolution will commonly be built into browsers in future, and the IDF is in active discussion to encourage this. The required functionality is currently provided in a number of ways.

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. It will recognize a DOI in the form “doi:10.1000/123”, and resolve it to a URL or other file type the browser recognizes. The user simply “clicks” on the DOI (or types the DOI into the address line in their browser) and the DOI is resolved directly.

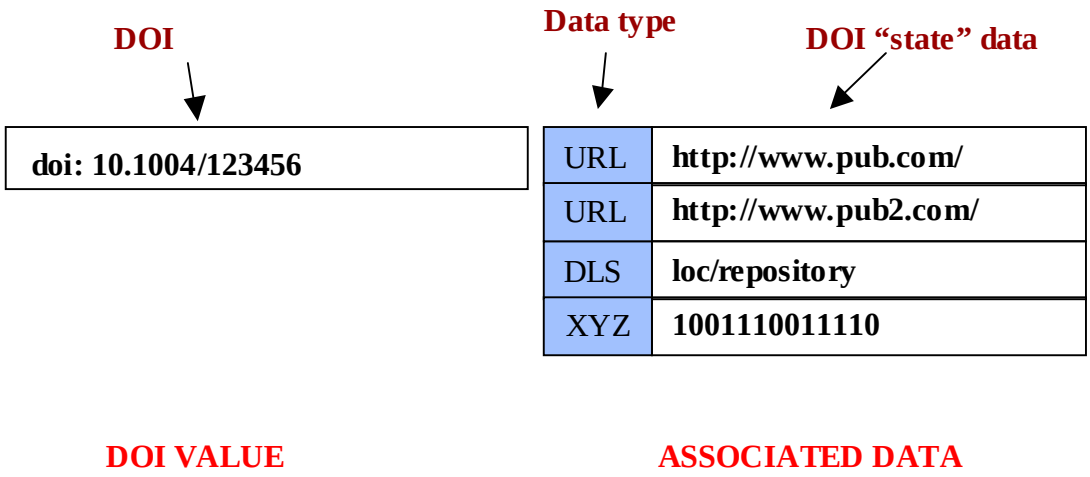
Alternatively, without the need to extend their web browsers' capability, users may resolve DOIs that are structured to use the DOI proxy server (http://dx.doi.org). The resolution of the DOI in this case depends on the use of URL syntax: the example DOI we have been using (doi:10.1000/123) would be resolved from the address: “http://dx.doi.org/10.1000/123”. Any standard browser encountering a DOI in this form will be able to resolve it.

The use of the proxy server and an unextended browser provides the more common user interface to the DOI today. However, it has significant disadvantages when compared with the resolver plug in. The disadvantages include both performance and functionality. Inevitably, direct resolution will often be quicker than resolution using a proxy server. Furthermore, the development of additional services which depend on utilizing the full multiple resolution potential of the DOI (and the Handle System technology) will necessitate the user being able to manage DOI resolution directly.

It is also possible to conceive of the required functionality being delivered to a browser by means of a scripting feature, such as JavaScript. However, to date we have not encouraged this as a key component of any long range DOI/Handle strategy. Reliance on scripting is unlikely to be assured of support by browsers in the medium to long term; for example, many security specialists are currently urging computer users to turn off JavaScript in their e-mail system preferences.

6.6 The maintenance of DOI “state data”

The effective operation of the DOI System depends on accurate resolution of a DOI to the appropriate URL or other data type.



The maintenance of the "state data" is an essential element of the responsibility of the Registrant of the DOI. Currently, only the Registrant or a service organization acting with the authority of the Registrant is permitted to maintain state data. More sophisticated models of permissions and access to DOI state data records within a DOI record are conceivable and the requirements for these are currently being investigated by the IDF.

The data types to which a DOI can resolve are fully extensible within the Handle System, to permit the DOI to resolve to any data that is accessible on the Internet.

6.7 The development of services

The development of services that make use of the potential of the DOI and multiple resolution will be the responsibility of commercial organizations that can identify appropriate business opportunities. We would anticipate that this development is likely to involve both suppliers of technology (including Registration Agencies – see Chapter 11) and groupings of registrant organizations that recognize a common need.

The IDF is keen to encourage the early implementation of many services that fully utilize the DOI and Handle System technology; it sees itself primarily as a catalyst, bringing together organizations that may have a common interest and actively championing and facilitating the development of useful applications.

7 Policies

An essential element of the work of the IDF lies in establishing policy. This chapter describes the process of policy formulation within the International DOI Foundation (IDF), and sets out the more important policies as they stand at present (February 2001).

7.1 The significance of policy

Like any other system of equivalent complexity, the DOI System requires rules for its management. These ensure that the system behaves in ways that are predictable and consistent. These rules are to some extent implicit in the technology and its implementation; but it is important for Registrants, Registration Agencies and users of the DOI system that its rules should be explicitly stated.

The formulation of policy is in many ways more complex than the management and development of the technology. The way in which the DOI is implemented could have significant impact on the way in which the intellectual-property based businesses that use the DOI will operate in the network environment. Policies with respect to metadata access and exploitation, for example, can have a considerable influence on the business models of Registration Agencies (and therefore on how the costs of managing the DOI System as a whole are to be borne).

The IDF is a young organization. However, a very significant framework of policy already exists; it is the detail of policy that is continually being developed to meet the requirements of implementation.

7.2 Policy formulation

The IDF ensures that its members are fully involved in all aspects of policy formulation and have the opportunity democratically to affect its outcome.

Policy is ultimately the responsibility of the members of the Board of the IDF (see Chapter 12 for a more detailed account of the Governance structure). The secretariat of the IDF is headed by a full time Director who is responsible (among his other duties) for producing draft strategy documents and for establishing processes for consultation on matters of policy with all members of the IDF and other interested parties.

Strategy papers (that prefigure the development of policy) are produced as required and circulated to members for comment. In response to member comments, a further draft is prepared reflecting the consensus view of the membership and is published widely, for comments from other interested parties. A final draft is presented to the Board of the IDF for approval and forms the basis for official Board policy.

7.3 Current policies of the IDF

The following top-level policies have already been agreed by the IDF:

- A DOI can be used to identify any intellectual property entity. Our definition of intellectual property is a broad one, following that of WIPO agreements and similar international instruments: this includes both physical and digital manifestations, performances and abstract works. An entity can be identified at any arbitrary level of granularity.
- The primary focus of the DOI is on the management of intellectual property entities, but this does not preclude (for example) issuing a DOI to an entity that is in the public domain.
- The use of the DOI system for resolution of a DOI is free to users; the costs of operation of the system should be borne directly or indirectly by the Registrants. The IDF will

provide support for the costs of the system until such time as Registrant fees alone can provide this.

- All DOIs must be registered with the global DOI registry. Registrants are responsible for the maintenance of state data and metadata relating to DOIs that they have registered (see Chapters 5 and 11).
- The syntax of the DOI follows a standardised syntax (see Appendix 1 for guidance).
- In use, the DOI is an opaque string (dumb number).
- Registration Agencies will be established to manage the assignment of DOIs, their registration and the declaration of the metadata associated with them. The business model adopted by an individual Registration Agency is a matter for the Agency alone, so long as it complies with overall IDF policy.
- Each entity registered for a DOI will be assigned by its Registrant to at least one DOI Application Profile; the rules relating to metadata declaration for that DOI-AP will be followed.
- A minimum kernel of metadata will be declared for all DOIs registered. The kernel metadata, the minimum required to permit basic recognition of the entity to which the DOI is assigned, must be publicly available, so that a basic description of the entity that DOI identifies can be accessed by any user (that is to permit look-up *from* a DOI *to* the declared metadata).
- Reverse look-up (from metadata to a DOI) is not a function of the DOI system itself. Reverse look-up may be offered by other services as a value-added feature. Individual applications or registration agency services will offer this service by agreement with their registrants and suppliers on commercial terms, not determined by IDF.
- Extensions to metadata beyond the kernel will not necessarily be publicly accessible, dependent on the DOI-AP.
- IDF will not consolidate DOI state data or kernel metadata for resale or re-use. This data is held by IDF solely for the purposes of permitting look-up from a DOI to the declared metadata by any user.
- DOI data deposited with a Registration Agency will be held in escrow under contractual terms between the Registration Agency and IDF; that is, the data will be available to the IDF in the event of cessation of the Registration Agency.
- Usage statistics and information about individual DOI resolution will not be released by IDF to any party. IDF will only release statistics relating to the aggregate activity of the DOI system.

7.4 Current priorities for policy development

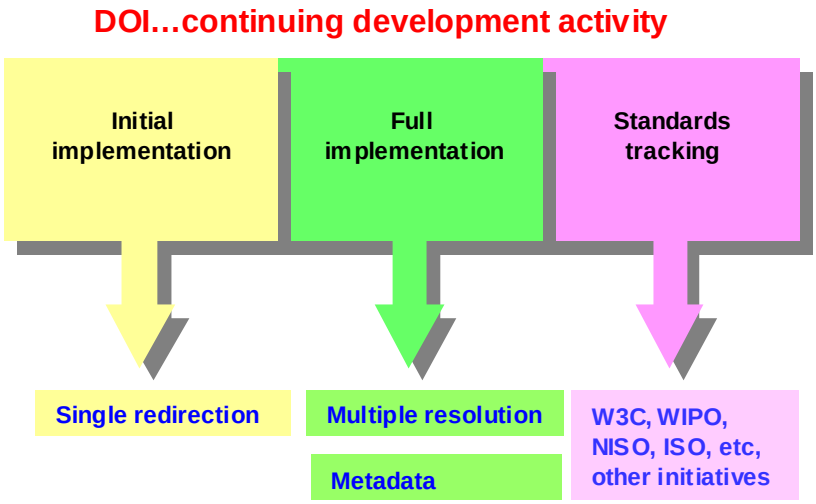
The most urgent current (February 2001) work on policy is in the development of more detailed policies covering the relationship between the IDF and Registration Agencies; some of the issues involved are discussed in Chapter 11. This also involves the development of policies relating to the establishment of DOI Application Profiles (see Chapter 5 and Appendix 2).

8 Development

This chapter looks briefly at the history of the development of the DOI and then describes current development efforts and where these are expected to lead.

8.1 The three-track approach to DOI development

The development of the DOI can be described as following a three-track approach:



The development of the initial implementation – the resolution of a DOI to a single URL, or "single redirection" – was the urgent first task. This was completed and demonstrated in 1997.

However, even as the initial implementation was being developed, work was in hand to develop the other strands of the DOI activity: the full implementation that is currently being deployed (allowing multiple resolution and mandating the declaration of metadata); and the close liaison with standards-making organizations and with other initiatives with adjacent interests.

8.2 Initial implementation

The initial implementation of the DOI was an essential, but limited, first step. Well over three million DOIs have already been registered and are in regular use, permitting simple resolution to a single URL.

However, this first step was always recognized as just that – a first step. As the full implementation of the DOI System is completed, the limitations of the initial implementation will become increasingly apparent.

This does not mean that all the DOIs already registered will be lost. At the appropriate time, Registrants will have the option to migrate them into the "Zero Application Profile" (see Appendix 2) or into another appropriate DOI-AP. A DOI in the Zero AP will have no associated metadata, and will therefore have very limited functionality.

There is no further development to be made of the initial implementation.

8.3 Full implementation

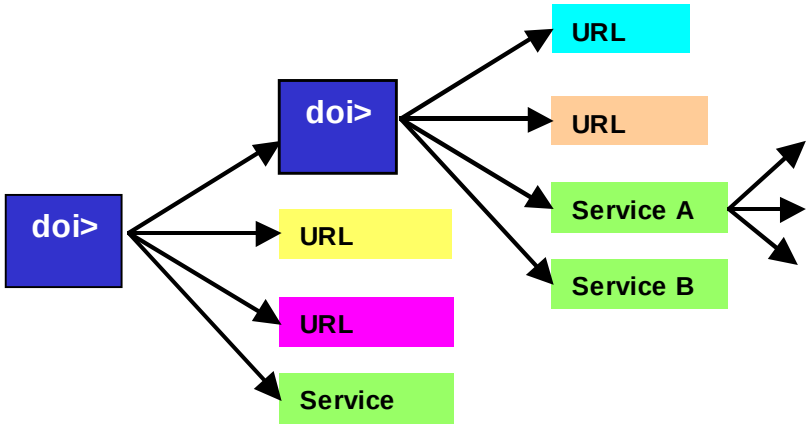
Work is currently (February 2001) continuing to develop the full implementation of the DOI System. This allows for a single DOI to be resolved in many different ways (see Chapter 6) and for complex automated services to be built on the platform of the DOI System. This

necessitates the mandatory declaration of metadata describing the entity that each DOI identifies (see Chapter 5).

The principle of multiple resolution of a single DOI using Handle System technology has already been demonstrated. At present, this simply presents the user with a pop-up window offering the alternative resolutions – but establishes the necessary platform for the development of automated services.

The IDF is in discussion with Registration Agencies and potential Registration Agencies about establishing the necessary infrastructure to provide for the declaration of metadata alongside the DOI, and the establishment of the simple service linking each DOI to its kernel metadata. This work includes development of mechanisms for metadata management. These are likely to involve the implementation of XML-based tools for the initial declaration and subsequent maintenance of both kernel metadata and AP-specific metadata (see Chapter 5).

This will allow for the development of services that reflect the complexity of intellectual property and intellectual property rights.



Such services will not appear instantly, but will be developed commercially in response to real market demand. The IDF does not believe that the development of services is directly its responsibility in the long term, but the IDF will work closely with Registration Agencies, Registrants and others to encourage the development and deployment of useful services (and to facilitate any technical developments within the DOI System as a whole, including development of Handle technology, that are required to permit the implementation of new services).

8.4 Handle System development

In the spring of 2000, CNRI released the second major version of the Handle System. Feedback from existing users, including the IDF, as well as a detailed study of recent developments in related network protocols, resulted in the rewrite being a significant evolution of the system.

Complete technical specifications and other details can be found at the Handle System web site at http://www.handle.net. These public specifications are intended to be sufficient for the creation of both client and server software, enabling, for example, any interested organization to build client software utilizing DOIs and associated state data without reference to or permission from either CNRI or the IDF.

Much of CNRI's recent Handle System development work, reflected in the new release, looks beyond the basic resolution facilities (rapid, scalable, and reliable resolution of "handles" to multiple, typed values representing current state data) to improvements in administration and security. The Handle System protocol, which formerly supported only resolution, now supports both resolution and administration. This means that administrative as well as resolution clients can be built in accordance with the public specification and can be combined into a single client if desired.

This protocol simplification is also related to a new “trust model” for the Handle System, based on public/private key encryption. Each operational handle server now has its own public/private key pair that can be used to authenticate server-to-server transactions. Additionally, each handle has one or more defined “handle administrators”. Administrators are themselves identified by handles, and each administrator may have their own public/private key pair for authenticating administrative permissions.

These two new pieces of Handle System infrastructure form the basis, among other things, for a completely distributed administrative system. Any individual or process, with valid credentials, can administer handles on the relevant handle service without the need for previously established login accounts or passwords. An additional feature of the new system is the ability to include a reference in a handle value, again using public key technology, to authenticate that value with a third party, a type of seal of approval. This is usefully distinguished from authentication of the server *(“I trust that this value came from a given handle server”*) and from the administration of that server *(“in addition to trusting that the value came from a given server I trust that the server's administrative facilities are not compromised”*); it is a way to authenticate the information itself (*“in addition to trusting the complete delivery mechanism, I trust that the data is true”*).

While its exact relation to the future of the DOI is not clear, this new trust model positions the Handle System as a reasonable candidate for an improved PKI (public key infrastructure), the existence of which seems vital to the evolution of managing intellectual property on distributed communication networks.

CNRI continues to be dedicated to the development and evolution of the Handle System as a useful part of the overall architecture for the management of digital objects on the Internet. This evolution will continue to be driven by feedback gained from the growing use of the Handle System across a variety of domains, including the DOI constituency.

8.5 Standards tracking

The DOI does not exist in a developmental vacuum. It is simply one component of a fast developing technological infrastructure for the management of intellectual property in the network environment. There are many different players involved in the development of that infrastructure, ranging from technical organizations to the “content industries” themselves. They include (in no particular order):

- WIPO (the World Intellectual Property Organization)
- ISO (the International Standards Organization)
- NISO (the National Information Standards Organization)
- IETF (the Internet Engineering Task Force)
- W3C (the World Wide Web Consortium)
- OEBF (the Open eBook Forum)
- MPEG 21

A significant element of the work of the IDF lies in tracking standards developments in related areas, understanding their significance to the context within which the DOI will operate, and establishing working relationships with the responsible organizations and projects to ensure that appropriate co-operation is fostered to mutual benefit (and that parallel developments do not remain in ignorance of one another).

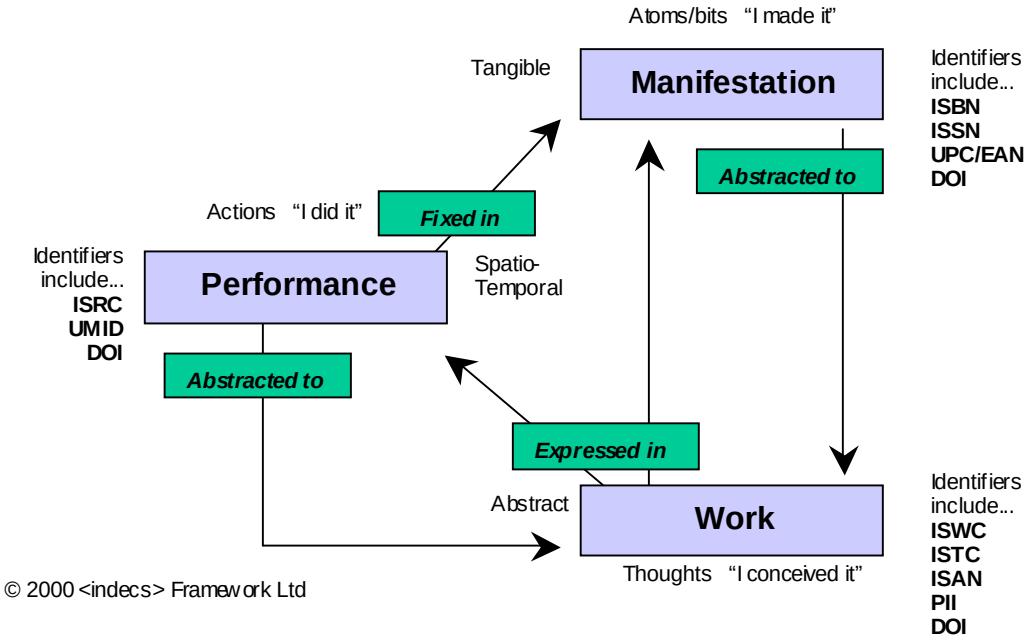
9 Application

This chapter sets out some guidelines for applying the DOI. What types of entities can a DOI be used to identify? At what level of "granularity" should it be applied? What impact does the DOI Application Profile have on the application of the DOI? Who can apply a DOI? It also describes the actual and potential examples of the DOI in action.

9.1 What entities can a DOI be used to identify

A DOI is a general purpose numbering scheme for any type of intellectual property, in any medium. It covers not only manifestations of intellectual property (books, recordings, electronic files) but also performances and "abstractions" – the underlying concepts (often referred to as "works") that underlie all intellectual property.

These "abstractions" are what enable us to recognize a performance of a song, or the words of a book, entirely separately from any particular performance or specific edition. The relationships between these different classifications of intellectual property are summarized in the following diagram developed by the <indecs> project:



The relationships implicit in this diagram can be fully supported by the DOI System, since a DOI can be resolved to another DOI (see Chapter 6). The DOI for an abstract work can (using the DOI System) be resolved to DOIs for many different manifestations of the same work (for example, many different digital manifestations in different formats). In the same way, derivative versions can all be linked back to the original.

The scope of the DOI is deliberately restricted to intellectual property – a DOI should not be applied to anything that falls outside this (already very broad) scope. However, the application of a DOI implies nothing about the status of the *rights* involved in the entity – in particular, there is no claim implicit in the registration of a DOI that the Registrant owns or controls the copyright. For example, a DOI can legitimately be applied to intellectual property that is in the public domain.

9.2 Functional granularity

The DOI can be applied at any level of granularity; in other words, there is no preset definition of the size or form of an entity that may be identified with a DOI. Rather the decision as to what a DOI identifies is taken by the Registrant on a purely functional basis – what is it that I need to be able to identify?

