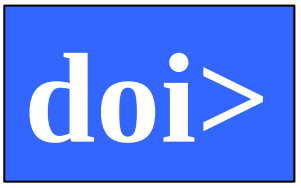
The DOI® Handbook

Version 2.7.0 February 2003



Please send any comments or questions to

Dr Norman Paskin
Director, International DOI Foundation, Inc.,
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

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. More precisely, we adopt the definition of the <indecs> metadata model: an item of metadata is a relationship that someone claims to exist between two entities. In this document we are concerned only with **DOI Metadata**, metadata that describes something identified with a DOI.

Ontology: an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them. The indecs framework produced such an ontology, which has been further developed and forms the basis of the **(DOI) TermSet**.

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

(DOI) TermSet: The dictionary used to normalise and map metadata elements from other schemes within the DOI Metadata system. Elements in the DOI TermSet are part of the indecs Data Dictionary, currently being developed as the Rights Data Dictionary within the proposed international standard MPEG-21 (ISO 21000) part 6. Formerly known as the

DOI Namespace: this terminology was dropped in favour of the more accurate TermSet, and so as not to cause confusion with Namespace in the XML etc. senses.

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

(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 International DOI Foundation 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.

"A developing trend that seems likely to continue in the future is an information centric view of the Internet that can live in parallel with the current communications centric view. Many of the concerns about intellectual property protection are difficult to deal with, not because of fundamental limits in the law, but rather by technological and perhaps management limitations in knowing how best to deal with these issues. A digital object infrastructure that makes information objects "first-class citizens" in the packetized "primordial soup" of the Internet is one step in that direction. In this scheme, the digital object is the conceptual elemental unit in the information view; it is interpretable (in principle) by all participating information systems. The digital object is thus an abstraction that may be implemented in various ways by different systems. It is a critical building block for interoperable and heterogeneous information systems. Each digital object has a unique and, if desired, persistent identifier that will allow it to be managed over time. This approach is highly relevant to the development of third-party value added information services in the Internet environment." (*What Is The Internet (And What Makes It Work) - Robert E. Kahn and Vinton G. Cerf*).

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, following the early acceptance of the principles of the system and its adoption in initial applications. At the publication of this version of the DOI Handbook, several million Digital Object Identifiers (DOIs) had been issued, with several hundred registrant organizations allocating DOIs, six DOI Registration Agencies have been appointed (with more planned to come), and DOI is well integrated into several related standards activities, with many applications actively under development. That development is evident in all aspects of the DOI System – technology, procedure and policy (see Chapter 3).

It is essential to codify the main elements of the DOI in a form that can be easily assimilated by the non-technical reader; at the same time to provide a central point of reference for more complex technical content; and enable updated information to be disseminated. This is the purpose of this Handbook.

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 Table of Contents).

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-publication release of this Handbook (v0.1) was published in June 2000. Edition 1 was issued in February 2001, edition 2 was issued in February 2002, and releases 2.1.0, 2.2.0, 2.3.0, 2.4.0, 2.5.0 and 2.6.0 were issued in April, July, August, September, November 2002 and January 2003 respectively. Version 2.7.0 has incorporated, where possible, material published over the past year in the form of Frequently Asked Questions, and other material previously published in separate documents such as “DOI Deployment”. Earlier versions of those documents are now superseded by this Handbook. The Handbook is intended to be the definitive source of information on the DOI.

If you find any errors, areas which are insufficiently clear, or have any other questions or suggestions relating to this Handbook, please do let us know by contacting info@doi.org; we appreciate your input which will help to improve future versions of the Handbook.

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

1.4 The International DOI Foundation

The International DOI Foundation was created in 1998 and supports the needs of the intellectual property community in the digital environment, by the development and promotion of the Digital Object Identifier system as a common infrastructure for content management. The Foundation is controlled by a Board elected by the members of the Foundation. The International DOI Foundation, Inc. is a non-stock membership corporation organized and existing under and by virtue of the General Corporation Law of the State of Delaware, USA. The Corporation is a "not-for-profit" organization, i.e. prohibited from activities not permitted to be carried on by a corporation exempt from US federal income tax under Section 501(c)(6) of the Internal Revenue Code of 1986 et seq. Although a US legal entity, the Foundation is wholly international in its membership and activities.

1.4.1 Membership of IDF

The activities of the Foundation are controlled by its members, operating under a legal Charter and formal By-laws. Membership is open to all organizations with an interest in electronic publishing, content distribution, rights management, and related enabling technologies. We also welcome comments and participation from non-members. The Foundation develops and establishes policies and procedures and oversees the successful operation of the System.

Members are able to participate in the governance of the Foundation, meetings and working groups, make recommendations to the Board, receive monthly briefings and information on the latest developments and issues. Members will join together with their most important fellow stakeholders to exchange views and to define electronic publishing. Through their dues, Members support the Foundation, and hence the high-quality operation of the DOI System, which will be integral to furthering the interests of the Foundation's members. Members of the Foundation may include:

- larger companies from the publishing, music, still image, broadcast, online news, software and other content industries;
- technology providers to the above listed industries;
- associations representing each of the above industries;
- internet technology companies;
- associations representing additional stakeholders such as authors, artists, libraries, secondary publishers, users;
- government agencies with a mission in a related area;
- others for whom Internet commerce is a critical business, social, cultural, political, or personal interest.

For a list of current members see http://www.doi.org/idf-member-list.html

1.4.2 Governance of IDF

The International DOI Foundation is governed by its members, through an elected board. All seats on the Board are held for a three-year term. There are currently fourteen board seats -- General Members hold four seats, Charter Members hold six seats, and Registration Agencies hold four seats.

The Board officers include a Chair and Treasurer (elected from the Board of Directors) and an appointed full-time Director responsible for carrying out policy formulated by the Foundation.

1.4.3 Benefits of IDF membership

Participating in an international effort to standardize Internet-related technology is perhaps one of the most important steps which can be taken by any organization which uses the digital environment to deliver content. Rather than rely on other organizations to develop the standards that your organization must deal with, participation offers the opportunity to be involved, to shape, influence and assist in making new technology and information standards serve the needs of your organization.

Content must be managed in the digital world; the current tools are insufficient. Doing nothing is not an option. There is clear demand from the market to present content accessibly and digitally. The content community must take the lead or technology companies will fill the vacuum, or will ignore or downplay the rights and intellectual property aspects. No forum does what the International DOI Foundation specializes in. Content and technology communities are represented; our forum is not a partisan development. We have a recognized position and support. Our excellent track record provides the perfect starting place to avoid costs re-inventing the wheel, and allows you access to results and plans (first mover advantage).

The International DOI Foundation focuses on structured responses (well thought through data models). Supporting our efforts guarantees your back-office systems are planned for and doesn't leave you to deal with non-interoperable costly solutions. The DOI System is extensible: you cannot forecast what all the applications will be in advance. In our working groups you have the opportunity to take account of intelligent reviews of others' activities.

Why collaborate? You need to inter-operate with others outside your business; supply chain, customers, and competitors that are adopting the key standards designed for inter-operability. Collaborating through the International DOI Foundation will reduce costs and prevent mistakes and dead-ends when trying to advance as a single company. We provide a common platform, and offer you the opportunity to build added-value services that ensure wider usage if designed to support your interests. We allow you to influence the course of our activities; participate in working groups, meetings, prototypes, stand for election and lobby board members.

The cost of membership for an organisation is low (equal to 2-3 days per month of a consultant), but you benefit immediately as well as for the long term. Members can suggest work in their area of interest and costs don't fall only on you. We provide monthly briefings on other activities, which you cannot afford to attend or monitor in detail (WIPO, W3C, IETF, MPEG, ISO, OEBF, SIIA, and others). You will be involved in a comprehensive effort that will expand your markets; international participation from the US, Asia, Europe and from multimedia industries, e.g., text, music, software, broadcast, images and news.

The more you participate in our forums and tap into our resources, the more you learn how to exploit and control content to ensure flexibility to do business amidst the changing winds of technology. By joining the International DOI Foundation you will soon recognize why participation in developing the DOI System is the prerequisite to digital trading, selling and protection of intellectual property.

Membership in the International DOI foundation (IDF) is open to all organizations with an interest in electronic publishing and its related enabling technologies. Members may be either for-profit or not-for-profit organizations.

1.4.4 Classes of membership

The International DOI Foundation has a membership of committed companies and organizations that participate in the development of the system and its applications. We are in the process of involving a broader community of organizations, which, by nature of their business or market, have a potential interest.

Through their dues, Members support the Foundation and the high-quality operation of the DOI system. Membership in the International DOI foundation (IDF) is open to all organizations with an interest in electronic publishing and its related enabling technologies. Members may be either for-profit or not-for-profit organizations. There are four classes of membership: General, Charter, Registration Agency, and Affiliate.

General Membership is offered to any organizations with an interest in electronic publishing and its related enabling technologies.

Charter Membership is only offered to organizations whose main activities are in the creation or production and dissemination of works of the mind, as defined by the Berne Convention, the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty. The Board of the IDF reserves the right to determine eligibility for the Charter membership category and may in particular refuse eligibility or exclude a member for any reason in the interest of the goals of the Foundation.

Registration Agency membership is only available to organizations, which have signed a Letter of Intent with the IDF. See chapter 11 for details on the appointment and role of Registration Agencies.

Affiliate Membership is a paid membership without voting rights or other full membership rights in the IDF, which entitles the Affiliate to nominate a representative to participate in a DOI Working Group. One Working Group is covered per affiliate membership fee (i.e., participation in two groups requires two fees).
Members are organisations. Membership by **individuals** is normally not open, but the Board reserves the right to allow this in exceptional cases.
Membership is open to other organisations which themselves have member organisations, but in this case the immediate direct benefits of membership as defined in the Rules for Members (such as eligibility to vote, participate in meetings and access to the member-only Web site) only extend to the staff and officers of those organizations, and do not flow through to their own members (who may decide to join the IDF under their own auspices).

1.4.5 Membership fees

The annual fee for General Members is $US 35,000.

The annual fee for Charter Members is $US 70,000, which is reduced to $40,000 if the Charter Member is also a lender of the foundation.

The annual fee for Members may be reduced at the sole discretion of the Board, subject to a minimum fee of $US 11,500 per annum. There are no differences in member rights and benefits between Charter and General, nor for those for which a reduced fee is payable. Criteria which will be considered in applications for such reduction include in particular any of the following:

- Significant role in the creation or ongoing support of the Foundation.
- Not-for-profit organizations, which have annual revenue, as measured by the most recent audited statement, of less than $US 10,000,000.
- For-profit organizations which have annual revenue, as measured by the most recent audited statement, of less than $US 10,000,000, and are either not majority-owned by an entity with over $US 10,000,000 revenue which would fulfil the criteria for IDF membership eligibility in its own right, or are a subsidiary of an existing Member of the Foundation

The annual fee for Registration Agency Members is $US 35,000.

The annual fee for Affiliate Members is $US 5,000 per working group.

Membership fees fall due annually on the anniversary of joining. Eligibility for Membership may be re-evaluated at each point the Member Agreement falls due for renewal.

The benefits of membership participation flow downward to subsidiaries of members. In the case of government agencies and departments, or educational institutions, this is interpreted to include sub-agencies, departments, laboratories, etc. When a subsidiary takes advantage of this provision, participation in IDF activities must be coordinated through the Member's IDF representative. Participants so authorized will officially represent the Member organization. The subsidiary may not indicate in its publicity that it is, itself, an IDF Member. Subsidiaries are, however, permitted to join the IDF in their own right. A Member may designate an employee of one of its subsidiaries to be its IDF representative. Such might be desirable when the member wishes its name to be listed, but a subsidiary is the focal point of all IDF activity.

To join the IDF, a membership application can be found at
http://www.doi.org/membership/membership-form.html

1.5 Intellectual property notices

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, and other types of resources where they are involved in intellectual property transactions (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. For more information about persistence, see Chapter 7: Policy.

A DOI is different from commonly used internet pointers to material such as the URL -- Uniform Resource Locator, the usual means of referring to World Wide Web material -- because it identifies an object as a first-class entity, not simply the place where the object is located. A first-class entity or object in the information infrastructure is stored on one or more servers and is accessible from these servers using a globally accessible identifier (URI). An entity is referred to as first class when it represents an object, not some attribute of an object; e.g. an address is an attribute of a thing, whereas the thing itself is a first class object.

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 more 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 (e.g. the “syntax” of the number itself – see Chapter 4 and Appendix 1). The DOI system makes use of other technologies (notably the Handle System and the indecs framework) which are themselves being further standardised (see Appendices 3 and 4). However, other aspects of the DOI are still subject to rapid change and development.

2.7 Who is using the DOI System?

Several million DOIs have already been registered, by several hundred different registrant organizations. These examples come mostly from traditional print-publishing companies that have already established major online publishing programs. This reflects the origins of the DOI in the text sector.

However the fundamental design of the system is applicable to any media or content. 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.

2.8 Benefits of the DOI to Publishers, Intermediaries, and users

2.8.1 External benefits (benefits in the distribution and sales life-cycle)

- More sales. Any hyperlink anywhere on the Internet which refers to a product’s DOI now becomes an active, dynamically-controlled, updateable sales & marketing tool – because clicking on it pops up a menu of all the actions or services which the Seller wants to provide to the consumer in connection with that book: purchasing it in any format from any retailer who carries that format, getting a free excerpt as an enticement to buy, reading reviews, visiting the seller’s own website or marketing page for that product, etc.
- These choices are a single click away for the consumer, making it as easy as possible for the Publisher to enable impulse purchasing and otherwise secure incremental sales by removing all the obstacles which previously stood between a consumer discovering a product, and making the purchase.
- Changing these choices dynamically is easy and cheap for the Publisher, because in order to add another retailer, add another review, run a special discount or temporary

promotion, or make any other change in the choices the consumer sees, the Publisher simply updates one central DOI record - at which point, thousands of links all over the Internet which refer to that DOI will now reflect the new or changed menu options, even if already stored, bookmarked, or printed.

- Content is "discoverable" via the DOI – easier to find on the Internet by consumers, easier to find on the Internet by Publishers' business partners, easier to find by search engines, Web portals, secondary publishers, reviewers, aggregators, online bookstores, etc.
- Sales can be tracked more easily and more precisely (at the "object" level if desired – or as aggregated into new collections of object-level items).
- Sales are tracked more cheaply and accurately. There is far less labour, errors, and costs when all transactions are keyed to an unambiguous identifier which computers can understand, instead of people trying to match the different IDs of each pair of partners, or trying to track sales based on descriptive metadata about the item instead of a numerical identifier.
- Royalty tracking and licensing revenue can be tracked more easily, cheaply and accurately, not only reducing costs but also capturing revenue more fully. (Currently, mechanisms for this are leaky and ineffective, especially for Internet-based licensing – e.g. licensing of content through special-interest portals or aggregator websites).
- Cross-linking becomes feasible and inexpensive on a wide scale, both among the publisher's own content and between the publisher's content and other sources. This increases revenue opportunities as well as enhancing the value of the content itself through its rich hyperlinks.
- Much wider dissemination of content, partly through enhanced "discoverability" for all third parties and also through greater use of metadata for targeting the content more precisely to different audiences whom it might not otherwise reach.
- Greater leverage over distribution channels, both by reducing the costs of switching distributors (because set-up costs are minimised when everyone shares a universal ID) and by increasing the reach and breadth of the channels available (again, because set-up and ongoing administration are much simpler via the DOI).
- Additional, incremental product revenue enabled by offering highly-targeted, customized information products (but all from the same asset base) to specific audiences who might not otherwise purchase that information. Again, this becomes possible because the DOI enables object-level control over digital assets - finding them, recombining them, distributing them, tracking their usage, etc.
- Additional revenue enabled via Digital Rights Management (DRM), including superdistribution – where each customer becomes an additional point of sale instead of a potential risk for free pass-along. DRM has been a commercial disappointment to date, and will remain so until the DOI is adopted more widely. Only then will the kind of interoperability be feasible which could provide legitimate alternatives to peer-to-peer in terms of a seamless, friction-free end-user experience.

2.8.2 Internal benefits (benefits in the production life-cycle)

- Unique, unambiguous, universal content ID – to identify content objects throughout the entire content lifecycle, i.e.:
  - pre-publication (content Authoring, Aggregation, Selection, Rights Acquisition)
  - post-publication (Distribution, Syndication, Sales, Superdistribution)
  - archiving/digital asset management (for later re-use and re-purposing)
- Ability to find content assets internally – in order to facilitate re-purposing, re-combining with other assets (internally or with external partners), future editions, etc.
- Reduced costs of determining ownership, clearing rights, etc.
- Reduced costs (and greater control) of licensing-out or syndicating-out
- Reduced costs (labour and errors) of sales tracking, channel management, P&L calculation, etc.
- Greater corporate-wide leverage over digital assets which are otherwise invisible behind separate divisional content management systems keyed with separate content IDs. (Alternatively – at the publisher's choice - instead of helping consolidate different content management systems, the DOI can even facilitate greater diversity in content

management systems tailored to particular businesses, because at least the content per se is now trackable universally.)

- Relatively cheap to implement: All existing CMS/DAM systems already have an internal content ID (as do all DRM systems), so it's easy to simply utilize this field for the DOI - as a standard, universal ID. Then the internal, proprietary ID also becomes usable by distribution/sales partners and all others – just like the ISBN for physical books.
- Overall effect: reduces costs, streamlines efficiency, and increases the functional capabilities of all internal systems which manage digital assets.
- Overall result of these improvements:
    - creates a foundation for digital asset management which allows those assets to be leveraged more profitably;
    - facilitates the creation of new, additional products over that same base of assets.

2.8.3 Quantified benefits: case studies

A white paper "Enterprise Content Integration with the Digital Object Identifier: a business case for information publishers", (http://dx.doi.org/10.1220/whitepaper5.) quantifies the business benefits for information publishers of implementing the Digital Object Identifier (DOI) to facilitate internal content management and to enable faster, more scalable product development, by delivering four key advantages in making it easier and cheaper to:

1. Know what you have (users able to look at catalogues of content available throughout the enterprise);
2. Find what you want (users able to search and browse for content to be used or re-purposed);
3. Know where it exists (able to see where the item exists within the organisation);
4. Be able to get it (users and production tools able to retrieve the content).

This is illustrated by four examples of cost savings, each of which is supported by a worked actual case study:

- Cost avoidance in cross-brand product development (example case study: $120K savings for a vertical market information publisher building a new cross-brand web portal);
- Scalable product development through repurposing (example: annual incremental revenue of $700K for a periodical publisher creating books from repurposed periodical content);
- Cost reduction in existing production processes (example: 94% reduction in staff effort = $400K for a textbook publisher building web sites to accompany textbooks);
- Increasing revenue and market share through tool integration example: >$1.2M incremental revenue for a financial information publisher selling documents through third party links from investment selection tools.

2.9 The use of the term "Identifier": Numbering schemes, specifications, and identifier systems

We need to make an important terminology distinction about the use of the word "identifier". As the use of numbering in digital networks has developed, the historical use of the word in this context has become expanded to the point where it is now used synonymously to cover several different things, all of which are useful but which actually carry different implications that need to be separated in a detailed understanding of practical Digital Rights Management (DRM) applications. It's important to understand the differences here; and to note that these are not mutually exclusive (one particular "identifier" may fit into one or all of these categories).

2.9.1 Identifiers as "labels": The output of numbering schemes

A numbering scheme is a formal standard, an industry convention, or an arbitrary internal system such as a an incremented production serial number etc., to arrive at a consistent

syntax for denoting and distinguishing separate members of a class of entities. The scheme is a specification for generating a number: this resulting "number" may include alphanumeric characters, but the accepted parlance is to speak of these as numbers (e.g. ISBN = International Standard Book Number). The intent is of establishing a one-to-one correspondence between the members of a set of labels (numbers), and the members of the set counted and labeled. The product of the process is enumeration, a cardinality judgement, and assigned numbers for each cardinal member. An example would be the ISBN, where a separate ISBN is assigned to each book edition. The numbering scheme may or may not be accompanied by some apparatus – for example, a registration agency and maintenance agency for the ISO TC 46 series of identifiers.

The important point here is that the resulting number is simply a label string (a "noun"). It does not of itself create a string that is actionable in a digital or physical environment (a "verb") without further steps being taken. It may be used (and probably will be used) in databases; or it may be incorporated into another mechanism later.

The most common standard numbering schemes of interest in DRM include those standardised by ISO:

- ISBN: ISO 2108:1992 International Standard Book Numbering (ISBN)
- ISSN: ISO 3297:1998 International Standard Serial Number (ISSN)
- ISRC: ISO 3901:2001 International Standard Recording Code (ISRC)
- ISRN: ISO 10444:1997 International Standard Technical Report Number (ISRN)
- ISMN: ISO 10957:1993 International Standard Music Number (ISMN)
- ISWC: ISO 15707:2001 International Standard Musical Work Code (ISWC)
- ISAN: Draft ISO 15706: International Standard Audiovisual Number (ISAN)
- V-ISAN: Draft ISO 20925: Version Identifier for audiovisual works (V-ISAN)
- ISTC: Draft ISO 21047: International Standard Text Code (ISTC)

Whilst these ISO TC46 identifiers were originally simple numbering schemes, of late they have also begun to adopt the notion of associating some minimal structured descriptive metadata with the identifier. Also relevant are the ISO- affiliated NISO standards including:

ANSI/NISO Z39.84 The Digital Object Identifier

2.9.2 Identifiers as "infrastructure specifications": Making labels actionable

"Identifier" is also sometimes used to mean a mechanism or syntax by which any label (as defined above) can be expressed in a form suitable for use with a specific infrastructure tool. This is sometimes known as creating an "actionable identifier" – meaning that in the context of that particular piece of infrastructure, the label can now be used to perform some action: e.g. in an internet Web browser, it can be "clicked on" and some action takes place.

Of particular relevance for DRM, the set of internet specifications known as Uniform Resource Identifiers (embracing URLs and URNs) provide mechanisms for taking labels and specifying them as actionable within the internet. The same principles can apply in the physical as well as internet environment – for example by prefixing an ISBN with the EAN sequence 978 or 979, the ISBN becomes a UPC/EAN identifier expressible as a physical bar code symbol, or a radio-frequency tag, for use in the physical supply chain.

Importantly, note here that such "identifiers" do not mandate a way of creating labels, they merely accept any labels: hence if one does not have an existing numbering scheme, it will be necessary to adopt or create one in order to form URIs. A URI specification merely ensures that a label follows the rules to become actionable in an Internet environment: a specification is not an implementation, with all the other aspects that a fully functioning identifier system (see below) may require: URI may for example specify the syntax, and specify a recording registration procedure, but not create a managed environment (e.g. by which registrations are "policed"), or carry any specifications of metadata or policy (which I consider to be the hallmark of a full *identifier system*). Some identifier specifications of this form may have

limited rules or requirements for implementation: so far this is limited to the URN specification including a proposed (not implemented) mechanism for resolution. The acid test one should ask of such a specification is: *what does specifying my label in this particular form get me, in practical terms, in a specific infrastructure?*

2.9.3 Identifiers as "implemented systems": Implementing labels in an infrastructure environment

The UPC/EAN is an "identifier system" in the physical supply chain; a DOI is an "identifier system" in the digital supply chain. ISBNs for example become implemented in the physical supply chain through UPC/EAN bar codes or RFID tags. This sense of "Identifier" denotes a fully implemented identification mechanism that includes the ability to incorporate labels, conforms to an infrastructure specification, and adds to these practical tools for implementation such as registration processes, structured interoperable metadata, and a policy/governance mechanism. Such a system is necessary for practical DRM applications; since DRM deals with digital entities, structured metadata will be an essential component of such a system. The DOI is one of the better developed, with several million DOIs currently in use by several hundred organisations.

Both ISO TC 46 and URN have published suggested lists of requirements for their identifiers - the first covering "labels", the second "infrastructure specifications". These suggest that a practical *identifier system* (which builds on both concepts) for digital use (DRM) should assume a combination:

- Unique "dumb" identification: unambiguous simple identification (label assignment) of a defined piece of information; opaque strings, not hard-wired with any specific application intelligence;
- Well-formed metadata: defined namespaces and controlled values within those namespaces for each value of a metadata element, defined by inherent structure not by their function in a particular application. A means of expressing an ontology to facilitate interoperability in many different functional applications;
- Support for arbitrary levels of granularity;
- Multiple, co-existing, labeling schemes should be possible, including support of existing (legacy) schemes; groups of content owners with common interests should be able to devise their own schemes which should then be interoperable in an open framework; multiple (overlapping) identification of content must be allowable. This implies extensibility: the ability to add within a scheme a particular namespace that defines that element.
- Links to distributed metadata: dumb identifiers pointing to specific repositories for different pieces of data, relating to different functions e.g. copyright, trading, EDI; details of medium, version, format etc. conveyed as metadata;
- Distributed (cascading) administration responsibility: once below a certain level, no central agency permission needed to assign unique numbers (sub-levels assigned by the owner of the higher level);
- Policy and governance process: a management structure design for the practical operation of the identifier registration and maintenance processes.

The three uses of the word "identifier" (*label, infrastructure specification, and implementation*) can become easily confused, since one particular string can be in more than one category. But to see why we need to be precise, consider the following statement:

"For use on the Internet, an ISBN label can become a URN specification; an ISBN label can be incorporated into a DOI, which is an implemented identifier system following the URI specification."

Replacing the more precise terms in this statement by the loose unqualified synonym "identifier" results in confusion:

“an ISBN identifier can become a URN identifier; an ISBN identifier can be incorporated into a Digital Object identifier, which is an implemented URI identifier”

(true, but only on close textual analysis!).

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.

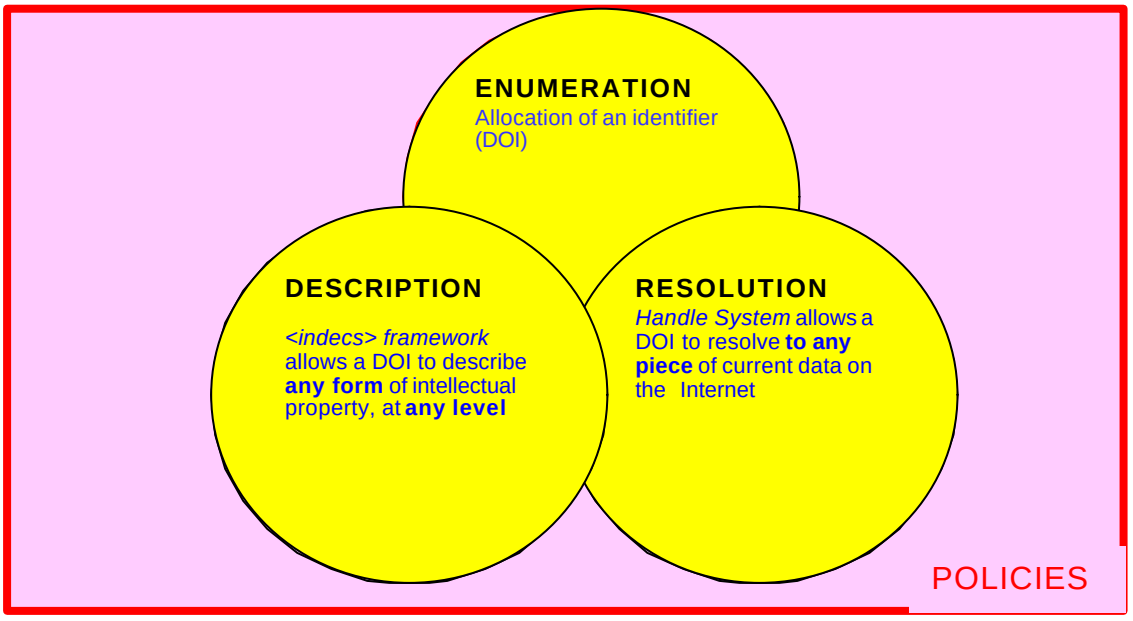
However, to understand the value of the DOI System, we must first deconstruct it. In this Handbook, we will view the DOI System as being made up of four primary components:

- **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, in a social infrastructure.



By combining a tool for naming "content objects" as first class objects in their own right with a mechanism to make these names actionable through "resolution", DOI offers persistent

managed identification for any entity. But that alone is not enough: managing resources interoperably requires appropriate metadata: creating a mechanism to provide a description of what is identified in a structured way allows services about the object to be built for any purpose. The IDF has outlined, and is actively developing in more detail, a standard way of not only doing this, but linking to existing standards such as ONIX, Dublin Core and so on, allowing each community to bring its own identifiers and descriptions into play. Finally, wrapping these tools into a social and policy framework, through the Registration Agency federation, allows the development of DOIs in a consistent quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

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 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, and 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 5). 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.

Identification requires that we understand precisely (unambiguously) what it is we are naming. And if we are to communicate and trade with others, we need to have a way of sharing that understanding – is what I call "a chapter" the same as what you mean by "a section"? This doesn't require that everyone use the same vocabulary (an impossible demand) but to accommodate the multiplying entities being traded we need an ontology ("an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them") which could describe each precisely: "Alice in Wonderland" as an abstraction, and the printed edition and so on as "manifestations" of that work in various forms. This could then provide a common analytic tool – a Rosetta Stone for the various ways in which things are described. That is now possible through further development of the <indecs> activity.

The need to describe precisely enough – unambiguously - for automation led to activities over the last few years on metadata (the "data about data") in various sectors, such as ONIX, the international standard for representing and communicating book industry product information in electronic form. During its development, ONIX – like DOI – benefited from the principles being outlined in <indecs>. ONIX has been widely successful and is now being extended into serials and audiovisual materials, and will be a central tool in developing DOI metadata in text applications, and seems likely to be one of the main metadata "languages" of the <indecs> Rosetta Stone.

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.

"Actionable" is an unfamiliar term, but familiar in practice to everyone from the actionable links of the World Wide Web: URLs (Uniform Resource Locators). The ease of assigning URLs was no doubt responsible in part for the expansion of the Web - but the fact that they are easy to create (and neglect) means they are not strong enough alone for a commercial basis. "Not found" link messages are a scourge across the Internet: the rate at which once-valid links start pointing at non-existent addresses - a process called "link rot" - is reported to be as high as one sixth of all links in six months. The fact that URLs change (technically, they are not "persistent") isn't a bad thing in itself: in fact, it is very helpful to separate names from locations – since location is only one property of (or piece of metadata about) a name which we might want to manage by the process of resolution. We want to be able to move things around – there are legitimate reasons such as change of ownership. The problem is that using URLs alone we can't track what's changed, or use one name persistently irrespective of where the item is. (Some services like Realnames and purls offer to solve part of this problem of persistence, though not the whole picture which DOI offers).

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. Even if there is no current location for a digital file, it might still be useful to know what it represented, or who owned it, or search for it elsewhere. Even if we have a location, we might want to offer other resolution results. Therefore *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.*

Using URLs alone is like using "221b Baker Street" as an identifier for an entity; when that entity moves to "Reichenbach Falls" we find it would have been more convenient to use a persistent name for the entity: "Sherlock Holmes" - the same name, at different locations. This treats Holmes, rather than his location, as what is technically called a "first class object". Managing digital objects as first class objects is much more akin to what we do in the physical world – managing books by title, not the pallets they live on in the warehouse.

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 3. 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 (or the Registration Agency he uses) 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).