This is an application of what the <indecs> analysis calls *Functional Granularity.* The principle of functional granularity proposes that *"it should be possible to identify an entity whenever it needs to be distinguished".*

A DOI can equally be used to identify a complete opera, an individual aria or a single bar of music. In the same way, it can be used to identify a journal, an individual issue of a journal, an individual paper in the journal, or a single table in that paper. However, it is not always possible to identify in advance which specific elements will need to be identified. It has to be possible to identify only those elements where there is a recognized need to do so – whenever that need is recognized.

9.3 The impact of the DOI Application Profile

Every entity that is registered for a DOI must be declared as belonging to at least one DOI Application Profile. The choice of *which* DOI-AP or DOI-APs is entirely a matter for the Registrant; again, the decision will be made on the basis of application functionality.

The differences between different DOI-APs are defined in part by differences in the metadata that must be declared for each; although the metadata scheme for every DOI-AP must incorporate the kernel metadata (see Chapter 5), most metadata schemes will mandate significant extensions to or elaborations of the kernel metadata that are appropriate to the management of the specific type of intellectual property to which they apply.

The metadata set for a DOI-AP is designed to provide the data elements necessary to support a specific application or set of applications. The selection of the data elements that are required is a matter for the specific DOI User Community that establishes the DOI-AP (see Appendix 2). Different data elements (and different sets of values for those elements) are likely to be required to describe a bar of music from those required to describe a table in a journal article. In exactly the same way, the metadata for that table in a journal article will be different from the metadata required for the journal article from which it comes and the metadata required to describe the journal itself. Each will require the establishment and definition of an appropriate DOI-AP. DOI Kernel Metadata, and the adoption of the <indecs> data model for the development of all DOI metadata, ensure that the metadata schemes used for different DOI-APs are capable of being used alongside one another.

An entity that is registered for a DOI may be declared as belonging to more than one DOI - AP; if this is the case, metadata will need conform to the requirements of each DOI-AP.

Decisions about which DOI-AP or APs to adopt for a specific entity will be made on the basis of the applications that a Registrant requires, the availability of DOI-APs able to deliver that functionality, and the ability of the Registrant to declare the necessary metadata. As more complex and sophisticated DOI APs are developed (including, for example, rights management data elements), increasingly sophisticated metadata schemes will be required to support them.

9.4 Who can register a DOI

Currently, the rules relating to the registration of DOIs are relatively loose; there is an implicit assumption that any organization that has acquired a DOI prefix is unlikely to register a DOI for an intellectual property entity in which it does not have a legitimate interest.

It is anticipated that DOI-APs (which will include procedural and commercial rules, as well as metadata structures – see Appendix 2) will ultimately have their own – different – rules relating to the "legitimacy" of Registrants.

10 The role of the Maintenance Agency

This chapter explains the DOI approach to standardization, and the role of the "maintenance agency" for the DOI that IDF fulfils.

10.1 DOI standards – flexibility and extensibility

The DOI System is like any other complex technical and social construct. Attributes of reliability and predictability can only be delivered, particularly in an automated environment, if the DOI System operates in conformance with technical and procedural standards.

However, standards cannot be established once and then forgotten. Particularly in a rapidly moving environment like the Internet, standards need continuous attention to ensure that they meet the real requirements of the market place – otherwise they fall into disrepute and are rapidly discarded.

This leads to a dilemma – how can anyone implement a technology in the face of so much uncertainty and change. The solution adopted by the IDF is to build a flexible and extensible framework of standards; the framework itself can remain unchanging, while specific market-driven developments can be incorporated and managed by extending the framework. For example, new DOI-APs can be readily developed to meet specific community needs (see Chapter 5 and Appendix 2); and new data types can be added, to allow continuing flexibility in resolution (see Chapter 6).

10.2 The IDF as the DOI Maintenance Agency

There is an increasing community of interest in the DOI – Registration Agencies, Registrants, users, and the members of IDF. Each of these groups needs to have a voice in the development of DOI technical and procedural standards, to ensure that they are genuinely market driven. However, there must ultimately be one organization that "holds the ring" and decides what should or should not be developed into a standard for the DOI System, a strong "force of convergence".

This "force of convergence" for the DOI System is provided by the IDF, as Maintenance Agency both for those aspects of the DOI that are put through external standardization procedures (the DOI Syntax, for example – see Appendix 1) and for those aspects of the DOI System that are considered more appropriate for purely internal standardization.

It fulfils slightly different roles in relation to external standards and internal standards. With respect to external standards, the role of the IDF as Maintenance Agency is laid down by the regulations of the external standards body. For internal standards, the IDF acts as final arbiter.

11 The appointment and role of Registration Agencies

This chapter explains the role of Registration Agencies in the DOI System, their relationship and obligations with the IDF, and the business models that they can adopt to provide the future financial support for the DOI System.

11.1 The role of DOI Registration Agencies

The primary (and minimum) role of Registration Agencies is to provide services to Registrants – allocating DOI prefixes, registering DOIs and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data (see Chapter 5).

This service is expected to encompass quality assurance measures, so that the integrity of the DOI System as a whole is maintained at the highest possible level (delivering reliable and consistent results to users). This includes ensuring that state data is accurate and up-to-date and that metadata is consistent and complies with both DOI Kernel and appropriate DOI Application Profile standards. All Registration Agencies will be expected to support registration of at least one DOI AP (as well as the Base AP).

The Registration Agencies will provide adequate security to ensure that only the Registrant (or someone acting with the Registrant's permission) is able to maintain both metadata and state data.

Finally Registration Agencies are expected actively to promote the widespread adoption of the DOI, and to co-operate with the IDF in the development of the DOI System as a whole.

Registration Agencies *may* choose to provide other DOI-related services to Registrants, without limitation (so long as they conform with IDF Policy). These services may include any combination of value added services in, for example, data, content or rights management. Registration Agencies may also develop services that exploit the metadata that they collect (although Kernel metadata must be publicly declared and freely available – see Chapter 5) Registration Agencies may (but are not obliged to) establish their own local Handle System (see Chapter 6 and Appendix 5).

11.2 Business model for Registration Agencies

Registration Agencies must comply with the policies and technical standards established by the IDF, but are then free to develop their own business model for running their businesses. Unlike the IDF, Registration Agencies may be run “for profit”. There is no appropriate “one size fits all” model and we anticipate that the following business models may involve:

- Direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, with volume discounts, usage discounts, stepped charges, or any mix of these;
- Indirect charging through cross subsidy of the basic registration functions from related value added services.

Registration agencies will determine whether charges are made for prefix allocation or on another basis. Prefixes obtained directly from the IDF are subject to a fee, currently a one-time payment of $1000 which entitles the registrant to an infinite number of suffixes; there is no annual fee; there is no limitation placed on the number of DOI prefixes that any organization may choose to apply for.

Our aim is to encourage the appropriate use of DOI prefixes via Registration Agencies without undue financial penalty, to encourage the use of multiple prefixes within a single organisation. This may prove administratively convenient, especially in large organisations or at a different level of granularity (for example, prefixes allocated to imprints, record labels, image libraries, magazines, journals) as appropriate to a particular Application Profile. Registration Agencies

are enabled to provide DOI prefixes as part of their overall package of services to Registrants. The charge and process for obtaining a DOI Prefix via the Agency is a matter for individual Registration Agencies

The IDF intends to place minimal constraints on the business models offered by Registration Agencies, and is willing to enter into discussions with any interested parties on the practical implementation of appropriate models. The IDF Board has established guidelines for the negotiation of agreements with Registration Authorities and these can be found in Appendix 4.

Permanent exclusivity of registration rights in either a specific geographic territory or for a particular DOI AP will not be granted. However, the IDF recognizes that there is substantial cost involved in establishing a Registration Agency and may be willing to offer exclusivity on the basis of, for example, a "first right to launch" in a sector in return for a guarantee to launch by a certain date, effectively granting exclusivity for a limited initial period. Registration Agencies are free to offer exclusive *services*, although the IDF intends to play a role in ensuring that standard services are properly co-ordinated.

On the related issue of the most appropriate scope of operation for establishing an RA – geographic or functional – no fixed rule can apply, but the most logical scope will become apparent on a case-by-case basis.

Registration Agencies may choose to provide other DOI-related services to Registrants, without limitation (so long as they conform with IDF Policy). These services may include any combination of value added services in, for example, data, content or rights management. Registration Agencies may also develop services that exploit the metadata that they collect (although Kernel metadata must be publicly declared and freely available – see Chapter 5). Registration Agencies may (but are not obliged to) establish their own local Handle System (see Chapter 6 and Appendix 5).

11. 3 Criteria for becoming a DOI Registration Agency

Applicants for Registration Agency status may be any profit-making or non-profit-making organization that can represent a defined ‘community of interest’ for allocating DOI prefixes to DOI Registrants. DOI Registrants can be any individual or organization that wishes to uniquely identify intellectual property entities using the DOI System.

Registration Agencies will become increasingly closely involved in the Governance of the DOI System. Equally, the IDF will be looking at commercial partners to provide future financial support of the central services on which the DOI System as a whole depends.

The procedures for applying to become a Registration Agency were formalized during 2000, but are still subject to some fine-tuning. Please see the "How to…” section at the beginning of this Handbook for the most recent guidance on making an application.

11.4 Formalizing the relationship between a Registration Agency and the IDF

The long-term aim of the IDF is to migrate from an organization supported entirely by its members, to one in which an operating ‘federation’ of appointed Registration Agencies have an increasing level of control (and an increasing level of responsibility). The initial appointment of Registration Agencies is therefore proceeding on the basis of an initial agreement between the IDF and candidate RAs which recognizes the developing nature of the DOI System and which is therefore a little less formal than we anticipate will be the case for the complete final contract.

The IDF Board will be responsible for considering all applications submitted by candidate Registration Agencies. The following initial conditions currently apply:

1. Each Registration Agency will be required to become a member of the IDF under the Registration Agency category of membership. The membership fee within this category is

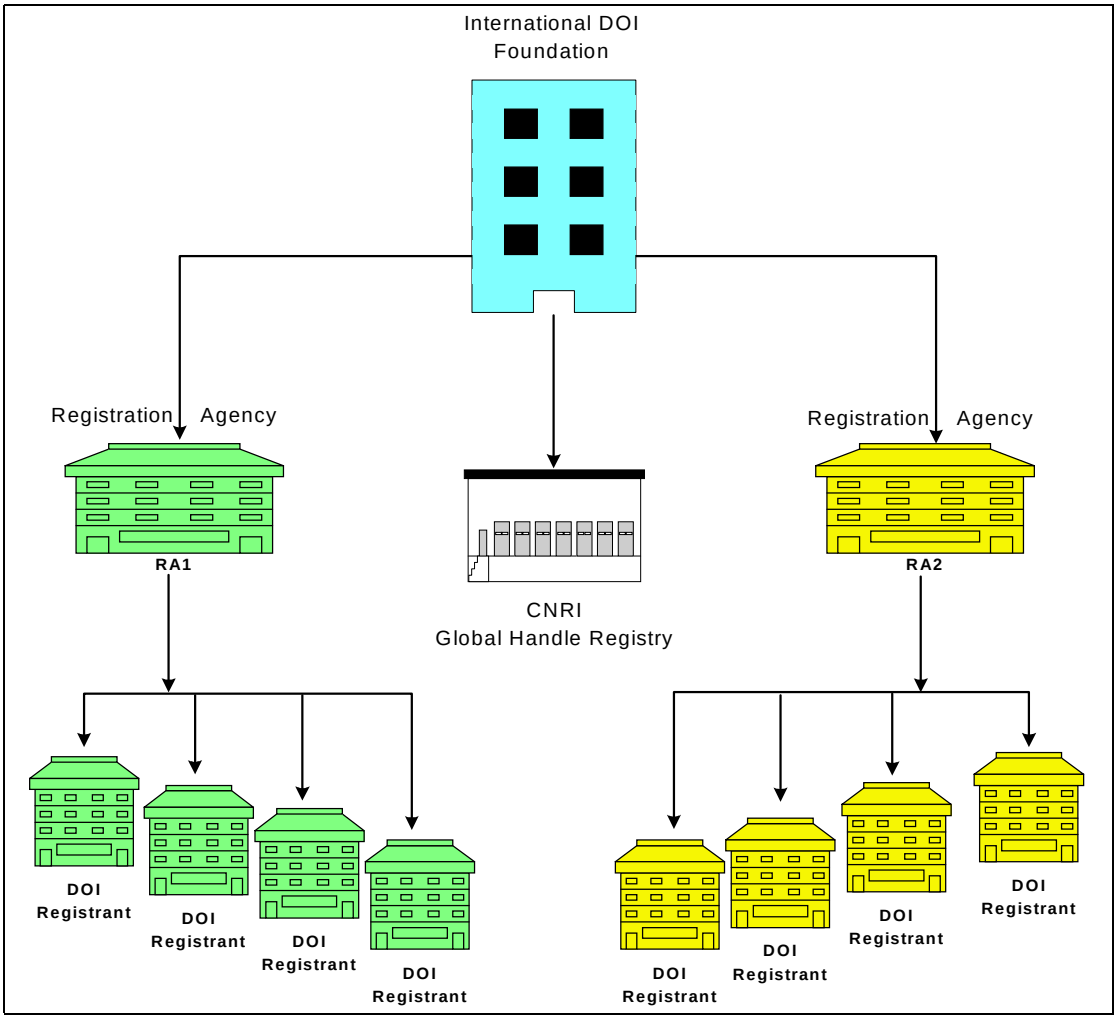
currently $30,000 and this amount must be paid in full before a new Registration Agency begins to operate.

2. A 'Letter of Intent' will be agreed between the Registration Agency and the IDF setting out the basis of the relationship and the terms of operation which have been discussed and agreed (see the "How to…" section of this Handbook)).

The IDF will establish a more formal contractual relationship between itself and all registration agencies in due course. However, in the early adoption period, it believes that a Letter of Intent allows greater flexibility and also provides all active Registration Agencies with the opportunity to become involved in the consultative process leading to consensus on the precise terms to be covered by the formal contract.

11.5 Registration Agencies and IDF Governance and Policy

This diagram illustrates the relationship between the various parties involved in the governance and policies of the DOI System in respect of the operational issues of the DOI System.



The IDF will remain a non-profit making entity, governing the DOI System on a self-financing basis with income derived from RA participation fees. The IDF determines policies and rules concerning the governance of the DOI System and standards for its management. The IDF controls the management of the DOI System through contractual relationships with CNRI for the provision of resolution services, and grants authority to Registration Agencies to become operators of the DOI System for the registration of DOIs.

Each Registration Agency will offer services to DOI Registrants as agreed on an individual basis with the IDF. There is no constraint on the business of a Registration Agency so long as it fulfils its agreed obligations to the IDF and provides an effective service to its associated DOI Registrants and the users of the System.

Each Registration Agency will provide its DOI Registrant community with a mutually acceptable framework for discussing matters of shared concern and interest. DOI Registrants will be contractually related to a Registration Agency. So long as a DOI Registrant complies with the rules and policies of each applicable DOI-AP, there is no restriction on it being a customer of, or in some other way affiliated to, more than one Registration Agency.

Finally Registration Agencies are expected actively to promote the widespread adoption of the DOI, and to co-operate with the IDF in the development of the DOI System as a whole.

Registration Agencies properly expect to become closely involved in the Governance of the DOI System and the IDF will need increasingly to take account of their needs and to provide them with the means to represent their views. To this end, it was agreed in July 2000 by the IDF Board to establish three additional Board seats in a new category, designated Registration Agency member. This first round of three seats is being allocated to RAs (contingent on the signature of a letter of intent which sets out the policies of the IDF) on a first-come first-served basis.

The proportion of RA representative seats on the Board is expected to increase in line with the financial contribution made by RA fees to the IDF's overall resources. Once RA fees generate the majority of the revenue, it is anticipated that the RAs will gain effective control of the IDF.

11.6 Fee structure for Registration Agencies

A fee is paid by RAs to the IDF in recognition of their participation in, and their ability to build a business using, the DOI System. RAs are free to establish their own business model and fee structure with their own customers. The principles of this arrangement are described in a discussion document on DOI Deployment (available from *www.doi.org*)

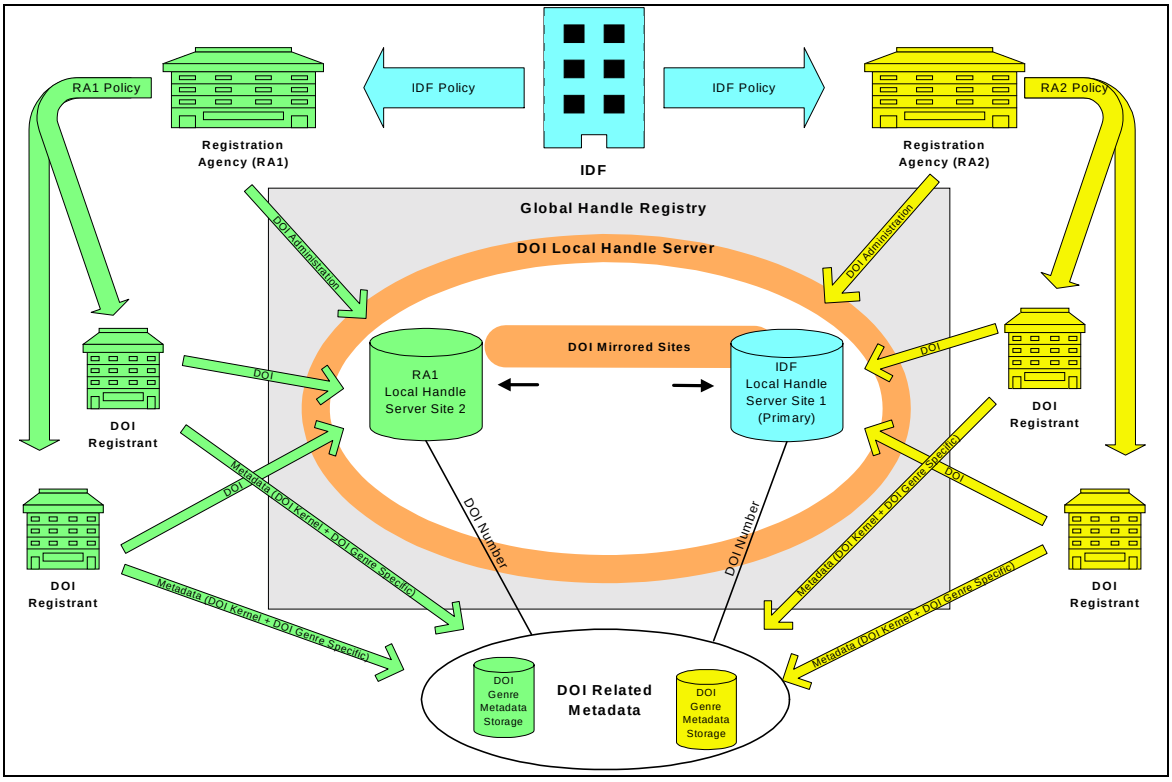
The migration from a membership organization to an operating federation of registration agencies cannot be achieved overnight; our aim is to outline the first steps towards this and to establish the mechanism for an initial start up period. A fee structure appropriate to such a start-up period has been agreed by the IDF Board and will be valid until 31 December 2001. The fee structure after this period will be determined by the IDF Board, a body on which the Registration Agencies are now being properly represented.

The charges for each Registration Agency during this initial period will be as follows:

- Each Registration Agency must be a current member of the IDF, the annual membership fee for which is $30,000. There will be no discretionary reductions in the membership fee for any members who subscribe within the Registration Agency category.
- A franchise fee will be charged. Initially (until December 31, 2001) the franchise fee will be calculated as US 2 cents per DOI registered (equal to $20,000 per million DOIs allocated), with a minimum and maximum as noted below. This will be a one-time fee for each DOI. Accounting will be made on 1.1.2001 and every six months thereafter (i.e. on 1.7.2001 payment is due for all DOIs newly registered in the last six months). Each Registration Agency will guarantee a *minimum* payment of $20,000 (equal to the registration of 1 million DOI's) in their first year of operation, irrespective of the number of DOI's allocated.
- Each Registration Agency will have a *maximum* franchise fee ceiling of $100,000, allowing for up to 10 million DOI's to be registered by a Registration Agency in the accounting period. In a situation where a Registration Agency registers in excess of 10 million DOI's the fee will be renegotiated.

11.7 Operational and Technical Requirements for Registration Agencies

This diagram broadly illustrates the operational relationships between the parties and the interdependencies that exist between them within the context of the DOI System.