3.6 DOI context: social infrastructure development

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

There is a widespread recognition of the advantages of assigning identifiers; and a widespread misconception that an abstract specification (like a URN or URI) actually delivers a working system rather than a namespace that still needs to be populated and managed. A common misperception is that one can have such a system at no cost. It is inescapable that a cost is associated with managing persistence and assigning identifiers and data to the standards needed to ensure long-term stability. This is because of the need for human intervention and support of an infrastructure. Assigning a library catalogue record, for example, will typically cost anything up to $25. Assigning an ISBN or ISSN or National Bibliography Numbers will also have costs, even if these are not paid directly by the assigner. Although a DOI is free at the point of use, there is a small fee to an assigner for creating a DOI (a few cents). This is because we have deliberately chosen to make the DOI a self-funding (though not for profit) system. Our task now is to show that DOI offers value for money as a tool which producers of information can use: CrossRef is one proven example of a registration Agency and Application Profile in text publishing; we expect to see other variants on this theme develop.

If adding a URL "costs nothing" (which itself ignores some infrastructure costs), why should assigning a name? It is indeed possible to use any string, assigned by anyone, as a name – but to be useful and reliable any name must be supported by a social as well as technical infrastructure that defines its properties and utilities. URLs for example have a clear technical infrastructure (standards for how they are made), but a very loose social infrastructure (anyone can create them, with the result that they are unreliable alone for long term

preservation use as they have no guarantee of stability let alone associated structured metadata). Product bar codes, Visa numbers, and DOIs have a tighter social (business) infrastructure, with rules and regulations, costs of maintaining and policing data – and corresponding benefits of quality and reliability (When a credit card is presented, we can be reasonably certain that the number is valid, and has been issued only after careful correlation with associated metadata by the registrant). It does not necessarily imply a centralised system – it may be a distributed system (like domain names), but it must have some form of regulation.

Such regulation of infrastructure for a community benefits all its members; funding the development of it is often a problem, and there is no “one size fits all” solution to how this should be done. But finding a workable model for the development of an infrastructure can yield obvious benefits. There are many modern examples – 3G telephone networks, railways – which are struggling with the right model for supporting a common infrastructure. The Internet was largely a creation of central (US) government; the product bar code, a creation of a commercial consortium. The IDF has chosen as its model the concept of Registration Agencies, based on market models like bar codes and Visa rather than on centralised subsidy: these Agencies effectively hold a “franchise” on the DOI: in exchange for a fee to the IDF, and a commitment to follow the ground rules of the DOI system, they are free to build their own offerings to a particular community, adding value services on top of DOI registration and charging fees for participation.

At the outset of the DOI development, a simple business model was introduced whereby a prefix assignment is purchased for a one-off fee. We are now in a process of migration to the long term aim of a wide variety of potential business models, using third part registration agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. The direct prefix purchase route is still an option, but our intention is that eventually all future DOIs will be registered through one of many Registration Agencies, each of which is empowered to offer much more flexible pricing structures. The pricing structures and business models of the Registration Agencies will not be determined by the IDF; each RA will be autonomous as to its business model, which could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc.

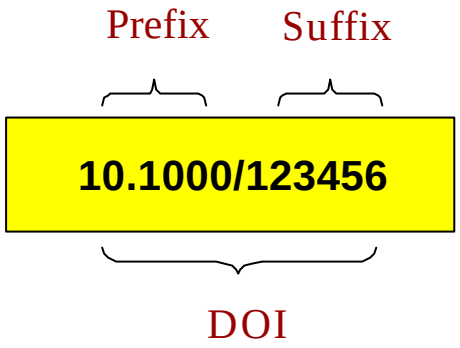
DOIs may be made available at “no charge”, if the costs of doing so can be met from elsewhere (there is no such thing as “free”, only “alternatively funded”). IDF itself is willing to allocate a DOI prefix free of charge to organizations or limited experimental non-commercial uses. For the longer term, the business model includes two separate steps: a business relationship between IDF and an RA (the "franchise fee"); and a business relationship between an RA and a DOI registrant (the "registration fee"). The two are not directly connected; this enables the RA to offer to registrants any business model whatever, which suits its needs. This could include assigning DOIs without charge. Hence DOIs can be used in both commercial and non-commercial settings, interoperably. Like any other piece of infrastructure, an identifier system (especially one which adds much value like metadata and resolution) must be paid for eventually by someone. So an organization could, if it wished, assign DOIs freely (registration fee zero to registrants) and subsidize this added-value service by paying a franchise fee to IDF from a central fund, as an acceptable cost for supporting the service.

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 Appendix 3). 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 Chapter 11).

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, for example:

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. A role of Registration Agencies is to provide a service to registrants which facilitates this. 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 Case sensitivity

DOIs are case insensitive. The Original DOI syntax (NISO Z39.84) is included in Appendix 1 and specified that DOI suffixes should be case sensitive, so that 10.123/ABC is completely

different to 10.123/AbC and the two are treated as two different identifiers. The Handle system is configurable by service so as to be either case sensitive or case insensitive and therefore allows this.

However a divergence with the written specification arose in the application due to experience: IDF decided to impose a further parameter on the specification, to remove case sensitivity which was causing confusion. All DOIs are converted to upper case upon registration, which is a common practice for making any kind of service case insensitive. The same is true with resolution. If a DOI were registered as 10.123/ABC, then 10.123/abc will resolve it and an attempt to register 10.123/AbC would be rejected with the error message that this DOI already existed.

The advantages of case sensitivity (librarian and publisher practice, human readability and expectations) were outweighed by considerations of data integrity. Case sensitivity practice across internet applications varies: DNS is not, the rest of URLs are except sometimes they aren't (this depends on the server), Unix vs PC/Mac file names (Microsoft Windows in general is not case-sensitive, Unix operating systems are always case-sensitive), markup language tags, etc. can all cause unexpected problems and one cannot guarantee that any particular piece of software will respect case sensitivity and not conflate two DOIs intended to be different. Some search engines and directories are partially case sensitive. Different web browsers may differ in case sensitive handling (Netscape have stated that “authors should not rely on case-sensitivity as a way of creating distinct identifiers, unless they are designing solely for a truly standards-compliant browser” - http://developer.netscape.com/evangelism/docs/technotes/css-class-id/)

This argued in favour of case insensitivity being the safer, and more robust, option for future evolution and development of the DOI system. The decision to do this was not easy, as the DOI syntax had already been set (though when the original standard was done, DOIs weren't being actively used in a big way and now they are). It is expected that this additional restriction parameter will be reflected in a revised version of the formal standard in due course.

As this restriction has been implemented from an early stage, we have not introduced any cases of two DOIs distinguishable only by ASCII case resolving to the same thing, as we haven't allowed that registration, even though strictly someone following the NISO specification would expect to be able to do that and would fail. We do not expect this to cause any problems.

4.6 Other prefix, suffix, and character set issues

When thinking about prefixes, suffixes and character sets, it is important to distinguish DOI from the underlying technology, the Handle System (see Appendix 3). 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.

4.7 Publishing DOIs in print

Since most publication of content is via a mix of digital and print media, there are often requirements for a DOI to be reproduced in print. A publisher might put the DOI in the document it names, and insure that the DOI appears whenever the item is downloaded or printed. It also might appear in the print version of a digital version. If the DOI is represented by a button on a Web page, the Web browser will display the full DOI at the base of the browser window when the cursor is moved over the button.

Whereas in a digital context a DOI might be assumed to be contextualized and updated (the active link it is referencing can be "wired" correctly), a print version cannot be updated or changed once released. Showing DOIs in print for e.g. journal articles tells people what an article's DOI is, but it doesn't tell people how to access it on the Web; readers will not necessarily know that the DOI is actionable. To do that, one may print the DOI in a readily recognised form such as the http proxy server (see chapter 6) URL form e.g. http://dx.doi.org/10.1002/prot.9999. There are however a couple of reasons to hesitate showing the URL form: the URL is not the article's identifier, the DOI is; and maybe the dx.doi.org form of the URL will not be the most persistent form, keeping in mind that these print copies will be around and immutable for many decades, even centuries.

In practice one can feel safe in using the dx.doi.org formulation. It should continue to work for many years even if and when it is common to use DOIs in some other formulation (for further information on persistence of the resolution system, see chapter 7). But if we are talking about centuries we will have moved beyond http:// as the most recognised route of access. So while it may be awkward, we recommend some convention of showing both the plain DOI and a way to resolve it online (a shorthand way of saying "the Digital Object Identifier (DOI) for this article is 10.1002/prot.999 and current information may be found on the web through http://dx.doi.org/10.1002/prot.999" or "…available via http://dx.doi.org/…".

e.g. DOI: 10.1002/prot.999
For current information see http://dx.doi.org/10.1002/prot.999

DOIs do not replace traditional bibliographical citations but are a very useful addition, especially if articles are published online with volume, issue, and page numbers. For example, in the CrossRef application, a citation of the Science article with a DOI would be:

Kornack, D. Rakic, P. (2001). Cell Proliferation Without Neurogenesis in Adult Primate Neocortex. Science. 294 (5549), 2127-2130, doi:10.1126/science.1065467.

A citation to a Nature article published in the Advanced Online Publication process without volume, issue or page number would be:

"Cell Biology: A cat cloned by nuclear transplantation" *Nature* AOP, Published Online: 14 February 2002, doi:10.1038/nature723.

Specific DOI implementations, such as CrossRef, may make additional recommendations appropriate to the particular applications concerned.

4.8 Use of Legacy Identifiers with DOI

4.8.1 Using existing identifiers as a DOI Suffix

An existing standard identification system number may be incorporated into a DOI, if the registrant finds it convenient to do so (it is course recommended that precisely the same entity be identified by the two systems). DOI is not alone in being a system that can incorporate existing identifiers: for example, physical bar codes can be used to express ISBNs.

For example, the prefix of a DOI might consist of the ISTC (International Standard Textual Abstraction Code number): the DOI would then identify the same entity (textual work) as the ISTC itself, with the added value of offering actionable resolution services which may be used to automate relationships (metadata); and interoperability with DOIs identifying related entities, such as manifestations of the textual work, or related textual works, even if these are not identified by ISTCs. The DOI/ISTC may then be parsed either according to the rules of the DOI system or according to the rules of the ISTC embedded within, depending upon the context.

The same mechanisms can apply equally well to the use of identifiers, which are not formal standards. For example, PII (Publisher Item Identifier) is an informal agreed standard among some publishers for simple identification of articles independent of format (it identifies articles at the level of textual abstraction, as does the forthcoming ISTC standard from ISO). For details see e.g. http://www.elsevier.nl/homepage/about/pii/. PII is used by several scientific publishers as an internal numbering system. (PII and DOI are two separate identifier systems. PII is not connected with the International DOI Foundation). So a publisher may use ISTC or PII in identifying article works.

Since any existing legacy identifier can be used within a DOI, a specific DOI implementation can create interoperability where none existed before. For example, in the CrossRef implementation of DOI, some publishers create their DOIs by incorporating PII as a suffix; others incorporate SICI as a suffix; others may in future use ISTC as a suffix, and yet others may use entirely proprietary internal production numbers as a suffix. By using DOIs, each publisher gains the benefit of interoperability of its data within the CrossRef system yet does not have to "re-number" entities which have already been assigned identifiers in another scheme.

Note that the kernel metadata for a DOI mandates the inclusion of "Identifier": "A unique identifier (e.g. from a legacy scheme) applied to the entity…it is normal to include a legacy identifier if one exists". Consideration of datasets which already include existing (legacy) identifiers shows why this requirement exists: it is so that the existing legacy scheme may be used by any automated processes which pick up structured metadata from a DOI service, using the kernel declaration of this element. Since, as we have stated earlier, DOIs are inherently opaque non-parsable strings, the legacy identifier will not be securely recovered from the DOI suffix itself (consider for example the heterogeneous collection of suffixes in the CrossRef application). Yet including the legacy identifier, additionally, as the suffix may be convenient, make the DOI more easily human readable, and be administratively desirable, even though it is not a requirement of DOI creation.

4.8.2 Example services from DOI with incorporated legacy identifiers

Relationships between entities may be expressed via metadata. For example a single chapter of a work is an *excerpt* (as expressed in ISTC metadata) of that work, and (if it needs to be identified as a work) can also have an ISTC. The precision with which one can/should express the parent/child or other relationships in the ISTC metadata is a matter for specific application (implementation) decisions based on functional granularity and other considerations. Once a specification is made of the entities, the relationship between them may be expressed as an item of metadata ("a relationship that someone claims exists between two entities").

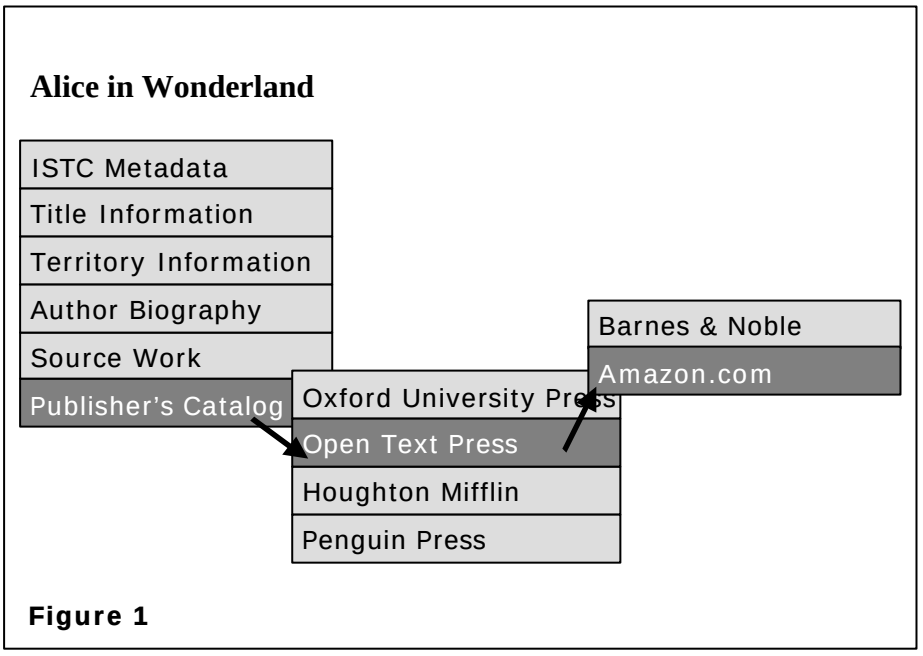


Figure 1

Figure 1 shows a possible DOI implementation of textual work identification, where the DOIs correspond to ISTCs (i.e. the ISTC may be incorporated as the DOI suffix), in this case implemented as pop-up cascaded windows in a web browser (note that DOI is not restricted to use in web environments or with windows). Since multiple resolution offers an unlimited number of possible implementation choices for data types, it is possible to express any defined relationship of an ISTC to some associated datum by means of a DOI resolution (from the DOI to the current value of the associated datum). The specific choices selected for implementation is not specified by ISTC or DOI rules, but a matter for application decisions: the options shown here are illustrative possible choices. Each entity may be shown in a user-friendly way (e.g. "Alice in Wonderland") with the associated DOI (e.g. "10.1000/ISTC0A9200212B4A1057") embedded in the application, e.g. as hyperlinks embedded in HTML pages. The following examples are shown:

- Get ISTC metadata: to return (e.g. in XML format) the complete listing of the ISTC metadata registered for the ISTC (Annex C of this standard).
- Get *title* information: to return (e.g. in plain text format) one specific element of the ISTC metadata, such as "title(s)".
- Get *territory* information: to return an element of metadata ("territory") not specified in the ISTC but provided in the application from other sources and registered with the DOI.
- Publishers catalog page(s): to resolve to a publisher's catalog, which may further offer options for selecting, purchasing etc manifestations (this may be offered as a cascading sub menu; one example is shown for illustration; in general, any of the service shown may

also be expressed a cascading submenus). The entities on this catalog page may be works or may be manifestations; whether ISTCs or not, each could have a DOI assigned.