In the diagram there are two Registration Agencies (RA1 & RA2), each with responsibility for their own DOI Registrants. For the purpose of illustrating different possible scenarios, RA1 is operating and hosting a mirrored DOI Local Handle Server, whilst RA2 does not host any Handle infrastructure. All the DOI Local Handle Servers are automatically synchronized by the Handle System (see Appendix 5).

The following operational relationships are implied:

- The IDF defines high-level operational policy and assigns the execution of this policy to the Registration Agencies.
- The Registration Agencies enforce their own operational policy, which is specific to their 'community of interest'. This specific policy will incorporate the IDF's high-level policy.
- Each Registration Agency administers the access rights and permissions for the DOI Registrants that form its 'community of interest'.
- DOI Registrants submit DOI's to the Handle system along with the DOI Resolution information.
- At the same time, the Registrant submits DOI related metadata to the appropriate repository for the relevant DOI Application Profile. This metadata incorporates the DOI Kernel information plus the metadata specific to the DOI AP.
- Each Registration Agency is responsible for managing the declaration of the metadata associated with their Registrants.

- The DOI number provides the key to linking the DOI related metadata with the registered DOI Handle.

An authorized Registration Agency issues DOI prefixes to DOI Registrants and requests the resolution system provider to register such new prefixes in the Handle System directories. The RA maintains the systems environment for storing a minimum set of descriptive metadata, as agreed with the IDF that can be integrated with the Handle System. In addition they may implement and operate a Local Handle Server to mirror the services provided by the Global Handle Registry as illustrated by RA1 in the above diagram.

Currently, CNRI implement and maintain the Global Handle Registry on behalf of the IDF. They are responsible for providing Registration Agencies with the necessary software and technical guidance to help them implement Local Handle Servers. As the custodians of the Handle System they are also responsible for the scalability of the system and, in consultation with the IDF, for implementing future developments leading to its growth and any improvement to its technical sophistication.

The service provided by each Registration Agency is expected to encompass quality assurance measures, so that the integrity of the DOI System as a whole is maintained at the highest possible level (delivering reliable and consistent results to users). This includes ensuring that state data is accurate and up-to-date and that metadata is consistent and complies with both DOI Kernel and appropriate DOI AP standards. All Registration Agencies will be expected to support registration of at least one DOI AP (as well as the Base AP).

The Registration Agencies must provide adequate security to ensure that only the Registrant (or someone acting with the Registrant's permission) is able to maintain both metadata and state data.

11.8 Developing the DOI Registration Agency Network

The first DOI Registration Agency, CrossRef, was appointed in 2000. Currently (February 2001), the IDF is in active negotiation with a number of potential DOI RAs. The development of a comprehensive DOI Registration Agency network is the primary IDFpriority for 2001.

12 Governance

This chapter explains the role of the International DOI Foundation in the governance of the DOI System, how the International DOI Foundation is currently funded and how it is anticipated that this will change as the DOI achieves widespread deployment over the next few years.

12.1 The International DOI Foundation

The IDF is a not-for-profit membership organization, established in the United States of America (under the laws of the State of Delaware). It has representative offices in Oxford, UK, Washington, DC and in Geneva, Switzerland.

The IDF was founded in 1998. More than 40 organizations were members of the IDF in February 2001. Membership is open to any organization, commercial or non-commercial. Membership currently includes content owners and publishers, technology companies and information intermediaries – a list of current members can be found at http://www.doi.org/idf-member-list.html and a description of the benefits of membership can be found at http://www.doi.org/membership/brochure.html.

The IDF has a Board of twelve Directors elected by its members. The Board is responsible for all aspects of management of the DOI, particularly policy formulation and standards maintenance. The members of the Board of the IDF are not remunerated for their services to the IDF. Members of the Board represent a wide cross section of organizations interested in the management of intellectual property in the network environment.

The small secretariat of the IDF is headed by a full time Director, Dr Norman Paskin, who represents the IDF in many different forums worldwide and is responsible for the implementation of Board policies and for day-to-day management of all aspects of the affairs of the IDF.

12.2 The funding requirements of the DOI System

The implementation of the DOI System adds value, but necessarily incurs some costs. The three principle areas of cost currently lie in the following tasks:

- Number registration; maintenance of resolution destination(s); declaration of metadata; validation of number syntax and of metadata; liaison with the Handle System registry; customer guidance and outreach; marketing; administration
- Infrastructure: resolution service maintenance, scaling and further development
- Governance: common "rules of the road"; development of the generic system

12.3 Current funding of the DOI System

All the expenses of managing and operating the DOI System are at present (February 2001) being paid by the IDF.

The IDF itself is for the time being supported primarily by the payment of member's subscriptions. This is seen as a temporary expedient to enable the IDF to be funded through its development phase. From the earliest implementation of the DOI, a modest, one-time fee ($1000) has been charged for allocation of a DOI prefix. Unlimited DOIs can be constructed using that prefix. This fee was not intended to provide continuing or sufficient revenue to support the DOI in the long term, but rather to stress that some charging for DOIs was inevitable.

12.4 Future funding of the DOI System

Ultimately, the DOI System will be funded entirely through its commercial application. The IDF proposes that:

- The system should be self-financing; costs for running the system should be recouped from those who benefit from the system.
- The tasks should be delegated to appropriate organizations that can offer appropriate expertise, economies of scale, synergy with existing operations, marketing presence, and similar advantages. In the case of DOI registration, these organizations are referred to as "Registration Agencies" (see Chapter 11)

It is anticipated that, over time, the costs of maintaining the central services will migrate from membership fees to “participation fees” paid (on a basis to be agreed) by the Registration Agencies. If this is the case, the IDF may be in a position to reduce or abolish its membership fees (or to direct this revenue in other directions).

12.5 Future models of governance

A change in funding implies a long-term change in governance structure. The IDF Board recognizes that Registration Agencies have a vital and increasing role to play in the governance of the DOI System. As Registration Agencies are being appointed, they are becoming actively involved in the formulation of policy and the development and maintenance of the standards that underpin the DOI System. Seats on the Board have been established for the representation of RAs.

The process of migration from the current model to the future model is likely to take a number of years, and will involve the development of some hybrid interim models of management. Registration Agencies are already represented on the IDF Board, where 3 of 12 seats are available specifically for Registration Agency members. It is expected that the proportion of RA members on the Board will increase in proportion to the numbers of Registration Agencies (or some related metric).

Appendix 1 ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier

Foreword

(This foreword is not part of the American National Standard Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-2000. It is included for information only.)

This standard defines "the composition and order of components of the DOI℠ (Digital Object Identifier), a character string used to identify intellectual property in the digital environment." The DOI was developed by the International DOI Foundation (http://www.doi.org/) on behalf of the publishing industry, 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. *Note:* DOI and DOI.ORG are registered service marks of the International DOI Foundation, Inc.

Some history on the development of this standard is needed to understand how and why the standard appears in its present form.

DOI System Background

The Internet is a new environment for information transactions and it requires new enabling technologies to provide services and to protect intellectual property. Systems must be developed to identify, authenticate, and protect content to ensure that what the user is requesting is what is being delivered. At the same time, the rights owner of the information must have assurances that copyright in content is respected and protected.

In considering the new systems required, international publishers realized that a first step would be the development of an identification system to be used for intellectual property in the digital environment. Such a system was launched at the Frankfurt Book Fair in October 1997: the Digital Object Identifier (DOI) System. The System provides a unique identification mechanism for content in all media, and a way to link users of the materials to the rights holders or their agents to facilitate automated digital commerce.

The DOI in Context (DOI System, Handle System®, IDF)

The DOI System is an implementation of the Handle System®, developed by the Corporation for National Research Initiatives (CNRI). The Handle System® is a distributed computer system that stores names, or handles, of digital items. It can quickly resolve those names into the information necessary to locate and access the items. It was designed by CNRI as 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 Library of Congress, the Defense Technical Information Center, the International DOI Foundation, and the National Music Publishers' Association are implementing 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 has issued a document outlining the DOI issues entitled "Guidelines for the Issuance and Use of DOI" (the most current version will be available at http://dx.doi.org/10.1000/25).

Basis of the DOI System

In the CNRI Handle System® the term "DOI" is used instead of "Handle" to describe the identifiers. Handle and therefore DOI is in conformance with IETF RFC 1737 Functional Requirements for Uniform Resource Names. (http://ds.internic.net/rfc/rfc1737.txt)

The Handle System web site (www.handle.net) includes an overview of the technology; protocol specifications, and a discussion of Handle System Scalability. The Handle System is an open specification described in two IETF Internet-Drafts: "Handle System Overview," July 1999, and "Handle System Namespace and Service Definition," July 1999.

These documents are in conformance with Section 10 of RFC2026. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt. The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html. An overview of the Handle System is available at http://www.handle.net/overviews/hs-version4.html

Resolution

The DOI System enables resolution of the DOI. A resolution system takes a URN and returns a list of services or instances of the information identified by the URN, commonly one or more URLs. Resolution is here used to mean the act of submitting an identifier to a network service and receiving in return one or more pieces of current information related to the identifier. 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.

DOI and Metadata

From the earliest development of the Digital Object Identifier (DOI), one of the most widely discussed issues has been whether or not the identifier string of which the DOI is composed should be meaningful. The syntax of this string, as defined by this standard, contains a set of components with very limited meaning.

- Each DOI string begins with a code indicating that within the Handle System of the Corporation for National Research Initiatives (CNRI) the string will be resolved by the Local Handle System reserved for the DOI.
- The next component of the DOI string notes the number assigned to the Registrant who originally created the DOI. It should be noted that 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. The Committee recognized that the lack of descriptive information would limit the usefulness of the DOI string. It was agreed that accompanying information that describes the object would be necessary to make the string meaningful. Such information is called "metadata." This metadata may be aggregated into databases together with the DOI string and used for specific purposes.

The following recommendations are accordingly included here:

1. No DOI string should be registered without an accompanying set of metadata describing the object being referenced in the syntax string.

2. The Maintenance Agency listed in Appendix B should provide the latest information about the relevant metadata schemas and any databases that aggregate metadata about DOI referenced objects.

NISO Involvement and Role of the International DOI Foundation (IDF)

The ongoing management of the DOI System is done by the International DOI Foundation, a not-for-profit, membership-based organization with offices in the United States and Geneva, Switzerland. The Foundation is responsible for licensing Directory Managers/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.

Goals of the Syntax Committee

The NISO DOI Syntax standards committee (SC AR) was established with the following goals:

- To formalize the syntax for the DOI identifier string to enable DOI registration.

It has been possible to register DOIs since 1998 but there has been concern that the syntax has not been conclusively set.

- To determine the elements of the DOI identifier string. Several proposals to add elements to the DOI identifier string had been made and the committee was charged with determining which ones should be included in the string.

- To limit the scope of the standard to the DOI identifier string. The DOI System is made up of a number of parts including the identifier string, the resolution mechanism, and the Directory. This standard addresses only the syntax for the identifier string.

While limiting itself to the DOI identifier string, the Syntax committee took into account the wider context of the DOI System, the Handle System®, the International DOI Foundation, and the Internet. This Foreword and the Appendixes provide important information and references for understanding the DOI System, how the DOI is being used, how it relates to other standards and the Internet, and where to get more information.

This standard was processed and approved for submittal to ANSI by the National Information Standards Organization. It was balloted by the NISO Voting Members July 15, 1999 - September 15, 1999. It will next be reviewed in 2005. 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. NISO approval of this standard does not imply that all Voting Members voted for its approval.

* * * *

Standards Committee AR

The following individuals served on the Standards Committee AR, DOI Syntax:

Edward Pentz, Chair
Academic Press

Duane Arenales
National Library of Medicine

Helen Atkins
Institute for Scientific Information

Julia Blixrud
Association of Research Libraries

Kimberly Douglas
Sherman Fairchild Library and
Caltech Library System

Michael Holdsworth
Cambridge University Press

John Mancia
Elsevier Science Inc.

Mark H. Needleman
Data Research Associates, Inc.

Norman Paskin
The International DOI Foundation

Carlen Ruschoff
Georgetown University

Albert Simmonds
albert simmonds development

Acknowledgment

The Committee thanks Larry Lannom of the Corporation for National Research Initiatives and Eric Hellman of Openly Informatics, Inc., for their part in creating this standard.

ANSI/NISO Z39.84-2000 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

Unicode Consortium "The Unicode Standard Version 2.0" ISBN: 0-201-48345-9 (http://www.unicode.org/) - equivalent to UCS-2 as defined in ISO/IEC 10646.

3. Definitions

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.

DOI prefix . The Directory and the Registrant codes issued by the Registration agency to a Registrant for use as the prefix 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. A Registration Agency may also be a Directory Manager.

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 limit on the length of a DOI string, or any of its components.

4.1 DOI Character Set

Legal characters are the legal graphic characters of Unicode 2.0 or greater. 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.

<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 Unicode. <DIR> <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 Unicode. The Suffix cannot start with */ where * is any single character. This is reserved for future use. The DSS is case sensitive.

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 the American National Standard Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-2000. It is included for information only.)

This appendix provides information on aspects of the DOI system beyond syntax 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 (See Section 4.1), Persistence and Character Encoding

<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://ds.internic.net/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 ISO 10646 UCS-2/Unicode 2.0 or greater 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 the American National Standard Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-2000. 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

Web site: www.doi.org

The International DOI Foundation

PO Box 233
Kidlington, Oxford
OX5 1XU, U.K.
Phone: (+44) 1865 843798
Fax: (+44) 1865 843967

* * * *

APPENDIX D Related Standards and References

(This appendix is not part of the American National Standard Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-2000. It is included for information only.)

Standards for item identification have been proliferating in recent years. The standard cited in Section 2 is required for the construction of the DOI syntax. This appendix includes secondary standards and references to standards in development, citations that may be useful with DOIs. Other references provide additional information on the DOI.

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

Secondary Standards

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.

ANSI/NISO Z39.9-1992, International Standard Serial Numbering (ISSN). Bethesda, MD: NISO Press, 1992.

ANSI/NISO Z39.23-1997, Standard Technical Report Number. Bethesda, MD: NISO Press, 1997.

ANSI/NISO Z39.56-1996, Serial Item and Contribution Identifier (SICI). Bethesda, MD: NISO Press, 1996.

ISO 2108:1992 Information and Documentation International Standard Book Numbering (ISBN). Geneva: ISO, 1992.

ISO 3297:1998 Information and Documentation International Standard Serial Numbering (ISSN). Geneva: ISO, 1998.

ISO 3901:1986 Information and Documentation International Standard Recording Code (ISRC). Geneva: ISO, 1986.

ISO 10957:1993 Information and Documentation International Standard Music Number (ISMN). Geneva: ISO, 1993.

Uniform Resource Identifiers http://info.internet.isi.edu:80/in-notes/rfc/files/rfc2396.txt

References to Standards in Development

NISO Book Item and Component Identifier (Draft released for comment January 6, 2000- April 7, 2000)

ISO/TC 46/SC9 ISO/DIS 15706, Information and Documentation International Standard Audiovisual Number (ISAN)

ISO/TC 46/SC9 ISO/DIS 15707, Information and Documentation International Standard Work Code (SIWC) Part 1: Musical Works

Other References

Guidelines for the Issuance and Use of DOI Version 3.1 http://dx.doi.org/10.1000/25

Document Object Model http://www.w3.org/DOM/

Dublin Core Metadata Initiative http://purl.oclc.org/dc/

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

Publisher Item Identifier http://www.elsevier.nl/home/about/pii

"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 the American National Standard Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-2000. 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. This standard uses ISO/IEC 10646 as the basis for the character set. In practice, only the UCS-2 set, which is equivalent to Unicode 2.0, will be used. The default encoding of UCS-2/Unicode 2.0 is that each character consists of 16 bits (2 octets). UTF-8 is a variation of the UCS-2/Unicode 2.0 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 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 practice is to imbed DOIs within URLs for transmission via HTTP to be resolved. The URL currently used for the resolution of DOIs is http://dx.doi.org/. A DOI inside a URL would be:

http://dx.doi.org/10.1006/rwei.1999.0001

DOIs are also primarily used in HTML pages. The DOI *10.1006/rwei.1999".0001* as 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)
}	(%7d)
^	(%5e)
[	(%5b)
]	(%5d)
'	(%6o)
|	(%7c)
\	(%5c)

Appendix 2: DOI Application Profiles: an overview

This Appendix provides an initial introduction to the concepts of the DOI Application Profile (DOI-AP) and the DOI User Community (DOI-UC).

Introduction

The only conceivable reason for anyone to assign a DOI to an IP entity is that they believe that there is an application of the DOI that will provide a service of genuine value to someone, value that can be captured and exploited somewhere in the value chain between creator and consumer of intellectual property. It is the definition of those *applications* with which we are concerned here.

A DOI Application Profile is the functional specification of an application (or set of applications) of the DOI System to a class of intellectual property entities that share a common set of attributes. The DOI-AP is in part defined in terms of a metadata schema that is always a superset of the DOI Kernel metadata schema. However, the definition of a DOI-AP is not limited to the definition of the metadata schema but is further refined by commercial and procedural rules (which can include, for example, rules relating to the exploitation of the metadata that is declared). The definition of the Application Profile includes the Registration Agency or Agencies responsible for its application, and DOI User Community responsible for its management.

The purpose of a DOI-AP is to enable implementation of the application (or set of applications) in a particular environment. Those applications can range from relatively simple discovery to complex e-commerce and rights management applications.

Each DOI-AP has a unique name, issued by the IDF. As an intellectual property entity in its own right, it also has its own DOI, registered by the IDF.

Defining a DOI-AP

A DOI-AP is initially defined in terms of a class of IP entities and an application (or set of applications). A DOI-AP could be defined for any IP entities that share any arbitrary group of attributes, but it is more likely that DOI-APs will, in general, describe classes of intellectual property entity that fall naturally and "intuitively" into a common definition.

The process of definition begins not with an abstract metadata schema but with a functional analysis of the application that the metadata schema and the commercial and procedural rules are designed to support. What service is the DOI-AP intended to provide to its users and what information about the item is necessary in order for this service to be provided? From this functional analysis, it becomes possible to define the metadata required.

To restate this process more briefly, the following four steps are required:

1. To name the intellectual property entity type to which the DOI-AP applies[2]
2. To describe this IP entity type, in terms of its limiting attributes (which will define the value or the constraints on the range of values used in the kernel and the extended metadata)
3. To name and describe the application(s)
4. To define the metadata necessary to enable these applications, and the procedural and business rules necessary to support them

[2] It is conceivable that DOI-APs may be devised in future that involve more than one IP entity type, in which those different entity types play distinctive and different roles. This would seem to imply the development of more than one metadata schema for a single DOI-AP. Nothing in this Handbook should be seen as ruling such a development out of scope.

Thus the CrossRef Application Profile deals specifically with the requirements of linking citations between journal articles[3]. What draws different IP entities together into an Application Profile is that they share a set of common attributes (commonly the **format** of their manifestation, the **mode** of their consumption, their **subject**, and – in the case of journal articles – the **continuity** of their publication).[4] More importantly, what also draws them together is that their registrants recognize that there is a common interest in the development of specific services relating to this class of IP entities; in the particular case of CrossRef, the service is citation linking.[5]

The CrossRef metadata schema does not provide for the generic description of "a journal article" that would be capable of supporting *any* (arbitrary) application that anyone might choose to develop relating to journal articles identified with DOIs. Nevertheless, it is self-evidently advantageous for the orderly future management of DOI-related applications by registrants that the CrossRef metadata schema is compatible with metadata schemas developed for any other future application relating to journal articles;[6] it is likely also to be the case that applications will develop, particularly in rights management, that need to cross boundaries between traditional sectoral classifications of intellectual property.

Both of these future developments mandate a careful adherence to an extensible generic data model such as the one that we have chosen to adopt.

A guide to developing metadata schemas for DOI-APs can be found in Appendix 3.

Registration

No Registrant will be compelled to register any intellectual property entity in a particular DOI Application Profile. All entities *must* belong to at least one DOI-AP (of which the simplest will always be the Base-AP – see below); entities *may* belong to many different DOI-APs, so long as the Registrant complies with the rules associated with each AP.

These rules will always involve the declaration of metadata including (but not necessarily limited to) DOI Kernel Metadata. DOI Kernel Metadata will always be open to public inspection.

Identifying the need for a DOI-AP

Any organization or group of organizations (formal or ad hoc) may identify a "community of interest" in the development of a new DOI-AP. The IDF prefers such groupings always to consist of a combination of more than one potential Registrant organizations with an interest in implementing the DOI-AP, and at least one potential Registration Agency. The Registration Agency should be willing in principle to register DOIs for entities in this Application Profile, and to develop the application(s) to support the requirements of the Registrant organizations. However, this is not an absolute rule and the IDF is willing to discuss the need for a DOI-AP with any interested organization or organizations.

Any organizations contemplating the development of a new DOI-AP are strongly recommended to approach the IDF at the earliest possible opportunity to avoid the duplication of effort which might otherwise result.

---

[3] It should be noted that CrossRef does not identify any specific manifestation of a Journal Article but rather the abstract IP entity (commonly called a "work") that may be expressed in many different manifestations.

[4] For definitions of these terms, see Appendix 3.

[5] There would be little value in the registrant of a DOI for a musical recording, for example, registering the DOI in the CrossRef AP. However, CrossRef does intend to extend its citation linking activities to other aspects of academic publication, and will need to develop its Application Profile accordingly.

[6] This could lead over time to the development of more generic metadata schemas for particular classes of IP entity, like journal articles, that become recognised as, say, "the Journal Article metadata schema". Specific applications, like CrossRef, would utilise only a subset of this more generalised schema.

DOI-APs and Functional Granularity

When is it appropriate to develop a new DOI-AP, and when is it more effective to extend the use of an existing one? This is a question to which the answer lies in the realm of "functional granularity".

The IDF will examine each proposal for the definition for a new DOI-AP with this question in mind. The IDF's primary guideline will be to minimize the number of similar DOI-APs, while maximizing the applications of the DOI that are available to registrants. To the extent that a new DOI-AP facilitates and encourages the development of new applications, the IDF will be inclined to approve its development.

The IDF will encourage dialogue between organizations with similar requirements, to explore the potential for convergence between DOI-APs.

Developing a new DOI-AP

The development of the technical, procedural and commercial rules for a new DOI-AP is the responsibility of the organizations that are seeking to register the DOI-AP; however, it is essential that the development work be done in close co-operation and dialogue with the IDF.

DOI-AP Governance

Each DOI-AP requires a formal mechanism of governance – a DOI User Community or DOI-UC – constituted by and out of the relevant "community of interest", to determine, subject to IDF approval, issues relating to the development of the DOI-AP. This may be related to a "parent" Registration Agency for each DOI-AP (who may, for example, act as chair). It may be a pre-existing group (such as an international trade body). Membership rules and procedures are to be determined by the DOI-UC itself (with IDF approval).

DOI-AP metadata schemas

DOI-AP metadata schemas are developed in accordance with the principles and using the overall structure of the <in*d*ecs> metadata schema. Before commencing any work on a DOI - AP metadata schema, developers are urged to read Appendix 3.

DOI-AP metadata schemas and interoperability

The reason for the IDF's insistence on conformance of DOI-AP metadata schemas with the <in*d*ecs> framework is simple. This mechanism will ensure interoperability between the metadata schemas of different DOI-APs, thus allowing the same entity to be registered in more than one DOI-AP without metadata conflicts.

DOI-AP procedural and commercial rules

Alongside the development of the metadata schema, the DOI-UC will be expected to develop their own rules of procedure for the management of the DOI-AP. These rules are expected to encompass issues such as access to and exploitation of metadata, as well the implementation of applications based on the metadata.

The procedural rules will also cover the central question of who shall be permitted to register a DOI and to manage the associated data.

The role of the IDF in development

Within this broad framework, decisions about the metadata schema are a matter for the DOI-UC that is developing it. The IDF will advise on development, but will not dictate requirements beyond the framework. The IDF will provide guidance and assistance to the extent that this is possible, but reserves the right to charge commercial rates for any consultancy it may provide.

The IDF itself may take the initiative in developing some DOI-APs where it can see that these may have widespread potential application. In these instances, the IDF will establish the appropriate DOI-UC.

Ownership of Intellectual Property in DOI-APs

Any intellectual property that may exist in a DOI-AP becomes the property of the IDF. The IDF will licence the *use* of that DOI-AP to any DOI Registration Agency without charge. However, in recognition of the costs involved in developing a DOI-AP and the applications which the DOI-AP enables, the IDF may be willing to negotiate a period of exclusivity for a Registration Agency, before others are licensed to use any particular DOI-AP.

There is no compulsion on Registration Agencies to licence the use of any applications that they may have developed or to licence their intellectual property in any context other than that of a DOI-AP.

Any other organizations may, with the permission of the IDF, choose to adapt either the metadata schema or the commercial and procedural rules of an existing DOI-AP to use in another DOI-AP.

Approval of DOI-APs

All DOI-APs must be registered with, and approved by, the IDF.

Metadata schemas drawn up for DOI-APs are subject to review by the IDF to ensure conformance with both IDF Kernel Metadata and with <in*d*ecs> principles. The IDF reserves the right to amend any proposed metadata schema to ensure compliance. No DOI-AP metadata schema can be declared as such until it has been approved by the IDF.

Any commercial and procedural rules relating to a DOI-AP are also subject to approval by the IDF to ensure that they are in conformance with IDF policy. The IDF will encourage as wide a variety of business models as possible, within the overall framework of the policy.

The IDF reserves the right to charge for the process of registration and approval of a DOI-AP, but does not anticipate charging for this service in the immediate future. Although the IDF will use its best efforts to ensure that a DOI-AP is appropriate to its purpose, and (when it has been approved) will widely promote its use, the IDF can take no overall responsibility for the effectiveness or the adoption of any particular DOI-AP.

The IDF will not formally register a DOI-AP unless at least one Registration Agency is identified as willing to implement it. Once registered, all DOI-APs are made publicly available by the IDF.

Maintenance of DOI-APs

For the same reasons as approval is required for new DOI-APs, all amendments to registered DOI-APs require the approval of IDF.

Special cases – the Zero-AP and the Base-AP

The Zero-AP

In order to be able to manage an orderly transition from the early implementation of the DOI (which did not involve metadata declaration), all DOIs that were registered without metadata declaration will be migrated to a special **Zero Application Profile**.

This will become a closed AP, available for the sole purpose of managing these early registrations. Once proper metadata declaration facilities are in place, it will not be possible to register a DOI without metadata declaration. Only very limited functionality will be possible with DOIs registered in the Zero-AP. There is no intention further to develop that functionality.

Registrants of DOIs without declared metadata are encouraged to register these DOIs in another DOI-AP (and to follow the rules of that AP with respect to metadata declaration).

The Base-AP

Any DOI can be registered solely in the **Base Application Profile**; the Base-AP requires declaration only of the Kernel Metadata and follows the general rules of the IDF. The Base-AP is available to any Registrant and maybe used in the absence of an alternative AP that is considered suitable (for whatever reason) by the registrant. All Registration Agencies will be expected to offer the Base-AP. Outline rules for the operation of the Base-AP will be established by the IDF, but in principle individual Registration Agencies will be permitted to operate their own business models for DOIs registered in the Base-AP.

The DOI-UC for the Base-AP will be established by the IDF Board.

Appendix 3. Defining metadata schemas for DOI Application Profiles: an outline guide and template

This Appendix is a guide for those seeking to define metadata schemas for new DOI-APs (or to extend the metadata schemas of existing DOI-APs).

1. Introduction

It has long been recognized that, for the DOI system to fulfill its potential in the management of intellectual property in the network environment, the registration and declaration of structured metadata would be essential.[7] For reasons we will discuss, some elements of this structured metadata – the DOI Kernel metadata – is publicly declared and freely available for all intellectual property entities to which a DOI is assigned. However, this metadata supports only the simplest of applications – the discovery, from a DOI, of enough information to be able to recognize what it is that the particular DOI identifies.

We are concerned in this Appendix only with defining metadata schemas for DOI-APs. We do not address questions of the formulation of commercial or procedural rules (except to the extent that they may have an impact on the design of the metadata schema). Such commercial and procedural rules are essential elements in the overall definition of a DOI-AP; but they lie outside our present scope.

Metadata is potentially infinite in scope. Through a chain of relationships, everything is connected (however remotely) to everything else. The *complete* description of anything, in terms of its detailed relationships to everything else, is quite simply an impossibility. What can be relatively simply achieved is a description sufficient to facilitate the operation of a specific application.

This application can be a relatively simple one, like the CrossRef application that uses a straightforward set of "discovery" metadata, to allow for the automated linking of citations in the reference lists of journals.[8] In time, much more complex applications will be developed based on the DOI System, including for example applications in the rights management domain. These will require the development of more complex metadata schemas.

To support such future applications, it is essential to build the principles of interoperability and extensibility into the DOI System from the outset. This implies that all metadata registered in support of any application of the DOI should conform to an overall metadata architecture to minimize the possibility of any barriers to interoperability in the future. No further attempt will be made here to justify either the principle of interoperability or the particular approach taken to the challenge of creating interoperable metadata. Acceptance of those arguments (made extensively by the <indecs> project) is assumed.[9]

Equally, there is no assumption that the metadata schema for a DOI-AP need necessarily be complex or difficult to manage. Indeed, our inclination is precisely in the opposite direction. The schema should embrace only such complexity as is entirely necessary to support the specific application that is proposed. Early implementations of the DOI support relatively simple applications (and thus have relatively simple metadata requirements). However,

---

[7] This document is entirely agnostic on the processes that will be utilised to manage the public declaration of metadata, whether for the "kernel" metadata or the extended metadata for a specific DOI-AP. It is intended that there should be public and automated access to the kernel metadata declaration through the use of a specific Handle data type (see the DOI discussion paper *One to Many* (August 2000) on www.doi.org). Whether this metadata is publicly declared in a central repository or on local websites maintained by individual registrants is a matter that remains to be determined (and may differ between registrants and between different DOI-APs).

[8] See www.crossref.org and also Section 12 of this Appendix for further information about CrossRef and its metadata schema.

[9] See Appendix 6 for an overview of the <indecs> project.

compliance with the structural rules is essential if metadata extensibility and interoperability is to be supported into the future.

2. The purpose of this Appendix

This Appendix is a guide for those seeking to define metadata schemas for new DOI-APs (or to extend the metadata schemas of existing DOI-APs). It is not a metadata primer; it assumes that the reader will have sufficient background in metadata analysis and data modeling to understand the fundamental issues that we will discuss. Nor can it be a complete introduction to the <indecs> metadata model, although DOI-AP metadata depends for its intellectual underpinning on the <indecs> Framework[10] and we cannot avoid covering many of the more significant elements of the <indecs> metadata analysis.[11]

Our intention here is to describe the over-arching structure to which all DOI-AP metadata schemas must adhere; our hope is to provide sufficient guidance to allow a metadata analyst, following this guide, to be able to define a schema that will be approved by the IDF. Nevertheless, we recognize that it will be challenging within many communities to locate the necessary resources to understand and implement some of the more complex metadata structures. The IDF is itself currently developing sufficient resources to enable it to offer any necessary consultancy in the definition of the metadata schemas necessary to support the functional specification for a particular application; these resources are available to assist in the definition of new DOI-APs and appropriate metadata schemas.

Close co-operation with the IDF in the definition of metadata schemas is essential if wasted effort is to be avoided.

3. The DOI System's dependence on declared metadata

Any application that operates on *any* identifier depends on metadata. The identifiers that we use every day – whether ISBNs or telephone numbers – would have little value unless we knew what it was that they identify.

Sometimes that metadata may be entirely proprietary – within the book trade, for example, the data supporting an ISBN in a bookstore's system and the data supporting the same ISBN in a publisher's system can be completely different (and often are, if not always entirely intentionally). The ISBN, a simple identifier, itself appears to be the only item of metadata "publicly" declared by the publisher.

However, it takes not a moment's reflection to recognize that without resources like *Books in Print* (or its equivalent), the ISBN itself would be without value as a tool to facilitate interoperability between different systems.

The IDF recognizes that metadata can have considerable value and that the realization of that value is potentially a matter of great sensitivity. Metadata has value precisely because it provides a tool essential to the delivery of applications and services operating on the identifier; these applications and services are designed to deliver value to users. *Books in Print* is a commercially valuable property, because it provides discovery services to its users – who will therefore pay for it.

The IDF has no wish or intention to restrict the commercial application of metadata. The entire concept of the DOI-AP is *application* driven – the extended metadata schema to support each DOI-AP is designed to support the specific *functional requirements of applications* that are recognized to be of value by the "community of interest" (or DOI User Community: DOI-UC) that establishes the Application Profile.

---

[10] See www.indecs.org for more information on the <indecs> framework; significant elements of this document are reproduced with permission of the <indecs> Framework Ltd.

[11] Additionally it is important to note that we are not here covering issues relating to the technical expression of DOI - AP metadata schemes. Technical standards for the management of DOI-APs are a separate strand of the IDF's current development work.

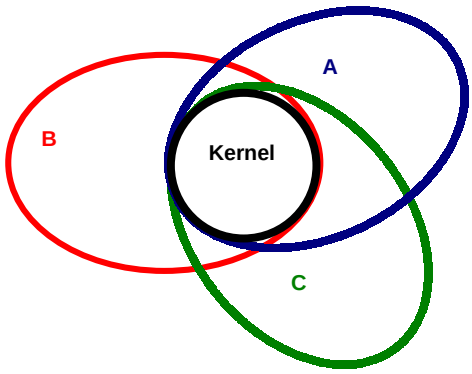
The rules about access to the extended metadata – those elements that go beyond the "DOI Kernel" – are a matter entirely for the DOI-UC itself to establish. In some cases, the DOI-UC may decide to restrict access only to other members; in others, the DOI-UC may decide to sell the metadata (or access to it) to support the costs of establishing and maintaining the application; another community may decide to provide free access to all metadata to anyone who wishes to make use of it in any way. The IDF does not dictate the policies of individual DOI-UCs.

However, one aspect of the IDF's metadata policy is mandatory. Because the DOI can be used effectively to identify *any* Intellectual Property entity, it is essential for a user to have access to some information about what it identifies. This one application is to be provided free to anyone by the IDF, by means of a look up of the DOI kernel metadata.[12] This metadata is available only for these simple look up purposes ("I have a DOI, what does it identify?").

The kernel is designed to provide the necessary information for adequate human *recognition* of the object identified, but not for its complete automated disambiguation from other similar objects. It is not anticipated that any other very useful application could be built using *only* publicly declared kernel metadata. The public record will, though, point users towards the DOI Application Profile (or Application Profiles) within which a particular IP entity is registered (and will therefore provide a starting point for a user who is looking for particular services).

As well as providing this simple but essential user service, the kernel metadata serves a further necessary service to the integrity of the DOI System as a whole, by ensuring that each DOI registered is indeed used to identify (persistently and uniquely) an IP Entity. It helps to avoid any use of the DOI as a simple Internet routing system (for which other solutions are available). The declaration of metadata enforces discipline that might otherwise be lacking in the DOI system as a whole.[13]

All DOI-AP metadata schemas are supersets of the kernel; to put it another way, the kernel metadata elements are the elements that are common to all DOI-AP metadata schemas:



The kernel is seen as a metadata schema that is likely to be useful for *any* class of intellectual property and for *any* application. The selection of the metadata to be included in the kernel has been based on this analysis.

[12] For the avoidance of any doubt, it is necessary to restate that the existence of an extended metadata schema to support a specific DOI-AP brings with it no presupposition that such extended metadata should be made freely available to anyone. Only the DOI Kernel (see Section 11 of this Appendix) is freely declared and publicly available.

[13] For example, the ISBN has commonly been inappropriately used to identify items of stationery – or even, notoriously, soft toys. Metadata declaration, while not eliminating the possibility of such inappropriate use of DOIs should help to minimise it.

4. A definition of metadata

An item of metadata is a relationship that someone claims to exist between two entities.

[From *The <indecs> Framework*]

The word metadata means many things to many people. So we begin from this definition of metadata that provides us with a concise paraphrase of much of the <indecs> framework. This definition stresses the significance of relationships, which lie at the heart of the <indecs> analysis. It underlines the importance of unique identification of all entities[14] (since otherwise expressing relationships between them is of little practical utility). Finally, it raises the question of authority: the identification of the person (individual or corporate) making the claim that a relationship exists is as significant as the identification of any other entity.

At the same time, it underlines the essentially boundless nature of metadata – the relationships between any entity and other entities are potentially infinite. Attempting to define a metadata schema for "all the metadata about something" would literally be an infinite task.

DOI-AP metadata schema definition relates primarily not to a rather abstract definition of what metadata *is* but rather to the requirement to support a specific application relating to the entity identified and described by the specific metadata schema. It is only what metadata *does* (or allows us to do) that is ultimately important to us.

5. Metadata and the principle of unique identification

The most significant of the <indecs> principles in support of metadata interoperability is the **Principle of Unique Identification**: *every entity should be uniquely identified within an identified namespace*. This applies equally to metadata elements as to any other entities. To the extent that metadata elements are not drawn from well-structured data dictionaries and controlled value lists, the cost of data maintenance will be increased and metadata interoperability will be compromised.[15]

Terms included in metadata schemas for DOI-APs are drawn, so far as possible, from controlled value lists[16] that are common to different schemas. We recommend the use of value lists from other authorities,[17] and devise our own wherever this is necessary in support of the general objective of facilitating future interoperability between DOI-AP metadata schemas.

6. Some basic definitions

Ultimately, this document (like any document about metadata) is about definitions. To some readers, it may appear that we are being unnecessarily pedantic; to others, we may be defining terms in ways that do not conform to their understanding of what the word "means". In a field as imprecise as natural language, it is inevitable that there will be disagreements on nuances of meaning.

In this document, words mean what **we** say they mean – and we ask readers to ensure that they consider our definitions carefully with this in mind. This does not represent a claim that

---

[14] An entity is defined in <indecs> as something which is identified – see Table 1.

[15] In really well-formed metadata, the only attributes of an entity not drawn from a controlled vocabulary are the names or titles by which entities are known; in some exceptional circumstances, even these may be controlled.

[16] The insistence on the use of controlled values – so far as this is practicable – in DOI-AP metadata schemas is intended to aid interoperability through the enforcement of a controlled vocabulary, rather than to create some artificial limitation on terminology. The <indecs> framework has been selected because of its flexibility and extensibility.

[17] Most notably ISO, but also from specialised data dictionaries where these are appropriate and rigorously developed and maintained. An example is provided by EPICS – see footnote 24.

our definition of a term is in some way "superior" to any other definition; it is simply one view of the meaning. However, *sharing the same view is essential if we are to avoid ambiguity and misunderstanding*.[18]

To avoid people assuming that they understand the intended meaning of a particular term, <indecs> has, in certain instances, adopted a somewhat unusual vocabulary; broadly we follow <indecs> usage, to avoid ambiguity.

We begin in Table 1 with definitions of the basic terms that we will use throughout this document.

Table 1: Basic metadata terms used in this paper[19]

Element	Definition
element	An item of metadata
entity	Something which is identified
attribute	A characteristic of an **entity**; something which an **entity** has
value	An instance of an **attribute**
type	A categorization of one or more **attributes** of an **entity** through which it belongs to a group of entities
relation	A connection between two or more entities.
situation	A static relation involving two or more **entities**; a relation in which the **attributes** of entities remain unchanged
Event	A dynamic relation involving two or more **entities**; a relation through which an **attribute** of an **entity** is changed, added or removed
Agent	The role of an **entity** acting in an **event** or sustaining a **situation**; a characteristic active **role** undertaken by an **entity**[20]
Intellectual Property entity[21]	An **entity** defined by law or international convention to be intellectual property. (This is the only type of **entity** that can be identified by a DOI.)

[18] This is in contrast to the approach originally taken by, for example, Dublin Core – where the "15 elements" were specifically chosen on the basis that "everyone knows what they mean". Subsequent developments in the DC community have demonstrated only too forcibly the extent to which even apparently simple words can have rather ambiguous meanings in different communities. DC has subsequently found it necessary to adopt a much more formal approach to definition, but the fact that no formal semantic structure was adopted *ab initio* is undoubtedly the cause of some difficulties that have been encountered in extending DC satisfactorily.