- Author biography: to resolve to a source of information linked through the ISTC metadata key “author”.

- Source work(s): to resolve to the DOI (ISTC) specifying the source textual work of which this work is a derivation. (The inverse relationship from *source to derived* may also be articulated in a similar manner). This also illustrates that DOIs may be nested, without limit: the source work may itself be a further derivation of a “grandparent” source.

And so on; any desired relationship may be expressed providing the appropriate metadata or specification is available. Rights expressions and transactions may be expressed using metadata, and the DOI system is participating closely in the work which has led to the adoption of the <indecs> framework as the basis for the ISO MPEG Rights Data Dictionary.

DOI provides a detailed technical framework/architecture for articulating multiple resolution as defined services (one DOI may resolve to many different data types, which could represent many services: see chapter 6), comprising principles, rules, and guidance framework for consistent development of multiple resolution across applications, and as a means of building functions and services into DOI resolution.

4.8.3 Benefits of services expressed using legacy identifiers within DOIs

There are some benefits, which are common to any DOI:

- Links appear as standard hyperlinks, but unlike URLs are persistent; changes to URLs defined as resolution points may be made without affecting the DOI and any links bookmarked using it.

- Once DOIs are registered, they are available to everyone who wants to use them.

- The registrant registers the multiple-resolution options along with the DOI, and may add or modify them at any time. Links pop up on the menu dynamically, in real-time, out of the global DOI directory - so that if the registrant then added an option to the DOI record, it would show up the next time a user clicked on a link based on that DOI, in his own application or anywhere else on the web.

- Users building applications of these DOIs are free to invoke all or only some of the multiple-resolution options - e.g., a DOI may have recorded with it all the publisher/manifestations of a textual work; if a user has its own exclusive or preferred relationship with only one of the publishers, it could show only those options. Therefore the whole downstream distribution and retail chain is still empowered to strike its own relationships and use the DOI for its own purposes – the Registration Agency offers registration of the full superset of options.

- Users might encounter the DOI in another way e.g. as text in e-mail rather than on a web site; DOIs may be cut and pasted as with any other link.

- The customer can select an individual chapter, or any other component the publisher wishes to assign a DOI to.

- The DOI could go on to help with all of the back office functions, which are not visible on the surface - e.g., allowing the retailer to track sales according to a specific #, then report sales back to the publisher according to that #, then have the publisher's internal systems tracking sales, performing accounting, calculating royalties, etc. - all because there is now a universal, machine-readable # allowing cross-system communication - just like the ISBN for physical books, or the CUSIP # for Securities. And note that DOI complements, rather than replacing, those existing identifiers such as ISBN, ISTC.

- The DOI metadata schema uses as its basis the <indecs> (interoperability of data in e commerce) framework. This has been adopted as the basis for the ISO/IEC MPEG Rights Data Dictionary. Conformance with this framework will therefore facilitate the use of ISTCs with MPEG-compliant tools for multimedia content management and digital rights management.

There are some benefits specific to the incorporation of an existing standard into a DOI:

- Allows the administration and structure of the entity identification, if required, to be diverse, points of registration to be multifarious, enabling authors and publishers of all kinds to issue identifiers at the most convenient point, close to source and carrying directly or indirectly the author's authority.

- DOI's have a common, controlled “front end”, but allow complete domain-specific structures at the "back end". They can have multiple registration authorities, and can incorporate established identifiers (like ISBN, ISAN, ISWC, PII, or any proprietary ID) to allow integration with existing systems. The same textual work can appear in many different guises - a poem may appear in a book, a newspaper, a film, on radio etc – it is necessary to promote longer-term interoperability through the use of a common identification scheme and practical interoperability mechanism. Use of DOI allows ready interoperability with existing abstraction identifiers, with associated manifestation identifiers and other metadata; with rights metadata; and builds on what is practical in each sector.

4.9 DOI as a Universal Resource Identifier (URI)

The DOI, as an identifier, stands on its own. When used in a web context, such as in a web browser plug-in or http-to-handle proxy, it can be put in the form of a URI with a scheme doi: In this sense it is no different than a telephone number, for which there is a URI scheme of tel:

A specification for DOI as a URI exists as an Internet Draft: Paskin, Norman; Neylon, Eamonn; Hammond, Tony & Sun, Sam; " The “doi” URI scheme for Digital Object Identifier (DOI) <http://www.ietf.org/internet-drafts/draft-paskin-doi-uri-02.txt>. This document defines the 'doi' Uniform Resource Identifier (URI) scheme for DOIs, which allows a DOI to be referenced by a URI for Internet applications.

A "doi" URI may serve as a pure name or may be de-referenced by a network service. When used as a name, a "doi"-based URI is independent of any service protocol and accordingly, is not network de-referenceable. When used within a network reference (e.g. within a hyperlink), a DOI identifier does not have a native resolution system, but is instead transported using a network protocol to a specific service (e.g. the Handle System or a HTTP request to a proxy), which may also include supplemental query components specific to that service. Note that in current practice all DOIs are Handles, registered in and resolvable through the Handle System.

The "doi" URI scheme conforms to the generic URI syntax as specified in RFC2396. UTF-8 encoding is mandated for any DOI transmitted between "doi" user agent and any DOI service. Syntax for DOI identifier within the "doi" scheme is defined in accordance with ANSI/NISO Z39.84 standard for Digital Object Identifier Syntax.

Some examples of syntactically valid "doi" URIs are given below:

(a) doi:10.1000/182
where "10.1000" is the prefix and "182" is the suffix

(b) doi:alpha-beta/182.342-24

where "alpha-beta" is the prefix and "182.342-24" is the suffix

(c) doi:10.abc/ab/cd/ef
where "10.abc" is the prefix and "ab/cd/ef" is the suffix

(d) doi:1.23/2002/january/21/4690
where "1.23" is the prefix and "2002/january/21/4690" is the suffix

(e) <element xmlns="doi:1.23/2002/january/21/4690">

The acquisition of DOI services can be achieved through the use other protocols as a proxy to transfer to dedicated networked service components. Examples of such use are given below:

(f) http://my.resolver.inc/resolve?id=doi:10.1000%2F182
is an OpenURL service request for "doi:10.1000/182"

(g) rtsp://service.net/query?doi:10.abc%2Fab%2Fcd%2Fef
is a service request for "doi:10.abc/ab/cd/ef"

4.10 DOI as a Universal Resource Name (URN)

The DOI can also be considered a Uniform Resource Name, as originally described in W3C and IETF architecture documents. The Technical Working Group of IDF is considering whether a formal specification of DOI as URN is advised or not. At the more general level (URN as persistent name as opposed to URL as current locator) DOIs are, particularly as implemented through the handle system, URNs. At the more formal level of IETF drafts and standards, however, DOIs are not URNs as doi is not registered as a urn namespace, e.g., urn:doi:10.1000/1. We see no reason to do so at the moment, but would consider it if there were a widely deployed and efficient URN resolution system.

The URL/URN/URI debates have gone on for a number of years and are not easily summarized. The situation is somewhat clearer, however, at the implementation level. URIs are not intended to rely on any additional network services. A software client either knows what to do with, e.g., ftp, or it does not. The URN architecture assumes an additional network service that would allow a client to deal with a previously unknown URN type, e.g., urn:doi. Specifically, a DNS-based middle layer (RDS) is used to find the specific service appropriate to the given URN scheme. URN resolutions are then delegated to that scheme-specific resolution service.

The original RDS mechanism proposed was NAPTR(Name Authority Pointer); more recently a variant of this, DDDS (Dynamic Delegation Discovery System) has been proposed. These are proposed DNS extensions that would use DNS to provide a regular expression for the namespace, e.g., turn urn:hdl:10.1000/1 into http://hdl.handle.net/10.1000/. But, to the best of our knowledge, this hasn't led to anyone actually using it in a production sense. There are no “CrossRefs” - practical implementations of large scale - in the URN space. There may be identifier strings being laid down as specifications, e.g., urn:isbn:123456789, but at this point there is no apparent advantage to that over the simpler isbn:12345678. In neither case is there a readily available well known global resolution service.

We subscribe whole-heartedly to the basic notion behind URN – that persistent and actionable object names are required for coherence in the emerging digital realm. This is the very reason that the DOI was originally brought forward. But we see nothing within the more formal URN space, either existing or on the horizon, that offers a long or short term advantage. We have adopted an existing name resolution system (handle) and are using two separate methods to reference DOIs on the Internet: as URIs (doi:10.123/456) and as URLs (http://dx.doi.org/10.123/456). Each string can stand on its own, as a pure unique name, or it can be resolved using some network service, e.g., the handle system. Resolution of the URI

form would require software not yet commonly found on users' desktops. Resolution of the URL form requires a proxy or gateway service out on the network. Both approaches are in use at the moment. Adding the additional layer of the URN namespace (urn:doi:123/456) would complicate this situation and would not appear to bring sufficient added value to either the naming or resolution process to justify the effort.

4.11 DOIs, URNs, and URIs

Fundamentally and somewhat simplistically, we think persistent names are a good thing, but we remained unconvinced of the utility of the current URN route: putting a resolution system in front of resolution systems, especially the one proposed. However, the issue of URI and URN specification is not the main issue. It is important to distinguish two issues:

1. The Internet specification of "what is" a URN and a URI
2. What this means for practical implementation

Irrespective of Internet specifications, as noted earlier a DOI is more than just a HDL (and hence more than a URI or URN). Especially, policy and added metadata are important components of DOI and are what will provide added value in practical implementations.

Consider what is needed to implement a DOI: policy and metadata rules, akin to the business rules for ISBN, EAN and Visa. A good example is specifying DOI metadata through mappings (see Chapter 5). Mapping ensures semantic integrity; it is a technical exercise through which a term is assigned a unique value in a data dictionary and the given a genealogy, and other information added. This is a painstaking, though once-only, activity, which requires intellectual input and effort which as yet cannot be fully automated – efforts to do so through activities such as the Semantic Web may in time yield tools to aid this but seem unlikely to fully replace such intervention.

5 Metadata

This chapter explains why interoperable metadata is an essential component of the DOI System and outlines the basis, rules and guidelines for metadata used with DOIs. This includes the concept of DOI Kernel metadata and DOI Application Profiles. Here as elsewhere, we use the word "metadata" in its most commonly accepted meaning – data that describes something.

Sections 5.1 discuss the principles of DOI metadata; 5.2 the management and administration of DOI metadata; and section 5.3 notes on specific sector applications which may have existing metadata usage.

Section 5.4 outlines how the principles described above are turned into practice and procedure, and is supplemented by additional technical detail in Appendix 2.

5.1 Principles

Read this section for background information and explanations of the principles underlying DOI metadata. For detailed implementation practice and procedure, go to section 5.4

5.1.1 Why does the DOI require metadata?

Identifiers are simply names: names that follow a strict convention and are unique if properly applied. 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 – for example, since "ISBN" is the acronym of "International Standard Book Number", 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. 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 by design an "opaque string", the user can tell nothing about what it identifies from just looking at the DOI. This makes it 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.

5.1.2 What metadata should be associated with a DOI?

Metadata is needed because a number alone does not impart anything useful (like a telephone number without an attached name): to use the identifier we need some additional data e.g.

- what is the creation that is identified?
- does it have another identifier I might know (e.g. an ISBN?)
- does it have a name (title)?
- who are the parties responsible for its creation or publication?
- what sort of thing is it? (abstract, physical, digital or spatio-temporal),
- what is its mode?(visual, audio, etc.)
- does it belong to a particular application type (e.g. article linking)?

We cannot list "all metadata" associated with an entity (by definition impossible). We must be able to use existing metadata sets, where that makes sense for a particular user community (ONIX, SCORM, SMPTE, DC). A DOI application will use a particular set of metadata: we call this an Application Profile. To ensure interoperability of DOIs we need to:

- Provide some minimum key metadata ("DOI kernel") sufficient for recognition, answering the questions above - akin to a telephone directory entry. If kernel metadata is to be commonly accessible by applications, common format(s)/schemas must be used and registered.
- Provide a way of building application profiles beyond this, using existing metadata sets. This implies a standard vocabulary or data dictionary for mappings to/from both the kernel and the wider application sets.

5.1.3 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 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 that 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 uses the analysis of the <indecs> project of the requirement for “well formed” metadata. 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.

5.1.4 Adoption of the <indecs> metadata analysis

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 <indecs> data model underlies all DOI metadata. The same analysis underlies ONIX International, rapidly becoming widely accepted as the metadata dictionary for the publishing industry internationally. Similar developments are now occurring in other media sectors (e.g. the adoption of indecs by MPEG-21).

Fundamental requirements defined within the indecs project and used within DOI are:

- Unique identification: every entity needs to be uniquely identified within an identified namespace;
- Functional granularity: it should be possible to identify an entity when there is a reason to distinguish it;
- Designated authority: the author of metadata must be securely identified;
- Appropriate access: everyone requires access to the metadata on which they depend, and privacy and confidentiality for their own metadata from those who are not dependent on it

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 <indecs> 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 <indecs> metadata model gave 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 <indecs> analysis, as this emerges over the next year or so. The <indecs> 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. The IDF, together with EDItEUR, is responsible for managing and further developing the output of indecs, and in a Consortium to build a Rights Data Dictionary - a common dictionary or vocabulary for intellectual property rights named to enable the exchange of key information

between content industries and ecommerce trading of intellectual property rights. Work done by IDF in developing the DOI Namespace (a data dictionary for DOI use, based on indecs) was used as input to <indecs>rdd, now renamed Contecs:DD. This data dictionary has been accepted as the basis for the ISO MPEG-21 rights data dictionary: see Appendix 4.

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.

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. DOI implements the indecs dictionary, creating a mechanism to provide a description of what is identified in a structured way and allowing services about digital content objects to be built for any purpose.

The IDF guidelines provide a standard way of doing this, and hence a means of mapping to existing standards such as ONIX, Dublin Core and so on, allowing each community to bring its own identifiers and descriptions into play. Wrapping these tools into a social and policy framework, through the Registration Agency federation, allows the development of DOIs in a consistent quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

5.1.5 Metadata and automation: an overview

A DOI is a *name* (identifier) for an entity in a network environment. *Resolution* is the process of submitting an identifier [of an entity] to a network service and receiving in return one or more pieces of current information related to the identified entity. Resolution can therefore be viewed as a mechanism for maintaining a relationship between two data entities. An item of metadata is a relationship that someone claims exists between two entities (<indecs> definition): therefore, such metadata relationships between entities may be articulated and automated by resolution.

DOI resolution is intended to return complex information, or *metadata*, related to the identified entity. Complex means not simple, and more than one piece of information.

For either humans or machines to make sense of, and so act on, that returned information, they must be able to understand the meanings of the individual pieces and the meaning(s) of the grouping or pattern of the pieces.

For any sort of shared use of DOIs those individual pieces of data (metadata terms, generally in attribute/value pairs) and the patterns (metadata schemas) must be standardized, at the level of both syntax and semantics.

DOI *application profiles* are the intended mechanism for associating a given DOI to one or more of these standard patterns of terms and relationships among the terms. That is, knowing that a DOI is of a given application profile will be the key to interpreting the rest of the data, validating that it is all there, and so on. XML is the current mechanism for expressing application profiles. RDF may be of further use as it develops.

Indecs provides an *ontology* (an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them) for talking about Intellectual Property transactions and so

will inform the creation of, or simply provide, the metadata terms for articulating application profiles. Does indecs cover all envisioned DOI metadata? Yes, in that indecs is a framework for creating (any) structured metadata. However specific implementations may need to be put into place before we can claim that “we have a transformation from ONIX to Dublin Core using indecs” or "we have indecs dictionary entries for DOI music recording entities". The DOI-Namespace (a tool for such implementations) is being built from (or as part of) Contecs:DD (a conceptual framework now being populated by some real terms).