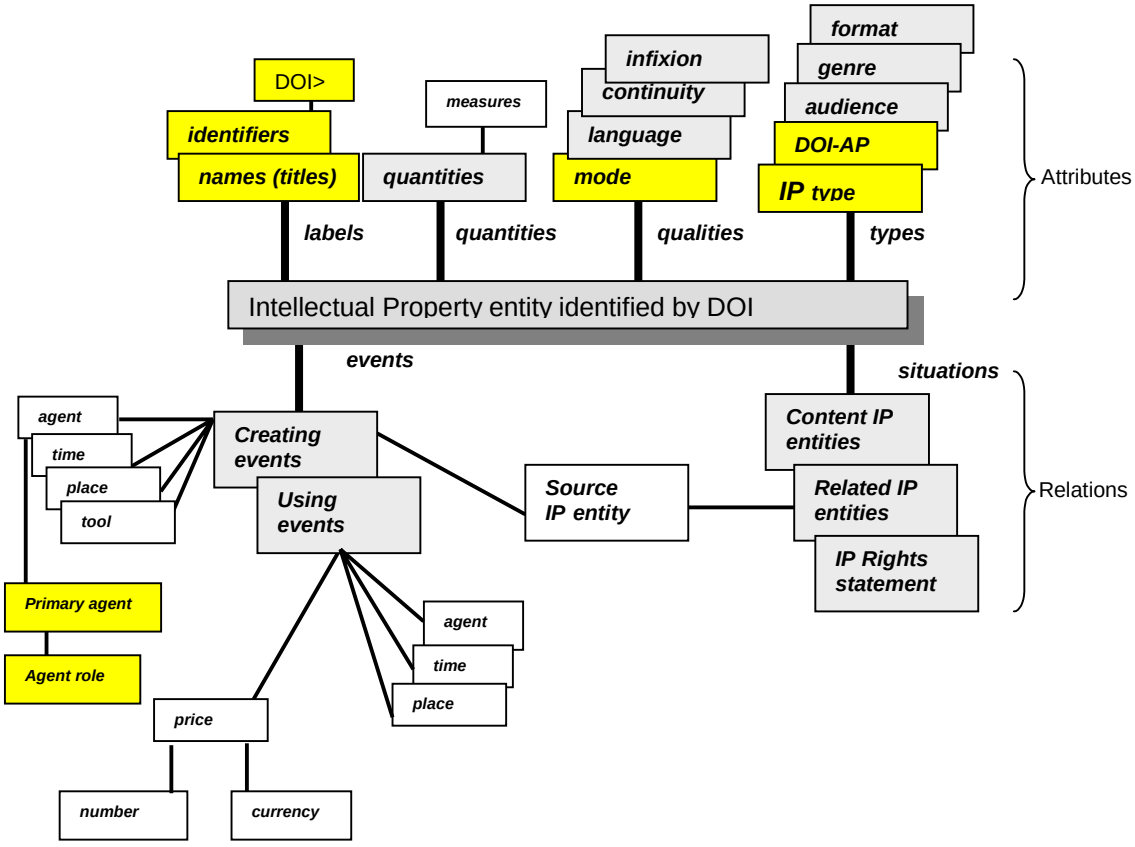
[19] Unless otherwise indicated, the definitions included in this document are adopted from the <indecs> data dictionary. However, in some instances they have been simplified; this document does not attempt to reflect the full range and complexity of the <indecs> Framework. Those seeking more information should see Rust G & Bide M (September 2000) *The <indecs> Metadata Framework* at www.indecs.org.

[20] *Agent roles* are normally fulfilled by people, or by organizations of people, although in principle anything capable of action may be an agent.

[21] The <indecs> framework uses the term "creation", which is any "*output of creative* [making] *activity*". This definition clearly goes beyond the scope of the DOI (which is specifically limited to the identification of IP entities (a subset of creations), and we have therefore preferred to use the term "IP entity" in this document.

7. A template for DOI-AP metadata schemas

The following diagram follows the <indecs> analysis to illustrating the generic classes of attribute that can be used to describe any *e*ntity, and the relations with other entities that are essential to the disambiguation of one IP entity from one another. It is on this analysis that DOI-AP metadata is based. The terms in the template are exemplary rather than exhaustive. Elements that form part of the DOI Kernel are highlighted in yellow.



We will first consider the top half of this diagram (moving from left to right) which looks at the generic attributes of entities; we will then consider the lower half of the diagram, which looks at relations with other entities.

8. Generic attributes

In the <indecs> framework, *any* attribute of *any* entity is identified as being of one of five types – *labels, quantities, qualities, types* and *roles*. Each of these has its own structure and behavior. This provides a *generic* structure for the development and interoperation of metadata schemas for diverse kinds of entity, and is the basic structure of all DOI-AP metadata.

These generic attributes have two characteristics.

The first, as shown in Table 2, is that the values of an attribute have a common form and may be qualified by other common elements. For example, quantity is a *number* value, and needs to be supported by a *measure* (such as *centimeters*) to create a complete attribute.

Secondly, it should be possible for any value from any namespace at any level to be substituted intelligibly as a value of one of its supertypes within the dictionary hierarchy (for

example, the *height* of a book given as *15cms* remains intelligible, though less informative, if shown as a value of the supertypes of *height*: *dimension* and *quantity*.[22]

These characteristics provide an essential part of the framework of interoperability that is designed to allow values originated in one namespace to be recognized and used in another, with greater or lesser degrees of precision. Interoperability is dependent on the use of controlled values for almost all attributes (the exception being labels, although in certain circumstances even these can be drawn from controlled value lists).

Table 2: Generic attributes

Element	Definition	Structure
label	A string whose function is to distinguish one **entity** from another	String
quantity	A number measuring some aspect of an **entity**	Number + measure
quality	A characteristic of the structure or nature of an **entity**; an intrinsic characteristic	Adjective
type	A categorization of one or more attributes of an **entity** through which it belongs to a group of **entities**; a characteristic **role** played by an **entity**	Noun
role	A part played or function fulfilled by an **entity** in relation to another entity or entities; a classification of an **entity** in terms of its external relations; an extrinsic classification	Noun

Names: the names by which things are known (commonly called titles when applied to IP entities) are the commonest form of label. An IP entity registered in the DOI System will almost invariably have a title[23]; it will commonly have more than one. Titles may be structured or unstructured. However, they are different from other attributes in that typically they are not drawn from a value list or from a controlled vocabulary.

Identifiers: identifiers are structured labels that adhere to a set of rules. An IP entity registered in the DOI System will always have at least one identifier – a DOI. Commonly, it may have one or more "legacy" identifiers (for example, an ISBN). The IDF expects to recognize the EPICS[24] namespace object identifier values (see EPICS Element 0001 *Object identifier type values*) as the value list for identifiers for use in DOI-APs.

Quantities: quantity attributes of IP entities may, for example, be a measurement of their physical form (the physical format of a book, for example); or they may be temporal (the length of a recording) or digital (number of bytes in a file). The IDF has recognized the need for a definitive structure for units of measure for use in DOI-APs and will develop this as required to meet the needs of specific APs.

Qualities and types: Both "qualities" and "types" contribute to the process of *classification* of an entity. The distinction between qualities and types is a very fine one that it may not always be necessary or helpful to make. Qualities are adjectives; typically involve single attributes; and are more intrinsic to the entity being described. Types are nouns; typically involve two or

[22] Not all the terms that are used in this document are included in the definitions given. A more broadly-based data dictionary is included in the <indecs> Framework document, but this also requires further development – see Section 14 of this Appendix)

[23] It may only be a catchword or an opening line.

[24] EPICS, the EDItEUR Product Information Communication System, provides an extended data dictionary for product description that has been developed to support e-commerce in the book trade; it provides the underpinning of the ONIX International bibliographic data interchange standard which is gaining rapid acceptance (see www.editeur.org for further information). EPICS was developed in parallel with <indecs> and shares a common intellectual analysis. The IDF will therefore be adopting certain key value lists from EPICS where these are appropriate for use in DOI-AP metadata schemas.

more attributes; and are more extrinsic to the entity being described. <indecs> found this distinction useful to make not least because it creates a structure for analyzing the construction of value lists – are the different values within a list genuinely comparable (ie should they form part of the same value list)?

However, we recognize that in many instances, it is enough to speak simply of the "classes" to which an entity belongs.

There is potentially an infinite number of different classes of IP entities, both in terms of their qualities and of different types to which they might belong. The diagram shows a few of the more common qualities and types that it may be necessary to record to fulfill the functional requirements of a particular DOI-AP. (See Section 10 of this Appendix for more examples.)

The recording of three of these is mandatory for any IP entity registered in the DOI System – its primary IP type (abstraction, performance or manifestation – see Section 10 of this Appendix), the DOI-AP to which it belongs and its mode. Those defining a new DOI-AP should either draw from an existing <indecs>-based analysis of the appropriate classes needed for a specific application (for example, the DOI namespace itself – see Section 14 – or EPICS), or must alternately develop their own analysis appropriate to the particular Application Profile.

9.Relations

While many descriptors of an entity may be managed in terms of its attributes, if we are to provide all the metadata necessary to distinguish one entity from another it is necessary to describe the relations it has with other entities. Relations consist of two or more **entities** that play **roles** in relation to one another. This thesis is central to the <indecs> metadata analysis (see the <indecs> definition of metadata in Section 4 of this appendix).

In ascending order of complexity, it is possible to describe these relations in terms of **attributes**, **situations** and **events** (see Table 1 for definitions). In contrast to a conventional resource-based approach to commerce metadata, an event-based structure offers major advantages for metadata interoperability.

Even simple attributes, like names, can be described in terms of an event. It is entirely possible to define an event, in this case a "naming event", in which someone (in a given time and in a given place) gave the entity its name. Other simple attributes (like color, for example) *can* be described in exactly the same way. However, while it is entirely *possible* to describe the relationship between an entity and, say, its name in terms of an event, such a structure is entirely unnecessary for most purposes[25] and introduces a degree of unwarranted complexity. Attributes are all that is commonly needed for the description of such simple relationships. It is only the efficient management of more complex relationships that demands a more complex metadata structure.

Situations

In general, in defining the metadata structures necessary to support a DOI-AP, it is rarely necessary make full use of the complexity of structure that the <indecs> analysis provides for the description of events.[26] Relationships too complex for simple attribution can be described in terms of the role which one entity plays with respect to another in a **situation**. In such a description, the event that gave rise to the situation is implicit.

---

[25] But not for all purposes. It is possible to define some applications where knowing who named something and when might be of great potential significance.

[26] Although this will not be the case if the application for which the metadata schema is developed is directly related to intellectual property rights and transactions. Such complex applications will require more sophisticated metadata analysis than we will describe in this introductory paper.

Table 3: Some situation types

Element	Definition
association	A **situation** in which two or more **entities** are passively related; a relation based on the verb to be
iprStatement	A **situation** in which one or more parties possesses intellectual property rights

The following are examples of relations that are likely to be those most commonly required in the development of DOI–AP metadata.

Content relationships: most IP entities can be broken down into smaller, discrete IP entities (which may or may not have different rights associated with them). The relationships between an IP entity and its constituent entities can be expressed very simply (for example, in a table of contents) or in a more complex set of **associations** between IP entities (in which an IP entity *isPartOf* another IP entity).[27] Where more complex sets of associations are created, it is essential that each component is uniquely identified, preferably by its own DOI, and fully described in accordance with the rules of the Application Profile.

It is frequently true (but not inevitable) that the metadata schema for a component will be different from the metadata schema of the composite IP entity of which it is a part.

Associations with other IP entities: IP entities often have relations with other IP entities, aside from component relationships. These may be derivative relationships or they may be associative (for example, two different manifestations of the same abstraction have an association even if one is not derived from the other). As can be seen from the diagram, the relation with a "parent" IP entity may be expressed either in terms of an event (see below) or (more simply) an association.

The choice of which is more appropriate depends entirely on the application; for many purposes, the information, for example, that one IP entity *IsTranslationOf* another IP entity is all that is needed (and information about the circumstances in which the translation was made is not essential).

IP Rights Statement: an IP rights statement is the simplest of rights structures. It describes the ownership of an intellectual property right, or of some entitlement to agree to further exploitation of an IPR. iprStatements are based on "possessor" roles, such as owner or rights administrator; the IDF is developing the necessary role codes to allow the expression of simple IPR statements in DOI-AP metadata schemas.

Events

Events are a powerful tool for the construction of interoperable metadata structures. However, they are also complex, and a detailed description of the <indecs> events based structure is not possible within the constraints of this document. Any reader who wishes to get a fuller understanding of the events structure should read the <indecs> Framework document.[28] However, we will look briefly here at two types of event that we believe are likely to be central to the management of DOI-AP metadata in more complex applications.

"Creating" events: The relationship between an IP entity and its creator or creators may often be simply expressed in terms of the agent and the role the agent played in the creation. As with other **associations,** this is simply a more compressed mechanism for the expression of a relationship between the IP entity and its "creating event"; the model allows for the creating event to be described at any level of complexity (with as many different creation events and agents as may be necessary).

[27] The development of controlled vocabulary for the description of component and derivative relationships is a current priority for the IDF)

[28] www.indecs.org

Creating events, depending on the role of the agent, will often simultaneously be **"using" events** (translation is a good example). Again, it will often be unnecessary to describe such using events explicitly – their existence will be implicit in the existence of an association with an agent playing the appropriate role (*translator* in this example). Role codes in "creating" and "using" events are central to the <indecs> analysis, and the IDF is developing a set of defined role codes for use in DOI-APs.

"Using" events: It is in describing availability and pricing that it may first become essential to be able fully to express events within the DOI metadata structure. This will involve describing not only events in which an IP entity *has been* used in the past but also events in which the IP entity *may be* used in future (and the conditions of such use). Such events, which are called "offers" in the <indecs> framework, provide a framework for information on availability and pricing, elements that we anticipate are likely to prove to be essential components of future DOI-AP metadata schemas.

The IDF is developing standardised structures for the common management of structures of this type within DOI-AP metadata schemas, on the model of what EPICS calls "composites". These will greatly simplify the task of defining metadata schemas for new DOI-APs, since standard constructs will be available to developers "off the shelf"; such constructs will be designed to be extensible.

A note on roles and types

All the same elements feature as both **roles** and **types** (see Table 2): when *roles* are attributed to an entity outside of the setting of an event or situation they become characteristic *types*. For example, Beethoven was the *composer* of *"Fidelio"* in an event; so from this and other events it is established that Beethoven was an entity of the type *composer*. Similarly if a *translation* is the output of a translating event, then translation is a *type* of the output creation.

Any entity fulfilling a role in a relation may then be said to be of the type described by the role, although in practice such attribution tends only to occur when an entity is identified *regularly* with a particular role; if a person once played a brief part in an amateur stage production, it would be technically correct but misleading to characterize them generally as an "actor".

10. Classification of IP entities

Primary types

An **intellectual property entity**[29] can be one of three primary types: manifestations (physical "things"); expressions (or performances – intellectual property that is created in an event); and abstractions (commonly known as works).

[29] See Table 1 and footnote 21.

This relationship between these different types of intellectual property may best be described graphically:

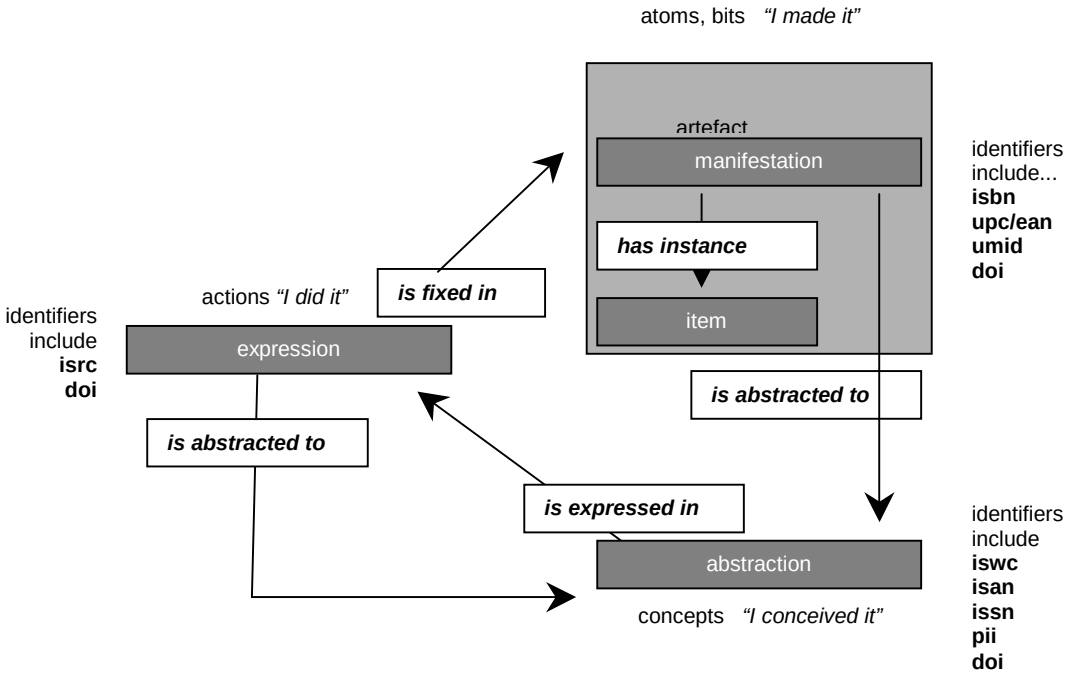


Table 4: Primary types of intellectual property

Element	Definition
artefact	Something that is made
item	A single instance of an **artefact**
manifestation	An **artefact** containing an **infixion** of an **expression**
expression	An **event** which involves the creation of an **intellectual property entity**
abstraction	An **abstract intellectual property entity** whose existence and nature are inferred from one or more **expressions** or **manifestations**

The distinction between different types of intellectual property is crucial to the proper management of IP entities. It is particularly important to be certain what type of IP entity is being identified in a particular DOI Application Profile, since often there is a significant difference between what people *think* they want to identify and what they actually want to identify.

Such decisions can seem counter-intuitive (and are often confused with the question of what a DOI identifies and what it resolves to).

We have found, for example, that it may sometimes be difficult to recognize that a digital file is an (intangible) manifestation; on the other hand, a cross reference or a citation may be made more appropriately to an abstraction than to any one particular manifestation. In performance, there is also a potential problem in confusing the performance itself (an expression) with a recording of the performance (a manifestation)

Other typing of intellectual property

A **type** is a categorization of one or more characteristics of an entity through which it belongs to a group of entities. It can be, for example, a typical role played by any entity. There are many ways of typing IP entities, some examples of which are included in Table 5.

Table 5: Some types of intellectual property

Element	Definition	Examples of possible subtypes
audience	The user or group of users intended to enjoy an IP entity in one or more modes	See EPICS Element
format	An **artefact** on which an **expression** may be infixed to create a **manifestation**	See EPICS Element 0501 Product form code values

Qualities of intellectual property entities

A **quality** of an IP entity is an *intrinsic* characteristic of that entity. There are many significant aspects of **quality** of an intellectual property entity, some examples of which are included in Table 6.

Table 6: Some qualities of intellectual property

Element	Definition	Examples of possible subtypes
continuity	The nature of dynamism of an **entity** over time	dynamic; static
completion	The status of an **entity** in the course of the creative process	draft; finished; etc.
infixion	The means of representation or fixing in which an **expression** of an **abstraction** is established in or on a **manifestation**	analogue; bitEncoded
language	A particular form of verbal or symbolic expression of an **abstraction** or an **expression**	ISO 639-2/B
mode	The sensory mode or modes through which an **entity** may be perceived	audio; visual; multimodal; audioAndVisual
origination	The process by which an **IP Entity** is made	See EPICS Element
substance	The form of the material of which an entity is made	physical; digital

11. Mapping the DOI Kernel metadata scheme to <indecs>

The following table sets out the DOI kernel metadata that must be declared whenever a DOI is registered.[30]

Table 7: The DOI Metadata Kernel

Element	Definition	Status	Number	Allowed values	Possible AP qualifications
DOI	A DOI	Mandatory	1 only	DOI	
DOI-AP	A class of entities with common attributes	Mandatory	1 minimum	From DOI-AP tables	
Identifier	A unique identifier (e.g. from a legacy scheme) applied to the entity	Qualified by AP	1 minimum	Any alphanumeric string but when present must include an identifier type, e.g. ISBN	Define in application: it is normal to include a legacy identifier if one exists.
Title	A name by which the entity is known	Mandatory	1 minimum	Any alphanumeric string	Define in application; a value of "untitled" may be allowable in certain APs.
Type	The primary structural type of the entity	Mandatory	1 only	From: *Abstraction* *Tangible Manifestation* *Intangible Manifestation*[31] *Performance*	
Mode	The primary sensory mode by which the entity is intended to be perceived	Mandatory	1 minimum	From: *Visual* *Audio* *Audio+Visual* *Abstract*	Define in application; a value of "unknown" may be allowable in certain APs.
Primary agent	The name or identifier of the primary agent(s) (normally but not necessarily the creator).	Mandatory	All primary agents. (1 minimum, but all entities fulfilling the same agent role must be included.)	Identifier or Name from an agreed namespace	The specification of the Primary Agent for any AP is determined by the DOI-AP rules.
Agent role	The role(s) played by the primary agent(s)	Mandatory	1 minimum	Role code from an agreed namespace	

Since the DOI Kernel metadata schema is designed to be generic for any IP entity, the Kernel-AP (to which all IP entities identified with a DOI belong) may be seen as the only DOI-AP that has a functional specification but no specification of the IP Entity type that it defines. The DOI Kernel metadata may describe any entity that can be identified with a DOI.

The DOI Kernel metadata schema is designed to fulfill the following functional specification:

[30] In addition to the minimal kernel metadata elements, a certain amount of administrative data – such as Registrant, date of registration, record version number – is also mandatory.

[31] The kernel makes a distinction between tangible and intangible manifestations quite deliberately. Intangible manifestations are commonly, but not invariably, digital files. This distinction is made in the kernel arbitrarily, on pragmatic grounds, since these two different classes of manifestation have such different characteristics and behaviour.

1. **[DOI]** To assign a unique DOI to the IP entity
2. **[IDENTIFIER]** To link the DOI to the principal local or legacy identifier of any IP entity (if one exists) to facilitate the integration of DOI-related applications and metadata with others
3. **[TITLE, PRIMARY AGENT, PRIMARY AGENT ROLE]** To enable a searcher[32] to identify the IP entity by its most common name and the party or parties responsible for its creation or publication.
4. **[TYPE]** To enable a searcher to distinguish the primary type of IP entity (abstract work; tangible or intangible manifestation; performance) and thereby also to distinguish between IP entities of different types with the same name.
5. **[MODE]** To enable a searcher to distinguish the intended mode of perception of an IP entity and thereby also to distinguish between different IP entities with the same name.
6. **[DOI-AP]** To enable a searcher to determine within which DOI-AP (or DOI-APs) a particular IP Entity belongs. This not only provides additional information about the IP Entity itself, it also provides information about which services are available that relate to that IP Entity.

The following table shows how the Kernel elements map to the <indecs> generic attribute model:

Table 8. The DOI Metadata Kernel mapped to <indecs> generic structures

<indecs> generic attribute	DOI Kernel Metadata Element
label	DOI
	Identifier
	Title
type	Type (primary structural type)
	AP
quality	Mode
relation	Primary agent
	Agent role

12. An example: mapping the CrossRef metadata schema to the DOI kernel and to <indecs>

The CrossRef metadata schema is, as we have discussed, intended to support a simple application – automated support for citation linking between journal articles. Its structure is therefore very simple and can be mapped easily to the DOI Kernel and to <indecs>.

[32] In this context, a "searcher" may be an individual human being carrying out a manual search, but may equally be an automated application.

Table 9. The CrossRef metadata schema mapped to the DOI Metadata Kernel and to <indecs> generic structures; implicit values apply to all IP entities registered in this AP. Based on CrossRef Rules V0.4 of 22 September 2000.

<indecs> attribute	DOI Kernel Metadata Element	CrossRef Metadata schema
label	DOI	DOI
		URL
	Identifier	Article identifier
		Identifier type
	Title	Title
		Enumeration
measure		*<not used for this application>*
type	Type (primary structural type)	<implicit> abstract work
	DOI-AP	<implicit> CrossRef
quality	Mode	<implicit> visual
relation	Primary agent	Publisher
	Agent role	<implicit> disseminator
creatingEvent		
agent		Given_name
		Surname
agent role		Author
		Author_sequence
isPartOf		Journal
		Journal title
		Abbreviated title
		ISSN
		CODEN

13. A metadata set for the “DOI Application Profile” AP

Since each DOI-AP is an IP entity in its own right, and is to be registered with its own DOI, we need a DOI-AP within which to register all DOI-APs. The functional requirement of the “DOI-AP” AP is to provide a registry of DOI-APs, with a history of their development, and the ability for a user to identify the type of IP entity that is described.

The functional requirement of this AP is to allow a searcher to locate an AP that may be appropriate to a particular class of IP Entity or to a particular functional specification. This functional requirement implies the need to name and define the application or applications for which the DOI-AP has been devised.

Table 10. The draft "DOI-AP" AP metadata schema mapped to the DOI Kernel and to generic <indecs> attributes.

<indecs> attribute	DOI Kernel Metadata Element	Draft "DOI-AP" AP Metadata schema
label	DOI	DOI
	Identifier	<none>
	Title	AP name
		Version Number
measure		*<not used for this application>*
type	Type (primary structural type)	<implicit> abstract work
	AP	<implicit> DOI AP
quality	Mode	<implicit> abstract
		Origination: original or derived
origination		[If derived, DOI of original AP(s)]
subject		Name of IP Entity type described
		Primary IP Type of IP Entity Type described
		Mode of IP Entity type described
		Format of IP Entity type described
		Other limiting attributes as required
		Name(s) and description(s) of application(s)
relation	Primary agent	Responsible organization
	Agent role	authority
creatingEvent		
agent		Author of schema
agent role		Author
context		Date and place of authoring of schema

14. Controlled vocabularies

The use of controlled vocabularies in DOI Metadata Schemas is an essential element in underpinning future interoperability. Some of these can naturally be adopted from existing ISO standards (languages, countries, currencies, time). Others we expect to adopt from EPICS.

However, we also recognize that there is a considerable task that needs to be undertaken in the development of controlled vocabularies in areas where there is currently no appropriate development available. This includes:

- Quantities

- Agent Role codes (essential for future rights management applications, but where the EPICS value list is currently limited and needs expansion if it is to meet the requirements of the IDF)
- Component and derivation codes (again, EPICS value lists might provide a basis)
- Rights ownership and management codes

Similarly the IDF needs to select – and where necessary create – appropriate generic constructs for use within DOI-AP metadata schemas to manage common “composites” (like pricing and availability).

Much can be achieved through work on two projects that have been initiated by the IDF. Firstly, the EPICS data dictionary is being explicitly mapped to the <indecs> data dictionary – the <indecs> structure is currently only implicit in EPICS. Secondly, the published <indecs> metadata dictionary is being expanded beyond the 200 or so definitions contained in the <indecs> Framework document. A considerable amount of additional work was completed during the lifetime of the project, but this work is now being reviewed and validated before it is released.

The outcome of this work, due to be complete in mid-2001, will be the development of a DOI namespace (DOI-NS) of controlled vocabularies for the construction of DOI-AP metadata schemas. The DOI-NS will be made available to anyone who is designing a DOI Application Profile; additional value lists will be added to the DOI-NS as requirements are identified.

Finally, it is particularly important for the IDF that work is started to map out the necessary semantics for rights management applications; this is an area where the work of the <indecs> project has provided the necessary signposts, but where a considerable task remains to be done. The IDF recognizes the need to collaborate with partners from many different sectors in the further analysis that is required.

15. Conclusions

The concept of the DOI Application Profile continues to evolve as new APs are implemented in practice; nevertheless, the basic principles of definition of a DOI-AP by a community of interest are now clear. The process of defining an Application Profile will always begin by:

- The naming of a class of Intellectual Property entities
- The definition of that class of IP entities in terms of its limiting attributes and attribute values
- The definition of an application (or a group of applications) relating to the class of IP Entities that will in some way create value
- The specification of the functional requirements of that application (those applications)

From that can be derived:

- A metadata schema to support the functional requirements
- The commercial and procedural rules which will support a robust implementation of the application

The metadata schema must adhere to the rules established in this Guide, in order to ensure future extensibility and interoperability (even if the need for both of these with respect to any particular implementation cannot as yet be identified). We actively encourage the greatest possible simplicity in the approach to defining DOI-AP metadata schemas, to ensure that the burden of compliance does not prove to be unmanageable.

The commercial and procedural rules must comply with overall IDF policy (including policy relating to metadata declaration) but are as lightly constrained as is possible,

Anyone – organization or individual – contemplating the definition of a new DOI Application Profile is encouraged to approach the IDF at the earliest possible opportunity. This will help to minimize duplicated efforts and will provide access to IDF resources.

Appendix 4: Overview of the relationship between the IDF and DOI Registration Agencies

This is a top-level indication of the anticipated issues that need to be agreed between the IDF and individual Registration Agencies. It is undergoing substantial further development in conjunction with interested parties.

Introduction

The IDF is moving to full deployment of the DOI in co-operation with Registration Agencies and potential RAs, not seeking to impose rules on RAs without discussion. The requirement to "implement mechanisms for quality control", for example, should not be seen as implying that a pre-existing set of criteria for quality control is already available; it is expected that Registration Agencies (RAs) will propose suitable and practicable criteria.

The relationship between RAs and the IDF will ultimately be contractual. These specifications are being considered as initial headings for the development of mutually acceptable contract terms.

Technical requirements

Registration Agencies will:

- Assign DOI Prefixes to new registrants in accordance with IDF standard terms; liaise with the resolution system provider to register prefixes in system directories.
- Ensure that DOIs under this prefix are loaded with corresponding URLs (or other data types) into a globally available resolution system nominated by the IDF (eg CNRI's Handle) in timely/accurate manner.
- Liaise with IDF and CNRI to agree definition of any necessary additional data types.
- Promote the use of native resolution protocols (HDL) in applications in preference to proxy server implementations

Information management requirements

Registration Agencies will:

- Ensure that appropriate minimal supporting metadata for each DOI entity (the DOI Kernel metadata) is declared and made freely available for look up.
- Support the “Base AP” (kernel metadata only)
- Support at least one additional Application Profile appropriate to the particular community of interest served and the applications provided to them. RAs may use any AP, but must have an agreed AP for each DOI registered.
- Manage AP metadata by use of, and in conformance to, a schema and data dictionaries agreed with the IDF.
- Deposit an escrow copy of data with the IDF under agreed terms.

General requirements

Registration Agencies will:

- Implement mechanisms for quality control of DOI resolution and metadata registration.
- Support and promote multiple resolution capability.
- If they wish to do so, subcontract or partner with others to deliver part of their services.

The IDF will:

- Control the relationship with the global resolution provider, on behalf of all RAs and the DOI community as a whole.

Commercial issues

- There will be no "one size fits all" fee for all applications of the DOI, and RAs are expected to develop their own commercial model. Fees need to be set at a realistic level.
- Permanent exclusivity (either of AP or territory) will not be granted to any applicant, although the IDF will work with RAs sympathetically to provide some comfort in investment and commitment.
- IDF will not become a profit-making entity (in the sense of "distributing profit to non-member shareholders") but it is desirable that the IDF generate a surplus to be reinvested in further development.
- Any agreement with RAs must allow for revision (possibly of a radical nature) as the system develops. In particular, a mechanism must exist for renegotiation in the event that one RA attempts to monopolize registrations by "excessively subsidized" strategies.
- The IDF will be neutral as to exclusivity of applications that are the basis of value added services, but will wish to co-ordinate standard services so far as possible.
- Resilience and reliability of the Handle System's Global Handle Service is a requirement for commercial levels of use. IDF works with CNRI to ensure this. In mid 2000, significant improvements were made by introducing redundancy for the proxy servers, and by moving the primary and secondary DOI handle server, the secure web server for DOI administration, the primary Global Handle Registry server, and a secondary global server to a commercial host (Exodus Communications Corp.) secure site with guaranteed 24 x 7 coverage and 99.97% uptime. Two additional secondary global servers are maintained in California. Further improvements to resilience will be made as required.

Appendix 5: The Handle System

This appendix provides a high level overview of CNRI's Handle System. For more detailed information, visit the Handle Web site at www.handle.net.

1. Handle System Overview

Laurence Lannom, Corporation for National Research Initiatives

An earlier version of this overview first appeared as an article in ICSTI Forum, No 30, April 1999 http://www.icsti.org/icsti/forum/fo9904.html

Introduction

The Handle System® 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.

History and Applications

The Handle System was originally conceived and developed at CNRI as part of the Computer Science Technical Reports (CSTR) project, funded by the Defense Advanced Projects Agency (DARPA) under Grant No. MDA-972-92-J-1029. One aspect of this early digital library project, which was also a major factor in the evolution of the Networked Computer Science Technical Reference Library (NCSTRL - see http://www.ncstrl.org/) and related activities, was to develop a framework for the underlying infrastructure of digital libraries. It is described in a paper by Robert Kahn and Robert Wilensky [1]. The first implementation, created at CNRI, was made available on the Internet in the fall of 1994. Subsequent work on the Handle System has been supported in part by the Defense Advanced Research Projects Agency under Grant No. MDA972-92-J-1029.

Early adopters of the Handle System have included the Library of Congress, the Defense Technical Information Center (DTIC), and the International DOI Foundation (IDF). Feedback from these organizations as well as NCSTRL, other digital library projects, and related IETF efforts have all contributed to the evolution of the Handle System. Current status and available software, both client and server, can be found at http://www.handle.net/. This web site, as well as the DOI site (http://www.doi.org/) also provides many examples of the use of handles.

The Handle System has evolved within the digital library community, but it was conceived and built as the naming component of an overarching digital object architecture, as described in Kahn/Wilensky [1] and subsequent papers [2, 3]. It has potential application not only beyond the early adopters such as the IDF, DTIC, and LC, but also well beyond the digital library area. As a general-purpose indirection system that resolves identifiers into state information, the Handle System can be used to advantage in any dynamic network environment as part of the overall process of managing digital objects. Interest has been expressed by organizations in application areas such as telephony (linking individuals with multiple phone numbers, 'telephone number for life', etc.), and crisis management (resource tracking). Any given

application area would have to build its own tools and approaches, but the Handle System, especially as part of the larger digital object architecture referenced above, can serve as an information management substrate for a wide variety of application areas.

Need for a General Purpose Naming System

The need for a general purpose naming system has increased with Internet growth. While there are existing services and protocols that cover some of the functionality proposed in the Handle System, and while we make no claim that the Handle System is the only such service that is now or ever will be needed, we do believe that the Handle System provides needed functionality that is not otherwise available.

There are several services that are in use today to provide name service for Internet resources, of which the Domain Name System (DNS) [4,5] is the most widely used. DNS is designed "to provide a mechanism for naming resources in such a way that the names are mappable into IP addresses and are usable in different hosts, networks, protocol families, internets, and administrative organizations" [5]. The growth of the Internet has increased demands for various extensions to DNS, and even its use as a general-purpose resource naming system, but its importance in basic network routing has led to great caution in implementing such extensions and a general conclusion that DNS is not the place to look for general purpose resource naming. An additional factor that argues against using DNS as a general purpose naming system is the DNS administrative model. DNS names are typically managed by the network administrator(s) at the DNS zone level, with no provision for a per name administrative structure, and no facilities for anyone other than network administrators to create or manage names. This is appropriate for domain name administration but less so for general-purpose resource name administration. The Handle System has been designed from the start to serve as a naming system for very large numbers of entities and to allow administration at the name level.

URLs (Uniform Resource Locators) [6] allow certain Internet resources to be named as a combination of a DNS name and local name. The local name may be a local file path, or a reference to some local service, e.g. a cgi-bin script. This combination of DNS name and local name provides a flexible administrative model for naming and managing individual Internet resources. There are, however, several key limitations. Most URL schemes (e.g., http) are defined for resolution service only. Any URL administration has to be done either at the local host, or via some other network service such as NFS. Using a URL as a name typically ties the Internet resource to its current network location, and to its local file path when the file path is part of the URL. When the resource moves from one location to another, for whatever reason, the URL breaks.

The Handle System is designed to overcome these limitations and to add significant increased functionality. Specifically, the Handle System is designed with the following objectives:

Uniqueness. Every handle is globally unique, within the Handle System.

Persistence. A handle 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 handle for convenience, the only operational connection between a handle and the entity it names is maintained within the Handle System. This of course does not guarantee persistence, which is a function of administrative care, but it does allow the same name to persist over changes of location, ownership, and other state conditions. For example, when a named resource moves from one location to another, the handle may be kept valid by updating its value to reflect the new location.

Multiple Instances. A single handle can refer to multiple instances of a resource, at different and possibly changing locations in a network. Applications can take advantage of this to increase performance and reliability. For example, a network service may define multiple entry points for its service with a single handle name and so distribute the service load.

Extensible Namespace. Existing local namespaces may join the handle namespace by acquiring a unique handle naming authority. This allows local namespaces to be introduced

into a global context while avoiding conflict with existing namespaces. Use of naming authorities also allows delegation of service, both resolution and administration, to a local handle service.

International Support. The handle namespace is based on Unicode 2.0 [7], which includes most of the characters currently used around the world, facilitating the use of the system in any native environment. The handle protocol mandates UTF-8 [8] as the encoding used for handles.

Distributed Service Model. The Handle System defines a hierarchical service model such that any local handle namespace may be serviced either by a corresponding local handle service or by the global service or by both. The global service, known as the Global Handle Registry™, can be used to dispatch any handle service request to the responsible local handle service. The distributed service model allows replication of any given service into multiple service sites and each service site may further distribute its service into a cluster of individual servers. (Note that local here refers only to namespace and administrative concerns. A local handle service could in fact have many service sites distributed across the Internet.)

Secured Name Service. The handle protocol allows handle servers to authenticate their clients and to provide data integrity service upon client request. Public key and/or secret key cryptography may be used. This may be used to prevent eavesdroppers from forging client requests or tampering with server responses.

Distributed Administration Service. Each handle may define its own administrator(s) or administrative group(s). This, combined with the Handle System authentication protocol, allows handles to be managed securely over the public network by authorized administrators at any network location.

Efficient Resolution Service. The handle protocol is designed to allow highly efficient name resolution performance. To avoid resolution being affected by computationally costly administration service, separate service interfaces (i.e., server processes and their associated communication ports) for handle name resolution and administration may be defined by any handle service.

Handle Name Space

Every handle consists of two parts: its naming authority, otherwise known as its prefix, and a unique local name under the naming authority, otherwise known as its suffix. The naming authority and local name are separated by the ASCII character "/". A handle may thus be defined as

< Handle> ::= < Handle Naming Authority> "/" < Handle Local Name>

For example, " 10.1045/january99-bearman " is a handle for an article published in the D-LIB magazine [9]. It is defined under the Handle Naming Authority "10.1045", and its Handle Local Name is "january99-bearman ". (see Fig. 1)

Figure 1

The handle namespace can be considered as superset of many local namespaces, with each local namespace having its own unique handle naming authority. The naming authority identifies the administrative unit of creation, although not necessarily continuing administration, of the associated handles. Each naming authority is guaranteed to be globally unique within the Handle System. Any existing local namespace can join the global handle

namespace by obtaining a unique naming authority, with the resulting handles being a combination of naming authority and local name as shown above.

Handles may consist of any printable characters from the Universal Character Set, two-octet form (UCS-2) of ISO/IEC 10646, which is the exact character set defined by Unicode v2.0. The UCS-2 character set encompasses most characters used in every major language written today. To allow compatibility with most of the existing systems and prevent ambiguity between different encoding, handle protocol mandates UTF-8 to be the only encoding used for handles. The UTF-8 encoding preserves any ASCII encoded names, which allows maximum compatibility to existing systems without causing naming conflict. Some encoding issues over the global namespace and the choice of UTF-8 encoding are discussed in [3].

Handle naming authorities are defined in a hierarchical fashion, i.e., a tree structure. Each node and leaf of the tree is given a label that corresponds to a naming authority segment. The parent node presents the parent naming authority of its child nodes. Unlike DNS, handle naming authorities are constructed left to right, concatenating the labels from the root of the tree to the node that represents the naming authority. Each label is separated by the octet used for ASCII character '.' (0x2E). For example, a naming authority for the digital library ("dlib") project at the Corporation for National Research Initiatives ("cnri") is defined as "cnri.dlib".

Each naming authority may have many child naming authorities registered underneath. Any child naming authority can only be registered by its parent after its parent naming authority is registered. However, there is no intrinsic administrative relationship between the namespaces represented by the parent and children naming authorities. The parent namespace and its child namespaces may be served by different handle services, and they may or may not share any administration privileges among each other.

Every handle is defined under a naming authority. The naming authority and the local name are separated by the octet used for ASCII character '/' (0x2F). The collection of local names under a naming authority is the local namespace for that naming authority. Any local name must be unique under its local namespace. The uniqueness of a naming authority and a local name under that authority ensures that any handle is globally unique within the context of the Handle System.