Metadata *registries* are a mechanism for recording the various schemes and associated terms, including a data dictionary or a way to reference a data dictionary.

By combining a tool for naming “content objects” as first class objects in their own right with a mechanism to make these names actionable through “resolution”, DOI offers persistent managed identification for any entity. Creating a mechanism to provide a description of what is identified in a structured way (metadata) allows services about the object to be built for any purpose. The IDF has outlined, and is actively developing in more detail, a standard way of not only doing this, but linking to existing standards such as ONIX, Dublin Core and so on, allowing each community to bring its own identifiers and descriptions into play. Formally integrating a DOI data model (an abstract model for what is returned from a DOI resolution) with the DOI-Namespace (built on the indecs data model) offers substantial benefits of synergy. Finally, wrapping these tools into a social and policy infrastructure framework, through the Registration Agency federation, allows the development of DOIs in a consistent quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

Without an ontology and structured framework, metadata terms and classifications become ultimately useless for anything other than the application the deviser had in mind, recalling the famous parable of Jorge Luis Borges: *“These ambiguities, redundances, and deficiencies recall those attributed by Dr. Franz Kuhn to a certain Chinese encyclopedia entitled Celestial Emporium of Benevolent Knowledge. On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance” ("The Analytical Language of John Wilkins").*

5.2 DOI Metadata: management and administration aspects

This section discusses the creation, management and use of DOI metadata within the DOI system.

5.2.1 DOI Application Profiles (APs) overview

5.2.1.1 Purpose of DOI-APs

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 (DOI-AP) 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 three main constituents of an AP are:

- Registration requirements
- Metadata requirements
- Available services

A DOI-AP is created to serve the requirements of a particular “community of interest” within the broader DOI Registrant community. 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. When the need for a particular AP is identified across multiple RAs, it is the intention to establish a DOI User Community (DOI-UC) as an organization that defines and implements the DOI-AP.

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.

The purpose in creating application profiles and then linking each individual DOI to one or perhaps more of them is to let you, where 'you' is probably an automated application, know what you can do with a given DOI. An application is able to recognise that “this is a Crossref Citation DOI and so it is reasonable for me to use it in a forward linking query or to offer some human user the option to do so”. That is based on structure and the ability of the application to use that structure directly or indirectly through some service which knows how to use that structure.

The application profile mechanism exists to allow controlled variability among DOIs, essentially by typing them, and that variability could include different kinds and extents of information (metadata) associated with each type of DOI as well as a variety of known services appropriate to each type of DOI. Metadata rules and guidelines are needed when you encounter DOIs of multiple types, or application profiles, so that you can map from one to the other and thus deal with the resultant information as a whole and not as two disjoint lumps of data.

Interoperability across DOI-APs is the main purpose of DOI metadata rules and guidelines. To understand even a single DOI requires that you understand the rules and guidelines, which were used to associate that DOI with its constituent metadata. Many would consider that to be the obvious reason to have rules and guidelines. It is the existence of multiple metadata schemas and views used in different application profiles that will require a framework for making sense of the world in which they co-exist.

The goal is to allow mapping of varying metadata sets to each other without doing a pair-wise mapping of each scheme to each other scheme. The core assumption is that this semantic interoperability will be required for widespread DOI uptake. The role of the DOI handle system in this context is to provide a network service that associates a given DOI with useful and current attributes of that DOI (current state data). DMS provides a framework and some specific mechanisms for creating and encoding some of those attributes: it does not however imply one-to-one mapping to distinct handle values (i.e. a DMS construct could be packed into a single DOI Handle value, or live someplace else entirely with the DOI as a pointer). This makes the interaction of DMS and multiple resolution a highly flexible and powerful tool.

5.2.1.2 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.

The following four steps are required:

1. To name the intellectual property entity type to which the DOI-AP applies. (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 i.e. more than one metadata schema for a single DOI-AP.)

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.

5.2.1.3 Interoperable metadata requirements for Application Profiles

The main purpose of the DOI Metadata Rules and Guidelines is to ensure that DOIs issued by different Registration Agencies supporting different Application Profiles can be used transparently in applications which operate with DOIs from more than one AP.

A secondary purpose is to provide RAs with guidance and resources to enable them to specify well-formed metadata, drawing on the experience of developing other APs, the IDF and the resources of the <indecs> metadata framework and Data Dictionary. This should benefit issuers and users of DOIs even in non-DOI based applications.

There are two basic types of metadata specified within an AP: Resource Metadata and Service Metadata. Resource Metadata Set comprises metadata of three types: descriptive, rights and administrative metadata (these are not clearly separated and logically parsable sets, but should be considered as useful views of common metadata). Service metadata describes the types of services covered by the AP.

5.2.1.4 Example

An illustrative example may be helpful. A CrossRef Application Profile deals specifically with the requirements of linking citations between journal articles (note that CrossRef does not identify any specific manifestation of a Journal Article but rather the abstract entity that may be expressed in many different manifestations). 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: for definitions of these terms, see Appendix 2.). 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; 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.

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

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. 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 "the Journal Article metadata schema". Specific applications, like CrossRef, would utilise only a subset of this more generalised schema. 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.

For a detailed technical analysis of this example see Appendix 2.D.

5.2.2 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 (i.e. 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, where held by IDF, is solely for the purposes of permitting look-up from a DOI to the declared metadata by any user.

DOI rules do not mandate *which* entities DOIs are assigned to: in accordance with the indecs principle of functional granularity ("it should be possible to identify an entity when there is a reason to distinguish it") this decision is entirely in the hands of the DOI registrant. Specific DOI applications (defined via application profiles as below), implemented by one or more Registration Agencies, may of course define rules for which entities are to be assigned DOIs, as part of that community activity.

5.2.3 Application Profile Registration and Governance Procedures

5.2.3.1 Approval and maintenance 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 <indecs> 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

when there is a reason to distinguish it.

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.

5.2.3.4 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). No DOI-UC are yet fully operational and rules and procedures are to be determined.

5.2.3.5 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.

5.2.3.6 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 license 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 license the use of any applications that they may have developed or to license 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.

5.2.3.7 DOI Registration in multiple DOI-APs

Conformance of DOI-AP metadata schemas with the <indecs> framework 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 (A DOI registered in more than one DOI-AP will therefore have available to it, potentially, the metadata consisting of the union of the individual APs). The practical business implications will however need to be discussed and a role of IDF may be for example to encourage of interaction of DOI-AP agencies or users.

5.2.3.8 Registration of DOIs within an AP

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.

5.2.4 Special cases – Zero-AP and Base-AP

5.2.4.1 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 (at a point to be determined) 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).

5.2.4.2 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.

5.3 DOI Metadata and specific sector applications

This section comments on some sectors which are using DOIs in conjunction with existing metadata schemes or where existing schemes are likely to be of particular relevance

5.3.1 DOI Metadata and text applications

5.3.1.1 Relation to ONIX

The ONIX (Online Information Exchange) metadata specification (http://www.editeur.org) is the international standard for representing and book, serial and video product information in electronic form. One of the key aims of ONIX is to provide a format for delivering structured data, and to that end that ONIX has a much more highly structured model for information than other descriptive metadata formats, such as Dublin Core. ONIX is of particular importance since many initial DOI applications are in the text sector, and therefore are based on ONIX or on other standards which are mapped to ONIX (e.g. SCORM (http://www.adlnet.org/Scorm/scorm.cfm), and potentially PRISM (www.prismstandard.org) etc. EDItEUR and the International DOI Foundation (IDF) are collaborating to ensure that users of ONIX and DOI will be able to easily achieve interoperability.

EDItEUR is committed to managing ONIX in such a way that any DOI Registration Agencies who choose to use it will get the support they need, and the user community can be assured that different applications of ONIX and DOIs will be consistent wherever they overlap. The International DOI Foundation is committed to ensuring that ONIX specifications can be easily used in developing DOI Applications.

EDItEUR will work closely with IDF on use of ONIX to generate DOI metadata sets as XML expressions. The current ONIX release goes only part of the way to enabling the description of some things that may have DOIs applied, such as content items (though the architecture is there to handle them): ONIX will therefore encourage anyone who wishes to do this to contact, and work with, IDF and the EDItEUR ONIX Support Team. In addition to this work, indecs consultants will review in detail IDF metadata documentation, and the DOI Data Model, which relates to service definition and how this relates to ONIX applications. IDF and ONIX are also committed to working with other metadata standards to create similar mappings.

In addition to the ONIX web site, text publishers should also consult the AAP 2001 Open eBook Standards Project (http://www.publishers.org/home/ebookstudy.htm) numbering and metadata recommendations (http://www.publishers.org/home/metadata.pdf, http://www.publishers.org/home/numbering.pdf), which provides in depth recommendations on the use of ONIX-based metadata and DOI-based identifiers for E-Books.

5.3.2 DOI Metadata and learning object applications

Learning Objects Network Inc (LON), a DOI Registration Agency, is currently working with a number of partners in a development project for Advanced Distributed Learning for the US department of Defense and others. ADL provides common standards for the application of learning technology in education and training. A mapping of the SCORM (Sharable Content Object Reference Model, which is a major output of ADL to date) to ONIX Release 2.0 has been completed by Francis Cave Digital Publishing on behalf of AAP. SCORM is a reference model that defines a Web-based learning "content model" (http://www.adlnet.org/Scorm/scorm.cfm). SCORM has an element ("general.identifier") to carry a unique, permanent identifier for a content object. A DOI is mappable to this element (and this is assumed in the Francis Cave mapping).

Central to standardisation in this area is the IEEE Learning Technology Standards Committee, now mirrored by an equivalent technical sub-committee (JTC 1/SC36) set up jointly by the International Organization for Standardization (ISO) and the International Electrotechnical Committee (IEC). The IEEE LTSC is responsible, among other things, for the Learning Object Metadata (LOM) specification on which the IMS Global Learning Consortium's Learning resource Metadata specification is based. SCORM implements the IMS specification.

In the UK, discussions have taken place between EDItEUR (ONIX) and the UK Metadata in Education Group on areas of mutual interest and prospects for future harmonisation.

The Curriculum Online project is a UK government supported project in the area of learning objects for school-level education. The project is now working on a specification to enable them to finalise a contractor to be the preferred developer of a detailed metadata scheme, which is to be based on the existing National Curriculum Scheme (http://www.nc.uk.net/metadata/index.html). IDF have been involved in discussions with the technical consultant advising on issues of identifiers and metadata for this project, who indicates that it should be possible to do appropriate mappings with DOI metadata. The preliminary draft of the Curriculum Online Metadata Scheme specifies that "Identifiers...for the learning resource could be expressed as a Digital Object Identifier and may be associated with a publisher code. It is intended that there is only one such identifier for any single learning resource, even if that resource is available for several uses or from several suppliers."

5.3.3 DOI Metadata and MPEG applications

DOI has taken an active role in introducing the concept of the digital object identifier to the MPEG-21 multimedia framework activity [http://mpeg.telecomitalialab.com/]. The indecs metadata framework, which IDF supports and recommends as a basis for well-formed structured metadata, is also a key component of the MPEG-21 framework.

The Moving Picture Experts Group (MPEG) is a working group of ISO/IEC in charge of the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard on which such products as Video CD and MP3 are based, MPEG-2, the standard on which such products as Digital Television set top boxes and DVD are based, MPEG-4, the standard for multimedia for the fixed and mobile web and MPEG-7, the standard for description and search of audio and visual content. Work on the new standard MPEG-21 "Multimedia Framework" started in June 2000. So far a Technical Report has been produced and the formal approval process has already begun for further parts of the standard. Several *Calls for Proposals* have already been issued and two *working drafts* are being developed

The MPEG-21 Multimedia Framework recognises that to achieve true end-to-end interoperability for digital exchange of content, more is needed than interoperable terminal architecture. MPEG-21's goal is to describe a 'big picture' of how different elements to build an infrastructure for the delivery and consumption of multimedia content – existing or under development – relate to each other. In setting the vision and starting the work, MPEG-21 has drawn much new blood to MPEG, including representatives from major music labels, the film

industry and technology providers; both IDF and the indecs consortium are now active participants.

The MPEG-21 world consists of *Users* that interact with *Digital Items*. A Digital Item can be anything from an elemental piece of content (a single picture, a sound track) to a complete collection of audiovisual works. What MPEG calls a "digital item" can be considered a sub set of what DOI calls a "Digital Object"; hence DOIs can be used to identify MPEG-21 Digital Items.

MPEG-21 seeks to use existing standards where possible, to facilitate their integration and to fill in gaps. Counting the MPEG-21 Technical Report as part number one, the second part of MPEG-21 will be ready in summer 2002. This is the Digital Item Declaration, a concise XML-based schema for declaring Digital Items. Arguably more ambitious is MPEG-21's third part: the Digital Item Identification and Description. This work solves the problem of uniquely identifying digital content in a global way, and giving a resolution mechanism along with the unique identification. The specification of identifier for the MPEG-21 DIID is: "Digital Items and their parts within the MPEG-21 Framework are identified by encapsulating Uniform Resource Identifiers (URIs), into the Identification DS. As noted in chapter 4, DOI is specified as a URI and therefore can be used in MPEG-21. DOIs are listed (under "Identification Systems for Digital Items and their Parts") as valid identifier systems for MPEG-21 in the DIID MPEG-21 Digital Item Identification and Description (DII&D) Committee Draft (approved), ISO/IEC JTC 1/SC 29/WG 11/N4532 (December 2001).

Further, IDF was a founding sponsor of a Consortium to develop a Rights Data Dictionary - a common dictionary or vocabulary for intellectual property rights – based on indecs. The dictionary resulting from this activity was adopted as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard. The MPEG-21 Rights Data Dictionary, based on Contecs:DD principles, will provide a key part of the architecture required to deliver interoperability between develop a digital rights management (DRM) standard systems. Since DOI has adopted the indecs approach from its outset and will implement the indecs dictionary, creating a mechanism to provide a description of what is identified in a structured way and allowing services about digital content objects to be built for any purpose, there should be easy mapping and interoperability between DOIs with structured metadata and MPEG-21 identification and description.

5.4 DOI Metadata System ("DMS")

This section outlines how the principles described above are turned into practice and procedure. The implementation of the DOI Metadata System (DMS) is in part dependent on the finalization of one of its principle tools, the first release of the <indecs> Data Dictionary (iDD), which is being developed as the basis of the international standard ISO 21000 ("MPEG-21") Part 6, Rights Data Dictionary. The structure of the iDD is on schedule to be completed by mid July; after this time the elements of the DMS XML Schema(s) and the specific procedures for mapping terms into the iDD-based DOI TermSet by Registration Authorities can be confirmed. This section is therefore work in progress.

5.4.1 Introduction

5.4.1.1 Scope and purpose

The ***DOI Metadata System ("DMS")*** describes the elements that will make up the system for providing well-formed and interoperable metadata to support the use of DOIs. The goal is to allow mapping of varying metadata sets to each other without doing a pair-wise mapping of each scheme to each other scheme. The core assumption is that this semantic interoperability will be required for widespread DOI uptake.

The DMS is a toolset to support individual Application Profiles, and to support interoperability between them. It has three functions:

1. To provide Registration Authorities (RA's) with a ***standardized format*** for metadata Declarations (of the Kernel and other metadata) from their Application Profile (AP).

2. To provide ***semantic integrity***: a way of relating the meaning of Terms in different RA's to one another by mapping through the indecs Data Dictionary (iDD).

3. To provides RA's with a ***template*** and a ***standardized dictionary*** of Terms to help them to develop their own DOI-related metadata services in an effective way.

The purposes of these three is provide for substantial *interoperability*: a way to integrate metadata from different APs (and non-DOI sources) to support the use of cross-domain tools and applications.