Handle System Architecture

The Handle System has a two-level hierarchical service model. The top level consists of a single global service, known as the Global Handle Registry. The lower level consists of all other handle services, which are generically known as local handle services. The global service is a handle service like any other and can be used to manage any handle namespace. It is unique among handle services only in that it provides the service used to manage the namespace of handle naming authorities, all of which are managed as handles. The state information of these naming authority handles is the service information that clients can use to access and utilize associated local services. The local handle service layer consists of all local handle services managing all handles under their naming authorities, providing resolution and administration service for these local names. Local services are intended to be hosted by organizations with administrative responsibility for the handles within the service or acting on behalf of the responsible organizations. The most convenient way to define local namespaces, and the most likely way to optimize overall Handle System performance, is by naming authority and it is anticipated that in most cases all handles under a given naming authority will be maintained by one service. This is not required, however, and it is possible for handles under a single naming authority to be split among multiple handle services. Handle services may be responsible for more than one naming authority. Another way of stating all of this is that the relation of handle naming authorities and handle services is allowed to be many-to-many in both directions, but that the relationship of naming authority to handle service is most likely to be one-to-one and that the relationship of handle service to naming authority is likely to be one-to-many.

A second important component of Handle System architecture is distribution. The Handle System as a whole consists of a number of individual handle services, each of which consists

of one or more handle service sites, where each site replicates the complete individual handle service, at least for the purposes of handle resolution. Each handle service site in turn consists of one or more handle servers. There are no design limits on the total number of handle services, which constitute the Handle System, there are no design limits on the number of sites that make up each service, and there are no limits on the number of servers that make up each site. Replication by site, within a service, does not require that each site contain the same number of servers, that is, while each site will have the same replicated set of handles, each site may allocate that set of handles across a different number of handle servers. This distributed approach is intended to aid scalability and to mitigate problems of single point failure. (see Fig. 2)

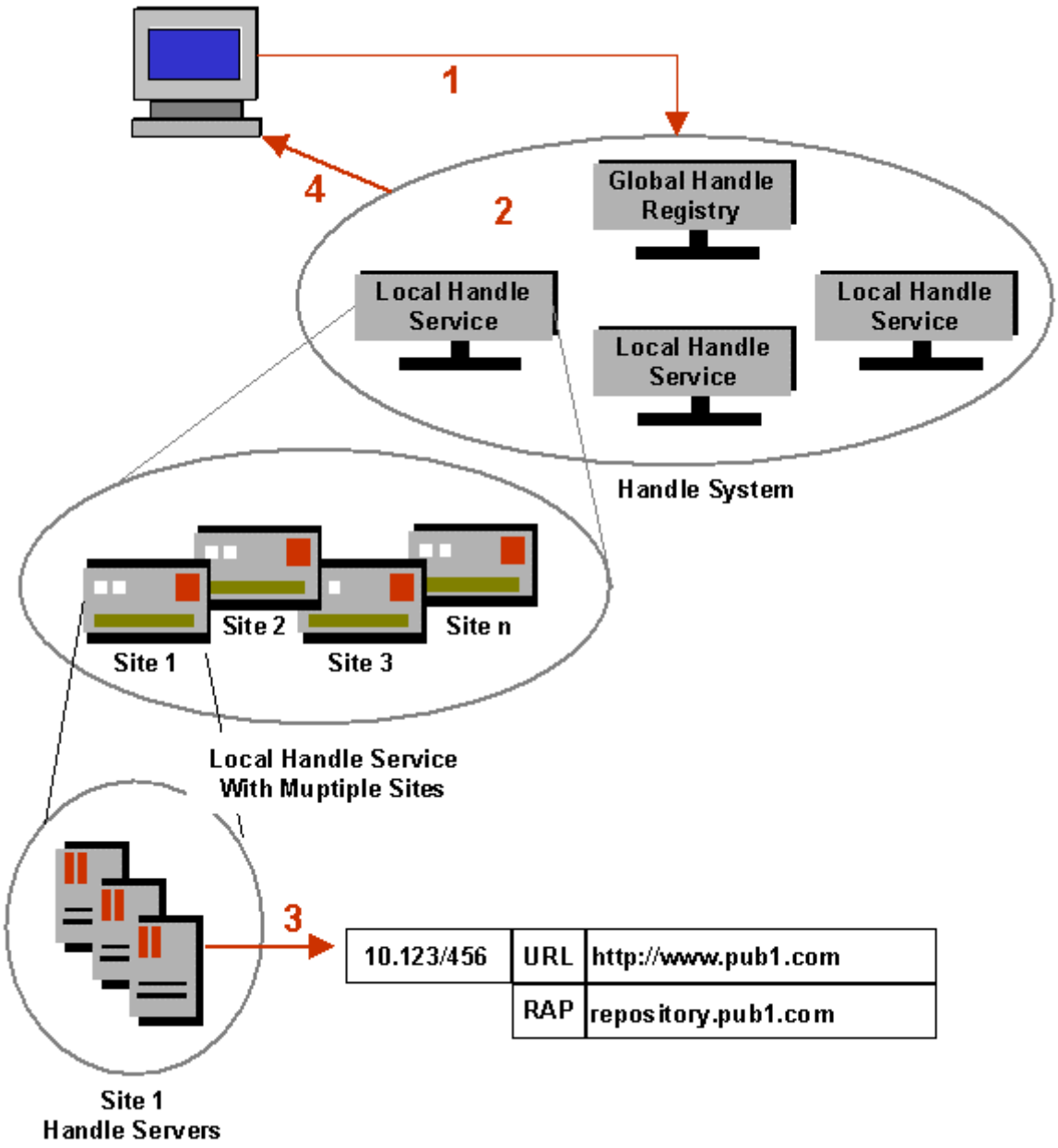


Figure 2

1. A client such as a web browser encounters a handle, e.g., 10.123/456, on the Internet or an individual intranet, typically as a hyperlink or other kind of reference. The client sends the handle to the Handle System for resolution. This can be done directly by a client that understands the handle resolution protocol natively or through a proxy server by a client that doesn't.

2. The Handle System consists of a collection of handle services, as described above. One service, the Global Handle Registry, is responsible for knowing the locations and name space responsibilities of all of the public Local Services. Each of these Local Services knows how to access the Global Handle Registry. This allows a resolution query to enter the Handle System at any point and to be routed to the specific service and server that knows the answer.

3 & 4. Each handle can be associated with one or more pieces of typed data. In this example, the handle 10.123/456 is associated with, and so resolves to, both a URL and a new protocol called RAP. This is the information that is returned to the client. Note that it would also be possible to associate multiple instances of the same data type, e.g., multiple URLs, with a single handle. The Handle System is a pure resolution system and carries no assumptions on what the client will or will not do with the resolution information, thus maximizing the flexibility of applications which use the Handle System as an infrastructure for naming. In the example shown, the client can presumably use either protocol to locate and retrieve the item, although again this is up to the client.

To improve resolution performance, any client may select to cache the service information returned from the global service, and/or the resolution result from any local service. A separate handle caching server, either stand-alone or as a piece of a general caching mechanism, may also be used to provide shared caching within a local community. Given a cached resolution result, subsequent queries of the same handle may be answered locally without contacting any handle service. Given cached service information, clients can send their requests directly to the responsible local service without contacting global.

Conclusion

Early deployment of the Handle System has served to confirm the basic design concepts, as described in this article, and significant progress has been made in understanding the complexities and issues involved in designing effective digital object naming and resolution systems. It is a large problem space, however, and a great deal of work remains in this area as well as many others as we attempt to navigate from the current world to one in which the primary sources of information are digital objects on networks.

This has been a very brief overview of the Handle System. Many more technical details, explanation, contact information, software, and updates are available at http://www.handle.net.

References

[1] Kahn, Robert and Wilensky, Robert. "A Framework for Distributed Digital Object Services", May, 1995. http://www.cnri.reston.va.us/k-w.html

[2] Arms, William Y., Christophe Blanchi, Edward A. Overly, An Architecture for Information in Digital Libraries, D-Lib Magazine, February 1997. http://www.dlib.org/dlib/february97/cnri/02arms1.html

[3] Sam X. Sun, "Internationalization of the Handle System - A Persistent Global Name Service", Proceeding of 12th International Unicode Conference, April, 1998, http://www.cnri.reston.va.us/unicode-paper.ps

[4] P. Mockapetris, "Domain Names - Concepts and Facilities", RFC1034, November 1987. http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1034.txt

[5] P. Mockapetris, "Domain Names - Implementation and Specification", RFC1035, November 1987. http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1035.txt

[6] Berners-Lee, T., Masinter, L., McCahill, M., et al., "Uniform Resource Locators (URL)", RFC1738, December 1994. http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1738.txt

[7] The Unicode Consortium, "The Unicode Standard, Version 2.0", Addison-Wesley Developers Press, 1996. ISBN 0-201-48345-9

[8] Yergeau, Francois, "UTF-8, A Transform Format for Unicode and ISO10646", RFC2044, October 1996. http://info.internet.isi.edu:80/in-notes/rfc/files/rfc2044.txt

[9] D-Lib Magazine, http://www.dlib.org/

2. Handle System Scalability

Scalability was a critical design criterion for the Handle System. The problem can be divided into storage and performance. That is, is there some limit to the number of handles that can be added? And, does performance go down, or do some functions simply break with increased numbers of handles, such that at some point the system becomes unusable? Specific details on this are given below, but it is important to keep two higher-level issues in mind. First, it is important here, as in many other places, to distinguish between Handle System design and any given implementation. Scalability in design may or may not work out as expected in any given implementation, but if the design is fundamentally scalable, specific implementation problems can be corrected as they are encountered. Secondly, use of the Handle System through some other service, e.g., an http proxy, may well introduce other scalability issues that the basic Handle System design does not and cannot address.

Storage

The Handle System has been designed at a very basic level as a distributed system, that is, it will run across as many computers as are required to provide the desired functionality.

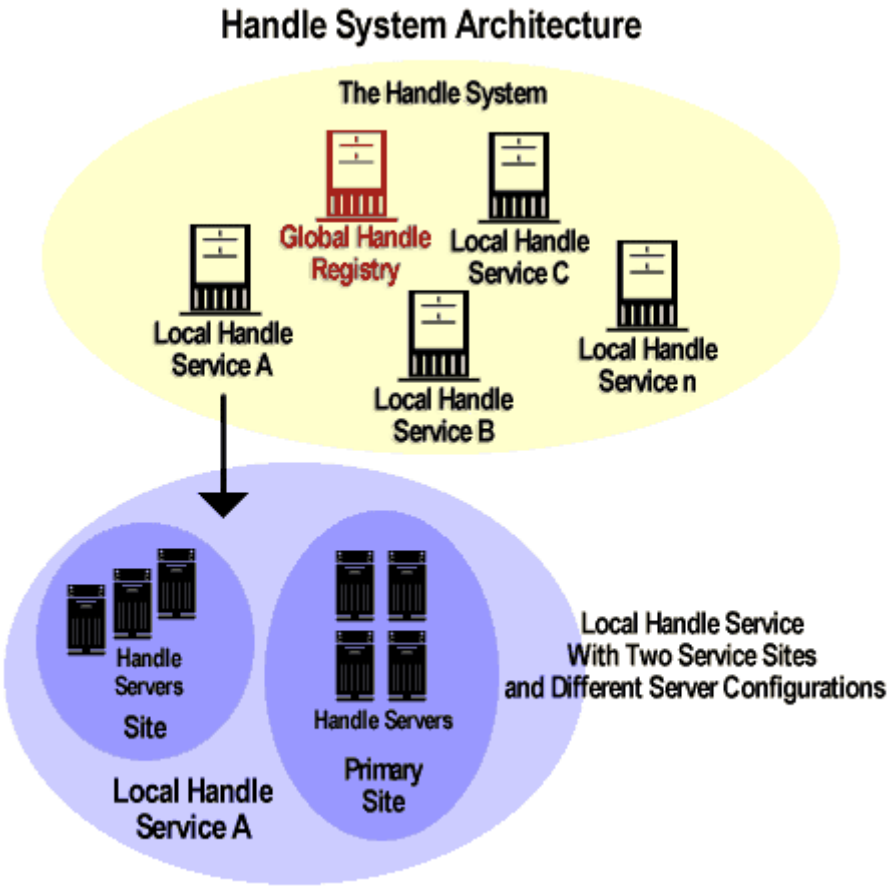


Figure 3

Handles are held in and resolved by handle servers and handle servers are grouped into one or more handle sites within each handle service. There are no design limits on the total number of handle services that constitute the Handle System, there are no design limits on the number of sites that make up each service, and there are no limits on the number of servers that make up each site. Replication by site, within a service, does not require that each site contain the same number of servers; that is, while each site will have the same replicated set of handles, each site may allocate that set of handles across a different number of servers. Thus increased numbers of handles within a site can be accommodated by adding

additional servers, either on the same or additional computers, additional sites can be added to a service at any time, and additional services can be created. Every service must be registered with the Global Handle Registry, but that service can also have as many sites with as many servers as needed. The result is that the number of handles that can be accommodated in the current system is limited only by the number of computers available.

Performance

Constant performance across increasing numbers of handles is addressed by hashing, replication, and caching.

Hashing, a technique well known to database designers, is used in the Handle System to evenly allocate any number of handles across any number of servers within a site, and allows a single computation to determine on which server within a set of servers a given handle is located, regardless of the number of handles or the number of servers. Each server within a site is responsible for a subset of handles managed by that site. Given a specific handle and knowledge of the service responsible for that handle, a handle client selects a site within that service and can perform a single computation on the handle to determine which server within the site contains the handle. The result of the computation becomes a pointer into a hash table, which is unique to each handle site and which can be thought of as a map of the given site, mapping which handles belong to which servers. The computation is independent of the number of servers and handles, and it will not take a client any longer to locate and query the correct server for a handle within a service that contains billions of handles and hundreds of servers, than for a service that contains only millions of handles and only a few servers.

The connection between a given handle and the responsible handle service is determined by naming authority. Naming authority records are maintained by the Global Handle Registry as handles, and these handles are hashed across the Global Handle Registry sites in the same way that all other handles are hashed across their respective service sites. The only hierarchy in Handle System services is the two level distinction between a single global and all locals, which means that the worst case resolution would be that a client with no built in or cached knowledge would have to consult Global and one local.

Another aspect of Handle System scalability is replication. The individual handle services within the Handle System each consist of one or more handle service sites, where each site replicates the complete individual handle service, at least for the purposes of handle resolution. Thus, increased demand on a given service can be met with additional sites, and increased demand on a given site can be met with additional servers. This also opens up the option, so far not implemented by any existing clients, of optimizing resolution performance by selecting the "best" server from a group of replicated servers.

Caching may also be used to improve performance and reduce the possibility of bottleneck situations in the Handle System, as is the case in many distributed systems. The Handle System data model and protocol design includes a space for cache time-outs and handle-caching servers have been developed and are in use.

---

3. Building Handle System Applications – tools

Handle System software is available for both **clients** and **servers**. On the **client side**, the choice of software components for download depends on the type of resolution services being offered.

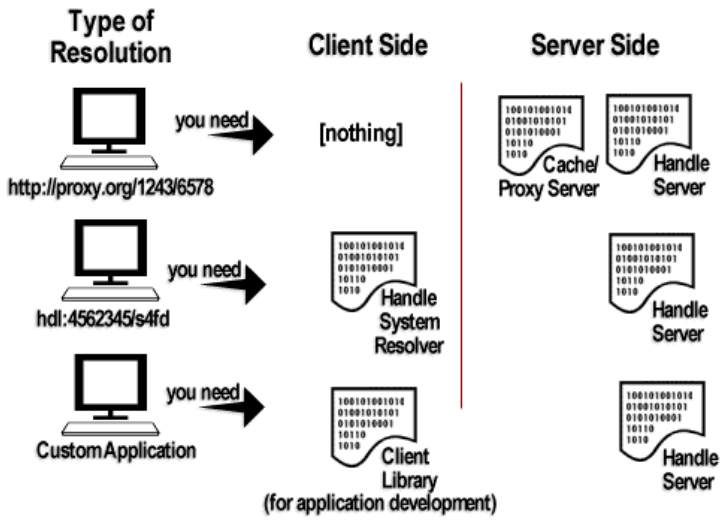


Figure 4: Client Side and Server Side Software

Currently available client side software includes:

- the **CNRI Handle System Resolver** (Version 2.0 beta) which enables web browsers to recognize the handle protocol and is now available for Windows 95, Windows 98, and Windows NT. (Versions for additional platforms are under development.)
- the Handle System **Client Library** (JAVA™ Version 5.1), a library of Java classes which understands the handle protocol and would form the foundation for Java-based custom client software development.
- the Handle System **Client Library** (C Version 5.1), a library of C functions which understands the handle protocol and would form the foundation for custom client software development. This library was used in the development of the Handle Resolver.

On the **server side**, handle service configuration can be customized. One site within a service is designated a primary site, and each site may contain one or more handle servers. The local handle server operates as part of the distributed system and enables specialized identifier, resolution and administration services on a single computer or multiple computers. All site configurations support mirroring, which increases reliability and performance by storing handle data on multiple computers, generally maintained at different locations.

Currently available server side software:

- **HSj** (Local Handle System Server Version 5.1) JAVA Version. This release includes a handle administrative client which enables:
    - administering handles (creating, deleting, and modifying handle data),
    - batch deposits, edits, and deletions,
    - creating naming authorities and homing naming authorities,
    - adding and deleting administrators and managing administrator permissions.
    - checkpointing and backing up the database
    - listing handles under a given naming authority

  Also included in the server distribution is code for a combined proxy server/caching server, with complete installation instructions.

Appendix 6 An Introduction to the <indecs> metadata framework

This introduction to the <indecs> framework draws heavily on the introduction to The <indecs> Metadata Framework*, a final project document from the <indecs> project published in July 2000 (reproduced with permission). This document summarizes the technical work of the <indecs> project. It describes the <indecs> metadata framework, a reference model.* **The <indecs> project***, and its successor the not-for-profit* **Indecs Framework Ltd***, was created to address the need, in the digital environment, to put different creation identifiers and their supporting metadata into a framework where they could operate side by side, especially to support the management of intellectual property rights*

<indecs> (the name is an acronym: **In**teroperability of **D**ata for **E**lectronic **C**ommerce **S**ystems) was a project that with the backing of DGXIII of the European Commission, brought together as partners and affiliates a global grouping of organizations with an interest in the management of content in the digital environment. The complete background and objectives of the <indecs> project are documented on the <indecs> website www.indecs.org. However, this brief introduction to the project deals with the question of **interoperability**: what does it mean in practice and why is it important?

A model of commerce

People make stuff. People use stuff. People do deals about stuff.

The stuff and the deals may come in any order, but neither come before the people.

This is the basic model of commerce that underlies the <indecs> framework and models. While the approach described here may be usefully applied in many domains, the main focus of <indecs> is on the use of what is commonly (if imprecisely) called “content” or “intellectual property”.

The model applies in many contexts, but is particularly useful in the digital and Internet environments, where the problems of metadata interoperability are becoming especially acute.

Commerce is used here in its broadest sense, not necessarily having financial gain as its object. The model applies equally to cultural transactions in places such as libraries in which people “make deals” that enable others to have free access to stuff for various purposes.

The <indecs> schema rests on certain fundamentals, or axioms, about electronic commerce.

*A*xiom 1: Metadata is critical

“Metadata is the lifeblood of e-commerce” (a phrase coined by John Erickson, then of Yankee Book Peddler). 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 (metadata, or data about data).

E-commerce requires the linking of identifiers that connect people with goods and services: *stuff*. In dealing with intellectual property these identifiers form complex and dynamic chains. All kinds of metadata elements find their way into them: where there is a gap or an ambiguity in these elements it is likely that the chains will be broken, or misrouted, and the required transaction will not happen, or will have the wrong results. As e-commerce grows, reliance on these metadata chains grows with it.

Axiom 2: Stuff is complex

The second axiom on which <indecs> rests is that, when dealing with intellectual property, *stuff is complex.* The generic <indecs> term for a piece of stuff that may carry intellectual property rights is a **creation**. While an apple bought at a market stall is a single physical entity owned entirely by one person, a single digital audiovisual creation may contain hundreds or even thousands of separate pieces of intellectual property. These may include moving pictures, recorded audio, still photographs, graphics, text and software applications, some only in part or in modified form. Each of these separate **manifestations** of intellectual property may have rights.

These manifestations are normally **expressions** of abstract works or **abstractions** in which there may be further rights; and those expressions may be through the medium of spatio-temporal **performances** in which yet further rights may exist. All of these rights may be owned or controlled by different people for different places and for different periods of time. The trading of one digital creation may involve rights transactions affecting thousands of people and companies, from whom permissions may be required and to whom payment may be due.

To take an example from music, an audio CD "greatest hits" compilation containing twenty tracks is in fact a manifestation, owned (let us say) by a record company. It contains twenty sound recordings, each of which embodies an expression or performance perhaps owned by different record companies or artists and in which, in some territories, each contributing performer has certain rights. Each performance in turn expresses one or more songs (abstractions) in which the composer(s) and publisher(s) have rights. Permissions and payments are required whenever the CD is bought or used. These deals (agreements) may be brokered individually or by collective licensing arrangements.

While this example is taken for music, similar kinds of complex relationship can be found in any other creation type. From type to type the importance and quantity of different elements may vary (for example, in text-based creations the *performance* element is often unimportant) but the functional requirements are the same, in structure if not scale.

Axiom 3: Metadata is modular

Because stuff is complex, *metadata is modular*. e-commerce metadata is made up of connecting pieces created by different people.

Each of the basic entities (people, stuff, deals or *parties, creations, transactions*) must have its own metadata set if creations are to be found and used and rights are to be protected and rewarded. If the rights in a complex creation come from many different people, so will the metadata. Constraints of cost, time and knowledge ensure that the multimedia producer is dependent on his suppliers of content also to provide him with the metadata on which its future exploitation depends. The same dependency is increasingly true for others in the chain, including non-profit-driven organizations such as libraries and academic institutions.

Metadata in the digital environment can therefore be viewed as a set of "modules", produced in different places and for different purposes, which must link together easily into complex forms to create new metadata modules for different stuff, people and deals*.* The result can be described as the *metadata network*, or in a narrower context the *semantic web.*

Axiom 4: Transactions need automation

In an increasing range of cases, *transactions need to be highly or completely automated*. In physical commerce much metadata complexity has been dealt with (if at all) in administrative systems within bounded organizations such as publishers or collecting societies operating their own local data standards and systems. The scale and nature of e-commerce has made it imperative that these local standards and systems can interoperate in automated ways with others.

For example, in the non-digital environment, securing copyright "permissions" is a complicated, time-consuming and often unsatisfactory process. Owners and publishers are

already often unable to cope with the volume of low-value permissions requests made in conventional ways.

In the digital environment the volume and nature of such uses is increasing exponentially. Because stuff is complex and technology is ingenious – and because the virtual world does not recognize national boundaries – the number of creations, agreements and potential rights holders multiplies rapidly and continually. Without automation, all but the most valuable permissions will become impossible to administer.

Interoperability

In the <indecs> framework, interoperability therefore means enabling information that originates in one context to be used in another in ways that are as highly automated as possible. *Commerce* does not necessarily mean the exchange of money: any environment where *creations* are *made* or *used* employing electronic means is encompassed by commerce in this sense.

The information that needs to interoperate here is *metadata:* data of all kinds relating to *creations*, the *parties* who make and use them, and the *transactions* that support such use. The problems to be overcome are often as basic as the fact that a term such as *publisher* has quite different meanings in two different environments which now need to exchange metadata; but they are also as complex as the fact that a single creation may contain a hundred distinct pieces of intellectual property, the rights of which are owned or controlled by many different people for different purposes, places and times. In the persistent environment of the Web, changes in the status or control of these rights which are recorded in different and unconnected systems will need to be communicated automatically to all appropriate points.

Types of interoperability

Interoperability in e-commerce has many different dimensions. As traditional sectors and business models break down, organizations increasingly face the need to combine or access information that comes in a variety of forms and from a variety of sources. The creator of metadata about a piece of intellectual property will want to be sure that the accuracy and effectiveness of the information s/he creates (often at substantial cost) can survive intact as it negotiates a range of barriers. A serious approach to the problem needs to support interoperability of at least six different types:

- Across *media* (such as books, serials, audio, audiovisual, software, abstract works, visual material)
- Across *functions* (such as cataloguing, discovery, workflow and rights management)
- Across *levels* of metadata (from simple to complex)
- Across *linguistic* and *semantic* barriers
- Across *territorial* barriers
- Across *technology platforms*

A good e-commerce metadata system therefore needs to be *multimedia, multi-functional, multi-level, multilingual, multinational* and *multi-platform.* Such an approach is said to be *well-formed.*

Failures to interoperate across each of these dimensions can be seen as the impending *trade barriers* to e-commerce interoperability. These barriers are not all yet critical, but only because the volume of e-commerce traffic in intellectual property is currently relatively modest: yet we are now seeing an unprecedented explosion in the development of intellectual-property-based metadata schemas. Listed alphabetically below are just some of the major initiatives where substantial metadata vocabularies, models, databases and/or interchange formats are currently being developed or deployed, showing the communities in which they currently operate or from which they were originated.

abc

CIDOC	(museums and archives)
CIS	(copyright societies)
DMCS	(recording industry)
Dublin Core	(library originated)
EPICS/ONIX	(book industry)
IFLA FRBR	(libraries)
IMS	(education)
IDF	(book/journal originated)
IEEE LOM	(education)
MPEG7	(audiovisual originated)
MPEG21	(audiovisual originated)
P/META	(audiovisual)
SMPTE	(audiovisual)

This is by no means a complete list, although it represents most of the major initiatives with which the <indecs> project communicated. These schemes, developing from different starting points, are converging on the “trade barriers” we have identified. To some degree, each is finding that it *has* to become multi-media, multi-function, multi-level, multi-lingual and technology neutral: and as convergence renders the traditional sector divisions increasingly meaningless, they will inevitably need to interoperate with one another substantially. In the future the same metadata about, for example, a web document, may easily need to be handled within each of these schemes – and many more.

The limits of technology

Web-driven tools such as XML (the Extensible Mark-Up Language) and RDF (the Resource Description Framework) and their derivatives will provide many solutions: but they can only go so far, and do not deal with the underlying issue of semantic identity. Ultimately it is only the deployment of unique identifiers across a wide range of critical pieces of metadata – well beyond what is currently practiced – that will allow the trade barriers between different metadata schemas to be surmounted without requiring an uneconomic level of human intervention and interpretation.

Such identification systems are more or less implicit in the schemes listed above: but as things stand today they will, unintentionally, find themselves competing to no good purpose. The <indecs> framework has been developed to provide a reference model for system implementers to avert this costly clash of standards and provide an underlying infrastructure for semantic interoperability between them. To be successful, the cost of compliance with this infrastructure must be low, its implementation relatively straightforward and it must facilitate, not obstruct, the development of local systems or schema like those listed above.

Such an infrastructure will depend on *semantic mapping* through *metadata registries*. The development of such tools is outside the scope of the project, and at the time of writing (July 2000) is in its very early stages. However, the implication of the <indecs> analysis is that powerful tools and systems for mapping and *transforming* metadata will provide the technical key to interoperability.

The project also recognizes that “make once, use many times” metadata is the only viable economic model for the future, and that as far as possible, such metadata needs to be automatically generated as a by-product of other processes.

Intellectual property metadata

The focus of <indecs> is intellectual property: “rights management”. However, this is not a domain that is separate from other metadata issues. While there are particular legal aspects involved in the establishment and use of rights, these are intimately connected with the everyday activities of the making and use of “creations”, so a well-formed system must provide mechanisms for the interoperation of this metadata if it is to enable automated rights management to be possible.

Intellectual property issues are wholly pervasive in e-commerce: every transaction involving the use of digital creations at any point in the “supply chain” is, in some sense, a rights

transaction, even where no money changes hand. Rights management is as important for the protection of legitimate “fair use” by libraries as it is for the protection of rights owners to exploit their intellectual property. The <indecs> framework is neutral on the merits or otherwise of any given right or practice but is concerned only with the mechanisms of describing the transactions that take place.

Characteristics of the <indecs> framework

The framework recognizes:

- metadata relating to any types of *creation;*
- the integration of *descriptive* metadata with *commercial transactions* and *rights;*
- that metadata should be created once, used many times for different purposes;

and proposes

- a *generic attribute structure* for all entities;
- *events* as the key to complex metadata relationships;
- a *metadata dictionary* for multimedia intellectual property commerce;
- unique identifiers (*iids*) to be assigned to all metadata elements;
- the need for *transformation* processes to express the same metadata at different levels of complexity for different requirements.

At the heart of the model lies the assumption that it is possible to develop generic systems to handle complex metadata for all different *creation* types. So, for example, instead of treating sound carriers, books, videos and photographs as fundamentally different things with different, albeit similar, characteristics, they are all recognized as creations with different values of the same higher-level attributes, whose metadata can be supported in a common environment.

The <indecs> framework is designed to help bridge the gap between the powerful but highly abstracted technical models such as that expressed in the Resource Description Framework (RDF) and the more specific data models that are explicit or implicit in sector- or identifier-based metadata schemes.

Selected Bibliography

This selected bibliography has been compiled for readers who wish to explore some of the themes covered by this Handbook. The list is partial and subjective, but may help readers to get started with wider reading. This is an area undergoing rapid development. Later references are therefore to be preferred for current information; older references may be useful for historical perspective.

Web sites of particular relevance

IDF	http://www.doi.org
CNRI	http://www.cnri.reston.va.us
CrossRef	http://www.crossref.org
Handle System	http://www.handle.net
<indecs> Framework	http://www.indecs.org

DOI specific papers and reports

Arms, William A. "Digital Object Identifiers (DOIs) and Clifford Lynch's five questions on identifiers". *ARL: A Bimonthly Newsletter of Research Library Issues and Actions* 194 (October 1997). Washington, DC: Association of Research Libraries. http://www.arl.org/newsltr/194/arms.html

Atkins, Helen et al. "Reference Linking with DOIs: A Case Study". *D-Lib Magazine, February 2000*. [doi:10.1045/february2000-risher] http://dx.doi.org/10.1045/february2000-risher

Bernstein, Paula. "DOI: A New Identifier for Digital Content". *Searcher: The Magazine for Database Professionals*, Volume 6, Number 1, January 1998. http://www.infotoday.com/searcher/jan98/story4.htm

Bide, Mark. *In Search of the Unicorn: The Digital Object Identifier From a User Perspective* BNBRF Report 89, Book Industry Communication, February 1998. http://www.bic.org.uk/unicorn2.pdf

Erickson, John. "Metadata Initiatives and the DOI: Implications for Electronic Commerce and Copyright Management Services". *Trialogue*, Summer 1998, Number 8. http://www.ybp.com/yrm/trialogue/898/898metad.htm

Erickson, John. *The Role of Metadata Supply Chains in DOI-Based Value-added Services* Yankee Book Peddler, April 1999. http://www.ybp.com/yps/papers/doiservicesapr99.htm

Erickson, John "The DOI and Rights Management: Tying up Loose Ends" *Trialogue*, No 11, Summer 1999. http://www.ybp.com/yrm/trialogue/1199/1199doi.htm

DOI Deployment IDF Discussion paper http://www.doi.org/deployment2.pdf

"DOI heads toward mainstream metadata, e-commerce initiatives". *Seybold Report on Internet Publishing*, Volume 3 Number 5, January 1999. Available to Seybold Report subscribers. [See http://www.seyboldreport.com.]

Maree, Jean. "The Digital Object Identifier System and its applications in the Online Information Industry: an overview for Southern African Information Professionals." *Musaion*, V. 16, No. 2, 1998. http://www.unisa.ac.za/library2/afdeling/tdienste/doi.html

Paskin, Norman. "The Digital Object Identifier System", *ICSTI Forum*, August 1998, No. 28.

Paskin, Norman. "DOI: Current Status and Outlook" *D-Lib magazine* May 1999 http://www.dlib.org/dlib/may99/05paskin.html

Paskin, Norman. *E-citations: actionable identifiers and scholarly referencing* [doi:10.1000/170]. Also available from "Learned Publishing", Vol 13, No 3, July 2000, [doi:10.1087/09531510050145308]

Risher, Carol A. and Rosenblatt, William R. "The Digital Object Identifier – An Electronic Publishing Tool for the Entire Information Community". *Serials Review*, Vol. 24, No. 3/4, 1998, JAI Press Inc., pp. 12-20.

Identifiers and identification

Connolly, Dan. "Names and Addressing: URIs". W3C Architecture Domain, 1998. http://www.w3.org/Addressing/Addressing.html

Green, Brian and Bide, Mark. *Unique Identifiers: a brief introduction*. Book Industry Communication/EDItEUR. http://www.bic.org.uk/uniquid.html

IETF Network Working Group. *Uniform Resource Identifiers (URI): Generic Syntax* August 1998. http://www.ics.uci.edu/pub/ietf/uri/rfc2396.txt

IETF URN Working Group. *Uniform Resource Names (urn)* http://www.ietf.org/html.charters/urn-charter.html

Lynch, Clifford. "Identifiers and Their Roles In Networked Information Applications". *ARL: A Bimonthly Newsletter of Research Library Issues and Actions* 194 (October 1997). Washington, DC: Association of Research Libraries. http://www.arl.org/newsltr/194/identifier.html

Paskin, Norman. “Toward Unique Identifiers” *Proceedings of the IEEE*, **87** (no.7) July 1999 pp. 1208-1227

Paskin, Norman. "Information Identifiers". Originally published in *Learned Publishing*, Vol. 10, No. 2, pp 135-156 (April 1997). http://www.elsevier.nl/homepage/about/infoident/

Payette, Sandra. "Persistent Identifiers on the Digital Terrain", *RLG DigiNews*, Vol. 2, No. 2, April 15, 1998. http://www.rlg.org/preserv/diginews/diginews22.html

Powell, Andy. "Unique Identifiers in a Digital World". *Ariadne*, April 1997. http://www.ariadne.ac.uk/issue8/unique-identifiers

Powell, Andy. "Resolving DOI Based URNs Using Squid". *D-Lib Magazine*, June 1998 [doi:10.1045/june98-powell]. http://dx.doi.org/10.1045/june98-powell

Reynolds, Regina R. "ISSN and Seriality". *Library of Congress CONSER,* May 1998. http://lcweb.loc.gov/acq/conser/issn.html

URN implementers “Uniform Resource Names: a progress report” *D-Lib Magazine*, February 1996. http://www.dlib.org/dlib/february96/02arms.html

Metadata

Baker, Thomas. "Languages for Dublin Core". *D-Lib Magazine*, December 1998, [doi:10.1045/december98-baker]. http://www.dlib.org/dlib/december98/12baker.html

Bearman, David et al. "A Common Model to Support Interoperable Metadata: A progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities". *D-Lib Magazine*, Volume 5, Number 1, January 1999. [doi:10.1045/january99-bearman] http://www.dlib.org/dlib/january99/bearman/01bearman.html

Borges, Jorge Luis *John Wilkins' Analytical Language*, (1942); translated in (for example) Weinberger, E (ed.), *Borges: Selected Non-Fictions* Viking, New York, 1999

Priscilla Caplan: "International Metadata Initiatives: Lessons in Bibliographic Control" (http://lcweb.loc.gov/catdir/bibcontrol/caplan_paper.html).

Erickson, John "The role of metadata supply chains in DOI-based, value-added services", *ICSTI Forum*, No.30, April 1999 http://www.icsti.org/icsti/forum/fo9904.html

Carl Lagoze “Keeping Dublin Core Simple” *D-Lib Magazine* January 2001 http://www.dlib.org/dlib/january01/lagoze/01lagoze.html

"The <indecs> metadata framework; Principles, model and data dictionary", June 2000, Godfrey Rust, MUZE Inc., and Mark Bide, EDItEUR http://www.indecs.org/pdf/schema.pdf

<indecs> "Summary Final Report", 2 August 2000.

Rust, Godfrey "Metadata: The Right Approach. An Integrated Model for Descriptive and Rights Metadata in E-commerce" *D-Lib Magazine* July/August 1998 [doi:10.1045/july98-rust] http://www.dlib.org/dlib/july98/rust/07rust.html

Sowa, John F. *Knowledge Representation: Logical, Philosophical and Computational Foundations* Brooks/Cole, 2000

Digital journals and reference linking

Helen Atkins, Catherine Lyons, Howard Ratner, Carol Risher, Chris Shillum, David Sidman, and Andrew Stevens; "Reference Linking with DOIs: A Case Study, D-Lib Magazine, February 2000. http://dx.doi.org/10.1045/february2000-risher

Caplan, Priscilla and Arms, William Y. "Reference Linking For Journal Articles", D-Lib Magazine, July 1999. [doi:10.1045/july99-caplan] http://dx.doi.org/10.1045/july99-caplan

Caplan, Priscilla; Flecker, Dale, "Choosing the Appropriate Copy; Report of a discussion of options for selecting among multiple copies of an electronic journal article", Digital Library Federation Architecture Committee, September 1999. http://www.niso.org/DLFarch.html

Davidson, Lloyd and Douglas, Kimberly. "Promise and Problems for Scholarly Publishing". The Journal of Electronic Publishing, Volume 4, Issue 2, December 1998 http://www.press.umich.edu/jep/04-02/davidson.html

Hitchcock, Steve et al. "Linking Electronic Journals". *D-Lib Magazine*, December 1998. [doi:10.1045/december98-hitchcock]. http://www.dlib.org/dlib/december98/12hitchcock.html

NISO, "Report of the Second Workshop on Linkage from Citations to Journal Literature", June 1999. http://www.niso.org/linkrept.html

NISO, "NISO/DLF/CrossRef Workshop on Localization in Reference Linking; Meeting Report", 24 July 2000, CNRI, Reston, Virginia, USA [http://www.niso.org/CNRI-mtg.html]

Paskin, Norman. "Digital Information Objects and the STM Publisher". STM Annual Report, 1997. http://www.elsevier.nl/homepage/about/diginfo

Paskin, Norman. "E-Citations: actionalble identifiers and scholarly referenceing". Learned Publishing, July 2000.

Van de Sompel, Herbert & Hochstenbach, Patrick. "Reference Linking in a Hybrid Library Environment", *D-Lib Magazine*, October 1999 [doi:10.1045/october99-van_de_sompel]. http://dx.doi.org/10.1045/october99-van_de_sompel

Digital libraries

Arms, William A. Digital Libraries. MIT Press, 1999

Arms, William A. "An Architecture for Information in Digital Libraries". *D-Lib Magazine*, February 1997, [doi:10.1045/february97-arms]. http://dx.doi.org/10.1045/february97-arms

Caplan, P. & Flecker, D. *Choosing the Appropriate Copy* Digital Library Federation Architecture Committee. http://www.niso.org/DLFarch

Cathro, Warwick. *Digital Libraries: a National Library Perspective.* Paper for the Information Online & On Disc Conference, Sydney, Australia, 19-21 January 1999. http://www.nla.gov.au/nla/staffpaper/cathro4.html

Van der Werf, Titia. "Identification, version control and future availability of electronic publications", Koninklijke Bibliotheek http://www.konbib.nl/persons/titia/publ/tap/tap.html

The Internet and e-commerce in Intellectual Property

Cox, Brad *Superdistribution: Objects as Property on the Electronic Frontier* Addison-Wesley, 1996

Evans, Philip & Wurster, Thomas S. *Blown to Bits: How the New Economics of Information Transforms Strategy* Harvard Business School Press, 2000

Gates, Dominic. "Update: Protecting Intellectual Property Online". *PreText Magazine*, December 1997. http://www.pretext.com/dec97/shorts/short4.htm

Gladney, Henry M. *Are Intellectual Property Rights a Digital Dilemma? Controversial Topics and International Aspects* Information Impacts, February 2000 http://www.cisp.org/imp/february_2000/02_00gladney.htm

Kahn, Robert E. and Cerf, Vinton G. *What is the Internet (And What makes it Work*)? Internet Policy Institute, December 1999 http://www.internetpolicy.org/briefing/12_99_story.html

Lessig, Lawrence *Code and Other Laws of Cyberspace* Basic Books, 1999

Levinson, Paul *The Soft Edge: A Natural History and Future of the Information Revolution*. . Routledge, 1997

Rosenblatt, Bill. "Solving the Dilemma of Copyright Protection Online". The Journal of Electronic Publishing, Volume 3, Issue 2, December 1997. http://www.press.umich.edu/jep/03-02/doi.html

Secor, Glen M. “Digital Rights Management Systems: Technology Update” Paper presented at the conference *Digital Rights management and Digital Distribution for Publishing*, New York, February 2000

Shapiro, Carl & Varian, Hal R. *Information Rules: A Strategic Guide to the Network Economy* Harvard Business School Press, 1999

Stefik, Mark J. *The Internet Edge: Social, Technical, and Legal Challenges for a Networked World*, MIT Press, 1999

WIPO (World Intellectual Property Organisation): "Primer on Electronic Commerce and Intellectual Property Issues" (May 2000) http://ecommerce.wipo.int/primer/index.html.