The DMS builds on the existing DOI metadata *Kernel* and provides a means of extending it to any degree required.

The DMS includes a set of Terms to be known as the *DOI TermSet* which in turn are a part of the *indecs Data Dictionary,* currently being developed as the Rights Data Dictionary within the proposed international standard MPEG-21 (ISO 21000).

Each RA must map its own Terms to the iDD and in that way create its own specific *TermSet* for an AP.

There will be an XML Schema for Declarations of Kernel metadata. There will also be an XML Schema (or Schemas) for standardized Metadata Declarations of any scope or complexity based on the structure of the iDD.

5.4.1.2 Obligations of Registration Authorities

DMS is designed to be a benefit, not a burden. Compliance with the DMS should not restrict RA's operational independence or place onerous administrative tasks upon them.

There are two mandatory DMS requirements for an RA:

1. Declarations of the DOI metadata Kernel must be made in the DMS format. This can be enabled through a one-off mapping to an RA's existing database or schema, and subsequently automated. This is not an exclusive requirement: RA's may also provide metadata in any other format according to the needs of their APs.

2. All Terms used in an AP's metadata (the *AP TermSet*) must be mapped to the iDD.

These two requirements are essential for the infrastructure of semantic interoperability. More extensive use of the DMS by a particular RA will depend upon any benefits it offers them.

The DMS is *complementary* to ONIX and other existing metadata schemes, and does not aim to replace them. RAs are free to use other schemes (the use of ONIX in particular is encouraged for RA's in appropriate sectors). However, the scope of ONIX and other schemes is focused on particular kinds of resources and massages, and will not meet the metadata needs of all DOI APs. The DMS is being developed as a more *generalized* scheme, to which others including ONIX can be fully mapped via the iDD.

5.4.1.3 The <indecs> Data Dictionary (iDD)

All Terms in the DMS, and all Terms used in a AP's metadata, will be mapped into in the <indecs> Data Dictionary ("iDD"). The iDD is being developed to support the MPEG21 Rights

Data Dictionary standard, and to provide interoperability between different schemes and dictionaries describing Intellectual Property Resources.

The iDD will provide the focal point for the interoperability of DOI metadata. All Terms used in public DOI Metadata Declarations will be mapped into the iDD, creating a network of equivalences and other relationships which can support different metadata functions, including the transformation of metadata from the Terms of one AP to another and the use of Terms from different APs together in cross domain applications (for example, in Kernel Metadata Declarations).

More information about the iDD is found in Appendix 2.

5.4.2 Metadata Declarations

Metadata Declarations are collections of metadata about a DOI-identified Resource which are made available by an RA to some group of users. That user group may be the general public, or may be restricted to specific parties (for example, subscribers to a service). The DMS places no restrictions of any kind on the type and scope of such Declarations, except in respect of the Kernel Declaration.

5.4.2.1 Kernel Metadata Declaration

Each Resource identified with a DOI must have a supporting ***Declaration*** which includes the elements of the DOI Kernel

Table 1: DOI Metadata Kernel elements

Kernel Element	***Description***	***Number***	***Structure / Allowed values***
DOIApplicationProfile	Name of the AP ("a class of entities with common attributes") under which the DOI is issued.	At least 1	Registered AP name
DOI	The DOI that identifies the entity.	1 only	Any DOI.
Identifier	A unique identifier other than the DOI applied to the entity (for example, from a legacy scheme).	At least 1 (if one exists in reality)	Any alphanumeric string with its associated identifier type (eg ISBN, ISSN, UPC, PublisherProductNumber). May also be a DOI from another AP.
Title	A name by which the entity is known.	At least 1 (if one exists in reality)	Any alphanumeric string.
Structural Type	The primary structural type of the entity.	1	Allowed values: as Categories (or Qualities): Tangible Fixation (Physical) Intangible Fixation (Digital) Performance (SpatioTemporal) Abstraction (Abstract)
Mode	The primary sensory mode(s) in which the entity is intended to be perceived.	1-n	Allowed values : Visual, Audio, Audiovisual, Abstract
PrimaryAgent(s)	The name or identifier of the primary agent(s) responsible for the entity, normally but not necessarily its creator(s).	1-n	Identifier or name from a declared namespace. The specification of what constitutes a primary agent for any given AP must be declared, but is at the discretion of an RA.
PrimaryAgentRole(s)	The role(s) played by the	1-n	Controlled value code or

	primary agent.		name from an agreed namespace (eg Author, Publisher, Editor).

These elements are to be publicly available to provide users and applications with a basic and unambiguous description of the entity identified with a DOI.

5.4.2.2 Other Types of Declaration

The DMS can be used to make Declarations of any extent or scope. The Kernel Declaration is the only one that is mandatory. The following are three other examples of Declaration which may be made:

5.4.2.2.1 Extended Metadata Declaration

An RA may require Declaration of other elements of metadata beyond the Kernel to enable disambiguation, or to support other applications served by an AP. For example, the Crossref AP requires details of the ***Year***, ***Volume*** and ***Issue*** of the publication in which the identified article appeared to be made available to users under certain circumstances. The Kernel may be extended for this purpose, and there is no restriction on the scope of such extensions. There is no obligation on any RA to produce an Extended Metadata Declaration.

5.4.2.2.2 Customized Declaration

An RA may wish to make one or more “customized” declarations, using the DMS but restricting the specific metadata carried by any given format. For example, a custom Declaration may concentrate on rights metadata, or limit descriptions only to summary levels and exclude content listings.

5.4.2.2.3 Restricted Declaration

An RA may wish to issue metadata Declarations to restricted groups, such as subscribers to a specific service.

5.4.2.2.4 Transformed Declaration

An RA may wish to issue metadata Declarations which can be transformed into other Schemas. For example, in conjunction with the Adobe DOI plug-in, an RA may require that a Declaration can be expressed according to Adobe’s own metadata schema.

5.4.2.3 DMS and metadata gathering

Note that the DMS is concerned with the *output* of metadata from Registration Authorities. While the iDD and the DMS Schema(s) may be used by an RA for gathering metadata from registrants, this is not its prescribed purpose. RA’s are of course free to gather and assemble metadata by any means they wish.

5.4.3 Components of the DMS

The DMS structure is explained in this document in natural language logic and presentation for general reading. An XML Schema covering the Kernel only is also included. An XML Schema covering the full DMS structure is to be added.

The DMS is designed as a simple, flexible and comprehensive structure for metadata which can be stored in a database or represented in XML or other messages and schemas. Its elements have been developed from the <indecs> framework metadata analysis, with significant input from many other specifications including ONIX and Crossref.

However, the DMS is not just a framework or reference model: it includes all the Terms used in DOI metadata, and standard XML schema(s) in which they can be expressed.

The DMS is designed to describe ***Resources***: that is, creations which are identified with a DOI. However, the attribute structure can be easily adapted to provide a framework for other kinds of entity.

More extensive explanation of aspects of the DMS structure will be found in Appendix 2 A-D.

5.4.3.1 DMS Tools

The three principal Tools of the DMS will be:

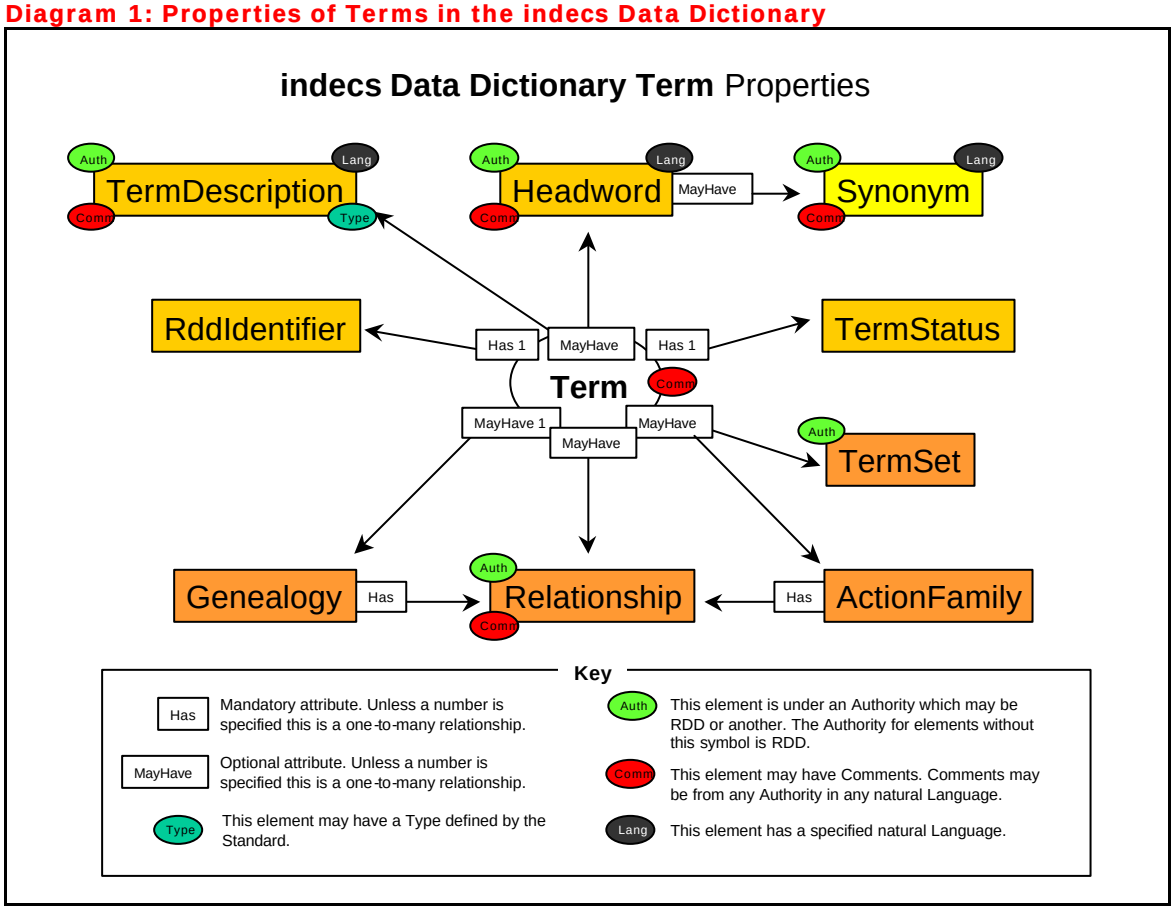
(a) the <indecs> Data Dictionary
(b) the DOI TermSet
(c) the Metadata Kernel XML Schemas, and further Schema(s).

The first draft of (c) is included in this document and the first versions of (a) and (b) will be available in July in line with the MPEG schedule described above. RAs are not expected to implement DMS until these tools are operational.

5.4.3.2 DMS Terms

All Terms used by DOI and its APs will be included as Terms in the <indecs> Data Dictionary which supports Headwords, Definitions, Synonyms and other attributes as illustrated in Diagram 1. A full account of these can be found in Appendix 2.E.

Diagram 1: Properties of Terms in the indecs Data Dictionary



5.4.3.3 Basic attributes

The DMS recognizes ten basic attributes of a Resource:

Table 2: Basic attributes

	Attribute Type	Description	Structure of basic Attribute Value
		Dependent Attributes	
1	**Identifier**	A unique label which makes the entity referable.	Uncontrolled alphanumeric string. Structure may be governed by rules.
2	**Name**	A non-unique label which makes the entity referable.	Uncontrolled alphanumeric string. Structure may be governed by rules.
3	**Quantity**	A numeric quantification of the entity.	Number.
4	**Annotation**	A textual description of, or note or comment about the entity.	Uncontrolled text string.
5	**Category**	A classification to which the entity belongs.	Controlled or uncontrolled string.
6	**Context**	An event or situation affecting the entity.	Identifier or Name of a related Context.
7	**Agent**	An entity acting in relation to the entity.	Identifier or Name of a related Agent.
8	**Relative**	Another entity with some relationship with the entity.	Identifier or Name of a related entity.
9	**Time**	A Time in relation to which something happened to the entity.	Identifier or Name of a related Time.
10	**Place**	A Place in relation to which something happened to the entity.	Identifier or Name of a related Place.

The basic attributes may be simply illustrated like this:

Resource

HasIdentifier=[Value]
HasName=[Value]
HasMeasure=[Value]
HasAnnotation=[Value]
HasCategory=[Value]
HasContext=[Value]
HasAgent=[Value]
HasRelative=[Value]
HasTime=[Value]
HasPlace=[Value]

The DMS is wholly based on types and combinations of the ten basic attributes. Often it provides two or three alternative ways of showing the same attribute.

Using these ten attributes the DMS is able to build up Metadata Declarations, starting with the Kernel and going to any level of complexity within a relatively simple common framework. One of the benefits of structuring a Declaration according to the ten basic attributes is that each has consistent behaviour.

More detailed descriptions and constraints on the basic attributes are shown in Appendix 2.A.

5.4.3.4 DMS Kernel Declaration

The illustration below (Figure 1) shows the Kernel expressed using DMS Terms and attributes. It is shown here in a non-XML format for the general reader. An XML Schema for this Declaration is provided in Appendix 2.E. Only one of the Kernel elements (***Identifier***) is a basic attribute in itself: the other elements are Types or Categories of basic attributes.

The development of this figure is explained in detail in Appendix 2.B.

Figure 1: DMS expression of Kernel Metadata

Resource

Has**Identifier** = [Value]
Has**IdentifierType** = [Value]
HasAllowedValueTermSet = [Value]
Has**DOI** = [Value]
Has**Title** = [Value]
Has**DOIApplicationProfile**
Has**Identifier** = [Value]
Has**IdentifierType** = **DOI**
Has**StructuralType** = [AllowedValue]
HasAllowedValue = **PhysicalFixation**
HasAllowedValue = **DigitalFixation**
HasAllowedValue = **Performance**
HasAllowedValue = **Abstraction**
Has**Mode** = [Value]
HasAllowedValue = **Audio**
HasAllowedValue = **Visual**
HasAllowedValue = **Audiovisual**
HasAllowedValue = **Abstract**
Has**PrimaryAgent**
Has**Name** = [Value]
Has**NameType** = [AllowedValue]
HasAllowedValueTermSet = [Value]
Has**Identifier** = [Value]
Has**IdentifierType** = [AllowedValue]
HasAllowedValueTermSet = [Value]
Has**AgentRole** = [ControlledValue]
HasAllowedValueTermSet = [Value]

5.4.3.5 DMS Declarations: principal features

DMS Terms and Declarations are developed through applying a number of basic processes to the ten basic attributes of a Term. These are summarized here, and expanded and illustrated in Appendix 2.B.

5.4.3.5.1 Types *of basic attributes (see Appendix 2.B)*

Basic attributes are developed through ***Types*** or subtypes: by specializing some aspect of an attribute to give it a more precise meaning. For example, a ***DOI*** is a Type of Identifier, a

Title is a Type of Name, ***Height*** is a Type of Measure, ***Description*** is a Type of Annotation, ***Creator*** is a Type of Agent, and so on. Types can be specialized to any level.

"Typing" something by precisely varying one or more of its attributes is one of the most important tools of the iDD for defining new Terms, explained in more detail Appendix 2.E. It provides one the main ways in which meaning is transferred after mapping.

5.4.3.5.2 Simplifying Declarations by substituting Types (see Appendix 2.B)

Types allow Terms and Declarations to be built on complex hierarchies (or "trees"). Semantic inheritance rules allow for Types to be substituted for their parent Terms as far up the "tree" as possible, thus "cutting out the middle man" and simplifying Declarations radically. The Crossref mappings in Appendix 2.D illustrate this characteristic.

5.4.3.5.3 Attributes of attributes (see Appendix 2.B)

Each of the ten basic attributes may have attributes drawn from the same list of ten. For example, a *Category* may have a *Name*; an *Agent* may have a *Category*; an *Annotation* may have an *Agent;* or a *Measure* may have a related *Time.*

5.4.3.5.4 Allowed Values (see Appendix 2.B)

Some attribute values are drawn from sets of *allowed* or *controlled* values or codes. This is principally true for Categories, but is not uncommon for Places and Agents. Allowed values (often referred to as *code lists)* are valuable for interoperability as they may be drawn from widely used standards (such as ISO Currency Codes), or they may be mappable to other code lists through the iDD. As far as possible, Category values used in DOI AP Declarations should be controlled.

5.4.3.5.5 Composite Attributes (see Appendix 2.B)

Attributes are often represented by two or more elements in a structured group, known in DMS (as in ONIX) as a ***Composite***. Standard composites for each of the ten basic attributes are provided in the DMS (See Appendix 2.A). RAs may define their own Composites which can be mapped to DMS like any other Term (for example, the Crossref AP has Composites for ***doi_data***, ***article***, ***journal*** and ***author*** among others). CompositeAttributes may also contain other CompositeAttributes.

5.4.3.5.6 Context Attributes (see Appendix 2.B)

The use of the sixth basic attribute ***Context*** (that is, an Event or Situation) provides a structure for the most complex of mappings, attributions and transformations.

Through the combination of these features and others implicit in the iDD, the DMS can build Terms and Declarations of any complexity.

5.4.3.6 The DOI TermSet

A TermSet is a group of Terms defined and mapped in the iDD. The DOI TermSet describes what has been referred to previously as the ***DOI Namespace (DOI-NS).***

5.4.3.6.1 Content of the DOI TermSet

The *DOI TermSet* comprises four groups of Terms:

- The set of Terms needed to support the ***Kernel*** (Table 1 above).
- The ***basic attributes*** (Table 2 above).

- The ***Types*** of basic attributes needed to support the mapping of all AP Terms (see 5.4.4, below).
- Certain ***standard Terms*** recommended for use by all APs.

The last group includes

(a) ISO 3166 Territory Codes.
(b) ISO 4217 Currency Codes.
(c) ISO 639 Language Codes.
(d) ISO 8601 standard formats for descriptions of Time.
(e) UCUM Unit Of Measure Codes.

The addition of other **standard Terms** should be agreed through consultation with the appropriate IDF group(s). However, it is proposed that wherever a Term is, or is likely to be, relevant to two or more APs, it should become a member of the DOI TermSet.

The first version of the initial DOI TermSet, based on the Foundation Terms of the iDD and the mappings to Crossref and ONIX, will be available by July, to coincide with the first published version of the iDD Foundation Terms at the MPEG meeting. Appendix 2.C is a place holder for this.

5.4.3.6.2 Functions of the DOI TermSet

Two principal functions of the DOI TermSet are expected to be:

(a) to provide the base Dictionary for inclusion in cross-domain XML DTDs and/or Schemas;
(b) to provide a start point for AP's developing new metadata schemes (see 5.4.4.4).

5.4.3.7 AP TermSets

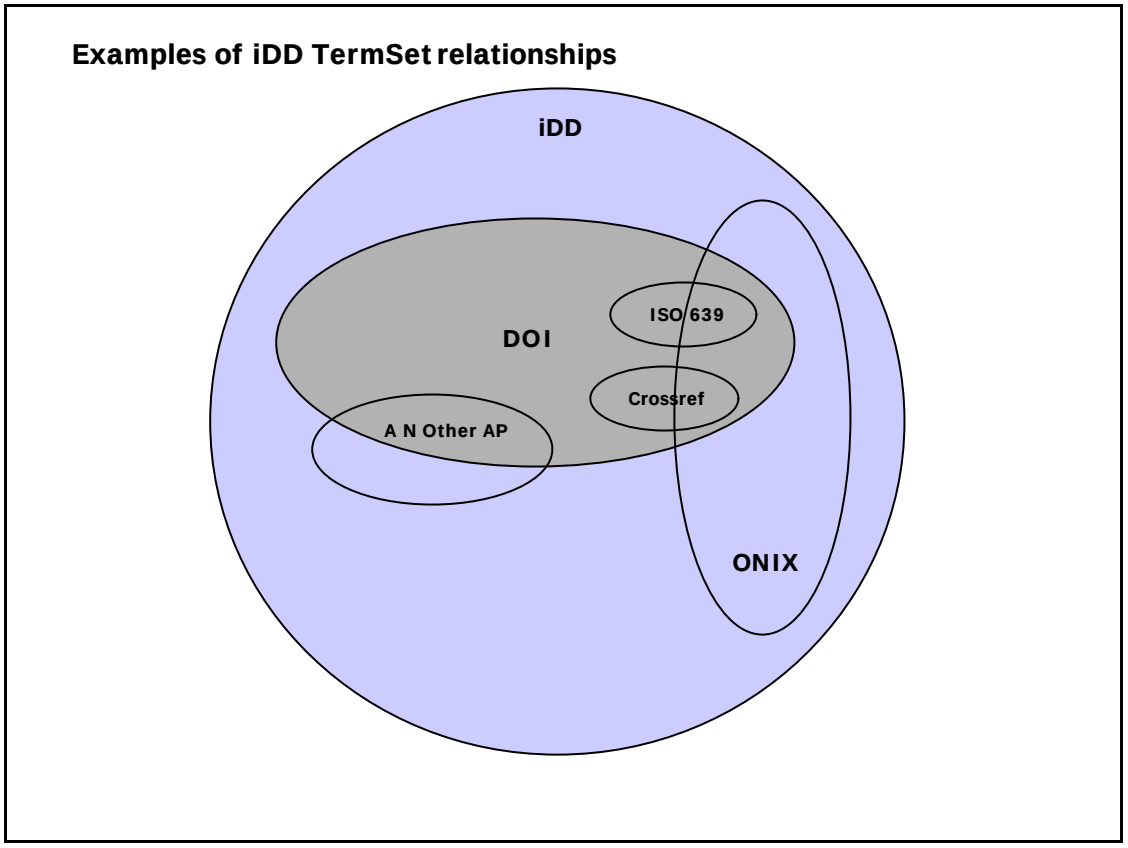
All of the Terms used by an AP in its public Metadata Declaration(s) must be mapped into the iDD (see 4, below). These Terms form the ***AP TermSet***. Two principal functions of an AP TermSet are expected to be:

(a) to provide the Dictionary for XML Schema Extensions supporting the AP (either alone or with other APs); and
(b) to provide the basis for scheme-to-scheme mapping.

5.4.3.8 How TermSets relate

TermSets are simply groups of Terms which are used by particular people or communities ("Authorities" in iDD terminology). Within DOI, these Authorities are the APs. Diagram 2 illustrates how within iDD different TermSets can be defined by any Authority, and may overlap with one another. Different Authorities may give different names for the same Terms in different Languages but this has no effect on the integrity of the iDD. Any number of TermSets can be defined for any number of reasons.

Diagram 2: Examples of TermSet relationships (shapes are arbitrary for illustration only).



5.4.4 Mapping

All Terms used by APs must be mapped into the <indecs> Data Dictionary. This mapping establishes the relationship between a Term and all other Terms used by APs, and is the way in which *semantic integrity* is achieved.

5.4.4.1 The mapping process

Mapping is a technical exercise in which:

(a) A Term is assigned a unique value within the iDD;
(b) the Term is given a ***Genealogy*** and in some cases a ***ContextDescription*** which together allow its meaning to be defined as precisely as possible using other DMS Terms. The Genealogy is a set of relationships that determine how the Term inherits meaning from equivalent, parent or other terms. The ContextDescription shows how the Term's meaning is specifically constrained by the context within which it occurs; and
(c) other supporting data is entered into the iDD in accordance with its structure as shown in Diagram 1.

The unusual aspect of mapping to the iDD is that a mapped term ***becomes a part of the Dictionary.*** The iDD structure is capable of recognizing any number of contextual meanings, and as new ones are identified in the course of mapping, they are placed in their appropriate place in the dictionary and the Genealogies and ContextDescriptions that hold them there.

Mapping is a ***consensual*** exercise. It requires agreement between the organization responsible for managing the AP and the organization responsible for managing the iDD that a given mapping is a correct interpretation of the meaning of a Term. This consent is registered on a Term-by-Term basis in the iDD.

Mapping can be described as a process of ***de-constructing*** the meaning of a Term and re-constructing it according to the iDD model. For example, a Term such as ***first_page*** in the Crossref AP refers to *the number of the first page on which a fixation of the article which is identified by the DOI appears in an issue of a journal.* The iDD allows for all of these constraints and conditions to be accurately described in common Terms in a standardized logic, providing the possibilities for integration between this term and, say, a Term from another scheme which identifies the Publisher of the Journal. An illustration of the mapping of Crossref metadata is given in Appendix 2.D.

When a Term is mapped to iDD, it may require several other ***bridging*** Terms to be created to "fill in the gaps" to existing iDD Terms. Examples of this are shown (in grey font) in the mapping of the Crossref metadata in Appendix 2.D. These Terms can be ignored by APs, but they provide essential links in the mapping chains within the iDD.

This is a painstaking process, but it is typically a ***once-off*** for each Term or scheme, with subsequent maintenance required only when new Terms are added, or amendments made. Mechanisms for modifying mappings, adding and deleting new Terms are provided for by the iDD, although of course the consequences of such changes can be serious.

Mappings are concerned with ***meanings***, not ***names.*** Terms can have different names in different APs, and the same word can mean different things in different APs. The DMS does not mandate the use of any standardized vocabulary outside of the Kernel metadata. All relationships within the DMS and iDD are described with unique identifiers for each Term.

Mappings further explained in Appendix 2.E.

5.4.4.2 Benefits of mapping through DMS/iDD

The level of granularity described above is unnecessary if only two or three schemes are being mapped. However, the fundamental assumption underlying the DOI Metadata System is that in time there will be many APs whose metadata requires integrating at various levels, whether simply at the Kernel level or to support more complex searching and processing.

Semantic integrity on such a scale appears unachievable without a central tool such as the iDD, for two simple reasons.

First, precise mapping depends upon at least one of the mapped schemes having a rich underlying model in which to precisely locate the others' terms. In general, schemes adopt data models which are tailored to meet their own particular requirements, and these are normally not rich enough to support unambiguous mapping. For example, a trial mapping recently between ONIX and a major metadata scheme from the educational world showed many approximate, unresolvable and ambiguous relationships. This is not because of any failings of either scheme, but because they were moving out of their original scope. Of course some one-to-one mappings can be very successful, if the schemes are well designed and operating in similar domains, but even here it is rarely adequate to support generalized automated processing. The mapping of ONIX and Crossref included in the current version of the Crossref XML Schema is strong but is still only a guide for developers who must provide their own contextualizations

The iDD is designed *for the purpose of* supporting unambiguous, contextual mapping: that is its primary job.

Secondly, the more schemes come into play, the more one-to-one mappings will be required, each of which is costly in resources and likely to be less than adequate for the reasons just given. With the rapid growth of metadata schemes this is becoming an increasing problem.

Diagram 3 below shows what happens if six schemes need to map to each other. Each scheme must do five one-to-one mappings: a total of fifteen mappings, probably with very mixed results. Any further scheme which joins this community then has to be mapped to each other

scheme; and the task grows by arithmetic progression. When there are N schemes, there are (N/2)x(N-1) one-to-one mappings needed.

Diagram 3: Many-to-many mapping

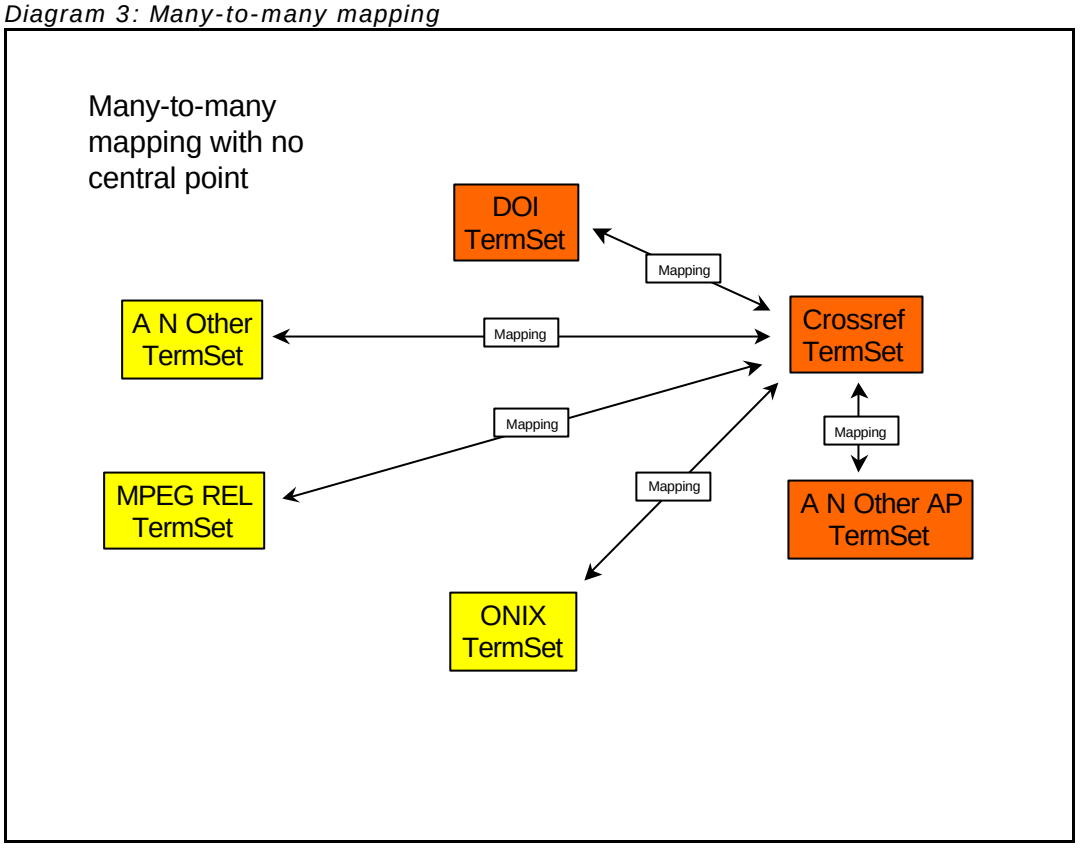
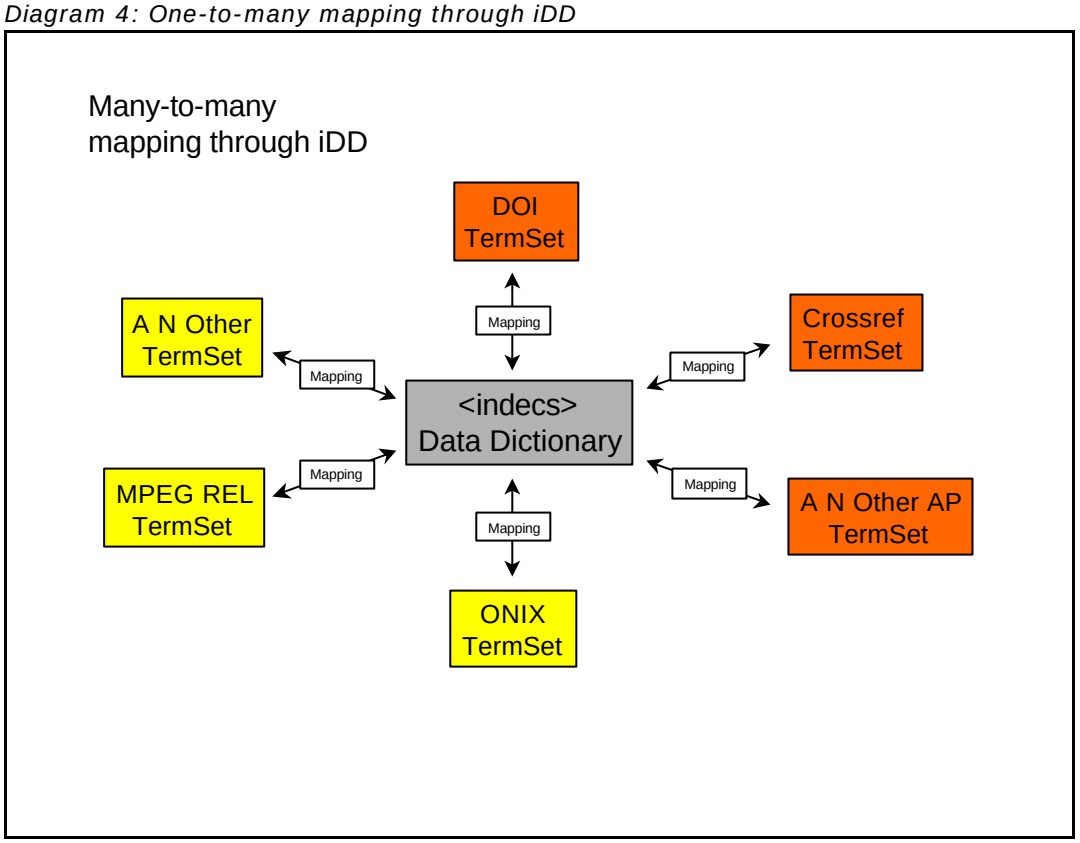


Diagram 4 shows the same mappings carried out through a central point. Each scheme requires mapping once (N schemes require N mappings) and thereafter it should be possible to create any required one-to-one mappings making use of the iDD ContextModel structure.

Diagram 4: One-to-many mapping through iDD



There are two important health warnings to make on this model.

First, iDD cannot produce unambiguous or precise mappings *if the Terms used by an AP are themselves ambiguous or imprecise.* iDD can accurately describe the ambiguity and leave the resolution to users. What iDD should be able to achieve is accurate mapping *as far as the source data allows,* producing considerably better results than a host of many-to-many mappings based on more limited models and varying techniques.

Secondly, the iDD contains the logic and data to support many kinds of processing, such as data transformations or the creation of scheme-to-scheme maps, but these will require the development of application software and business processes. Contextual mappings provide one of the necessary bases for semantic interoperability, but do not provide everything.

5.4.4.3 "Parallel Declarations"

One output of the mapping process is expected to be the development of specific maps to allow an RA to produce, with minimum effort, a standardized DMS version of a Declaration which is already being produced in another format, such as ONIX or a proprietary DTD or Schema.

5.4.4.4 Designing an AP TermSet

Up to this point this section has focussed on the interoperability of existing TermSets from existing APs. However, as RAs start to become operational, in many cases there will be no

established TermSet to work from, or partial and legacy TermSets which require substantial development.

The consensual mapping process can then be turned into a consensual ***design*** process to help RA's develop their own DOI Service tools. Existing DMS Declarations can serve as models, with Terms interchanged from the iDD as required, and new Terms developed following the best semantic practices. *An experiment with one new RA is to be carried out.*

A "template" for an Extended Metadata Declaration can be produced from the initial DOI TermSet to provide a starting point of reference for new APs.

As has been stated, there is no obligation on any RA to use DMS Terms or Declarations beyond the Kernel.

5.4.5 DMS XML Schemas

DMS Terms and Relationships are format-neutral and can be stored and expressed in any medium. However, XML Schemas are expected to be the principle means by which messages containing DMS Declarations will be communicated.

5.4.5.1 Kernel Declaration XML Schema

A first draft of an XML Schema for the Kernel Declaration is given in Appendix 2.E.

5.4.5.3 Generalized DMS XML Schema(s)

There will be one or more Schema(s) for other types of Declarations, based on a General XML Schema incorporating the ten basic DMS attributes and other DMS features described above. The structure of the Schema will draw heavily on the exemplary XML Schema being produced as part of the RDD Standard, and awaits that development. Further analysis or examples of required functionality are also needed, and the following issues need to be resolved in the development process:

(a) whether a single generalized Schema will support all Resource Description declarations;
(b) whether such a Schema will incorporate the Kernel Schema; and
(c) whether specific Schemas for specific APs or cross-domain applications can and should be produced as extensions or subsets of the Generalized DMS Schema.

5.4.6 Ongoing support for DMS

The ongoing development of the DMS is dependent on the maintenance of the iDD and the DOI TermSet once it is established. IDF, through its membership of the <indecs> Consortium, is playing an active role in discussions of this issue.

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?

A DOI is a name (identifier) for an entity in a network environment. Entities identified by a DOI may be of any form, including abstractions (e.g. as identified by ISTC). Resolution is the process of submitting an identifier [of an entity] to a network service and receiving in return one or more pieces of current information related to the identified entity. 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 DOI, using the Handle System as a reference implementation, the resolution is from a *DOI*, e.g., 10.1000/140, to one or more [hence "multiple"] *pieces of typed data*: e.g. URLs representing instances of (manifestations of) the object, or services such as e mail, or one or more items of metadata. Resolution can be considered as a mechanism for maintaining a relationship between two data entities; an item of metadata is a relationship that someone claims exists between two entities: therefore, such metadata relationships between entities may be articulated and automated by resolution.

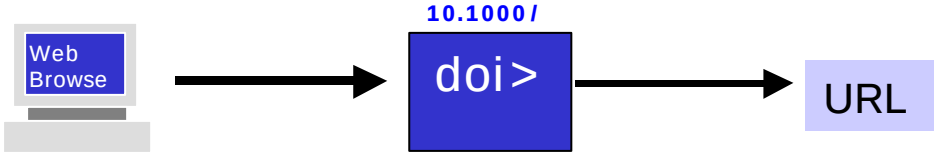
Using multiple resolution, a DOI can be resolved to an arbitrary number of different associated values: multiple URLs, other DOIs, or other data types representing items of metadata. Resolution requests may return all associated values of current information, or all values of one data type; these returned values might then be further processed in a specific "client" software application. At its simplest, the user may be provided with a list of options; more sophisticated automated processes would allow for the automated choice of an appropriate value for further processing.

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.

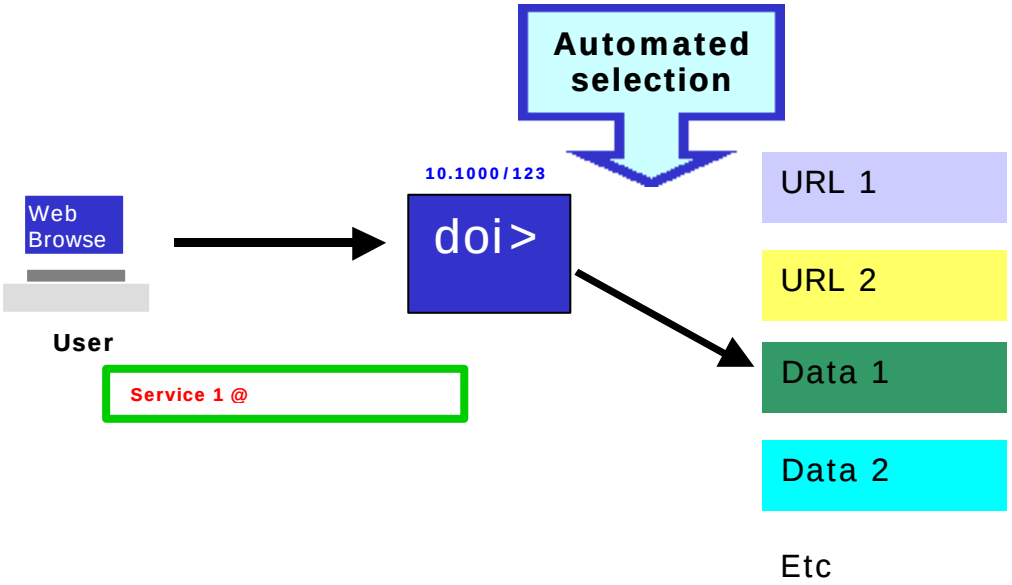
However, this lack of persistence of the URL is only the first – and the simplest – challenge that the DOI System was designed to manage.

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, and 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

6.4.1 Overview

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 3 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.

For further information on the Handle system, we also recommend consulting the General and Technical FAQs about Handle at http://www.handle.net/faq.html.

6.4.2 Technical support of DOI Resolution

CNRI continue to provide technical and operational support for the DOI system as a contractor. The following is a summary of the relevant technical and operational support which IDF guarantees through such contracts, and in turn is able to offer as a basis for Registration Agency operations. Further details of the relevant Agreement for Technical Services are available to potential and current Registration Agencies.

1. Technical and Operational Oversight

    - Monitor and advise on core DOI technical infrastructure configuration and performance.
    - Monitor research and commercial computing and networking developments that could enhance or limit the use of DOI.
    - Recommend, help initiate, and guide new prototypes to evolve the utility and expansion of DOI usage, excluding production-level support when prototypes transition into operational mode.
    - Represent the DOI in technical meetings and fora.
    - Manage and/or make significant contributions to technical working groups in the formation of IDF technical policy and the creation of IDF technical resources, such as specifications, code or demonstrations, as required.
    - Contribute to DOI technical documentation and provide editorial support as required, including technical and general editing of articles, handbooks and the preparation of briefing materials.

2. Management of DOI Directory Infrastructure

- Provision of 24x7 commercial hosting and secure facilities, back-up and disaster recovery, etc.
- Operational Management of the DOI Service (DOI Primary server (DOI Local Handle Service); Mirror (secondary) of the DOI Primary server; two proxy servers (dx.doi.org); related management tools)
- DOI Directory Custom Services

3. IDF Central Registration Agency (Directory Manager)

- Prefix administration/creation and DOI registration to requesters who are associated directly with the IDF and not with an IDF RA.
- Create and allocate Prefixes to IDF RAs. This includes discussing and implementing the process by which their requests will be submitted and how the Prefixes and related information will be provided to the IDF RAs.
- Configure the Prefixes so the Handle System knows which service is responsible for the DOIs under the Prefix (“homing”).
- Technical guidance and general assistance for Prefix owners, including guidance on use of administrative forms and batch loading, as well as status reports on their DOI deposits.
- administration tools to manage Prefix inventory
- statistical analysis reporting
- Technical Consultation

4. Analysis and Communications Services

- DOI web site and mailing list infrastructures

6.4.3 Software support for use of DOIs

There are a few programs distributed with the Handle System that users and programmers may find useful. The software section at www.handle.net describes the packages formally available to date. These include:

net.handle.batch.DOIBatch

A batch loader for DOIs.

net.handle.apps.admin_servlets

The servlets used for admin.doi.org. Useful if you'd like to allow DOI administration from a local web server.

net.handle.apps.simple

If you do decide to roll your own handle software, this package has a number of examples of how to use the handle library.

net.handle.apps.tools, net.handle.apps.site_tool

A number of utilities for low-level maintenance of a handle server. Make sure to check there before writing anything along these lines yourself.

Application Programming Interfaces (APIs).

In addition to Java, libraries are available for Python, Perl, and C. DOI specific libraries will soon be available with the Acrobat/DOI-services prototype.

Please pass on any unfulfilled needs you have for supporting software. We are always looking ways to make handles and DOIs easier to use.

6.5 Relationship between the DOI System and the Handle System

6.5.1 Relationship between the DOI System and the Handle System

The DOI system is an implementation (application) of the Handle System, which adds other features and functions, notably metadata, policy, and business rules.

The Handle System is a protocol specification (plus a reference implementation), not fundamentally a commercial application - anymore than DNS or any other low level network infrastructure is, and it is useful to think of it as analogous. As a result the Handle System does not offer commercial-level technical support from its originators CNRI, or anyone else, other than specific contractual arrangements with CNRI or other entities that may build on the base of this published standard protocol (the IDF has such arrangements which also cover its registration agencies). Applications may be built on top of Handle, but it does not provide sophisticated applications "out of the box", by design. Commercial licensing of the protocol may result in some possible offerings of this form (akin to DNS in routers, etc.).

The DOI system is an application of the Handle System to intellectual property. (The DOI Handbook is a starting point for information.) It is more than the Handle System: it adds to the Handle System an approach based on structured associated metadata, policies, procedures, business models and application tools. It is being developed by the International DOI Foundation. Initial implementations are now being supplemented by increasingly sophisticated value-added tools for metadata management and content management through the Handle System multiple resolution function.

The DOI system is also not fundamentally a commercial software offering, but is intended to be a community effort to provide enabling technology which others may build on. However we are building a self-funding system, based on a network of registration agencies who offer operational DOI registration services and applications in exchange for a fee paid to support the development of the Foundation. As a result of this there is a fee obligation for participation as a DOI registrant; in return the IDF provides tools (e.g., The DOI Handbook) to support DOI use, which offer added value. Additional information such as metadata guidelines (application profiles) is offered, and administrative tools for registrants are being further developed. A major advantage of DOI is that our structured metadata model ensures ready interoperability between applications, which is of increasing importance. For this reason we are already in early discussions with some other Handle applications about how we can encourage convergence of our approaches using Handle.

6.5.2 Relationship between DOI and other “actionable identifier” systems.

The relationship between DOI and non-actionable identifier systems such as standard numbering systems has been dealt with in Chapter 4. Chapter 4 also discusses the relationship with standard protocols such as URI and URN. This section is concerned with other actionable systems, i.e. systems which set out to deliver some or all of the functionality intended by DOI. Some of these may make use of the Handle System.

In considering how the DOI system relates to any other naming/resolution/metadata system, we might consider a decomposition of the DOI system into its four principle components:

- identifier (enumeration, and syntax)
- resolution of the identifier (using Handle)
- metadata associated with the identified entity (based on indecs analysis)
- policy (implementation and deployment of the complete DOI system)

It is then useful to consider which of these components are common and which are different in the case of DOI and XID (a notional other system, “X identifiers”). The more components are shared, the more easily interoperability can be achieved:

- No common elements (competing, alternative, or complementary systems). Note that minimal interoperability can of course be conferred by including XID as the DOI Kernel element “OtherIdentifier”
- Common use of identifiers and resolution system.
- Common use of metadata, in addition to identifiers and resolution system.
- Common policy, in addition to metadata, identifiers and resolution system.

We assume an interest in using common components. Most commonly, the following considerations arise; these are orthogonal, i.e. one is not dependent on another and any combination is possible:

- ***DOI-Application Profiles***
  DOIs, whatever their naming authority, could use a specific DOI-AP used by XID; that is, XID could define and make use of a DOI-AP.

- ***Naming authorities***
  Each registered handle type in the Handle System is itself represented by a handle of the form 0.TYPE/XYZ, where 0 is a system naming authority reserved for identifying system components, 0.TYPE is the naming authority for handle types, and XYZ is the character string of the type itself. The type “NA” is reserved for “Naming Authority”: hence the Global Handle System contains Handles of the form 0.NA/10.1080 which resolve to "where do I find 10.1080"? Local Handle Systems then consist of HDLs in the form 10.1080/456 which resolve to information about that specific HDL.

  It is logically possible to envisage that e.g. "XIDs are DOIs in the form of a separate naming authority". Note that there is no difference between X being allocated prefixes 10.XXXX, 10.100.XXXX or 20.XXXX; all HDL naming authorities are "peers". 10.100 is not a hierarchy below 10, all NAs are opaque. A HDL client would just know which LHS 10.100 belonged in. However they look different (20.XXX looks very different from 10.4567) - which could offer an opportunity of "branding" with true interoperability within the same LHS. The naming authority issue is largely a business policy decision.

- ***Local Handle System (LHS) usage***
  The question of where the NA information lives. The mapping of NA to LHS is most efficient if done one-to-one (1 LHS to many NAs is allowable; 1 NA to many LHS is not recommended). Hence it is possible that XIDs could live in a different LHS. This seems to be largely a function of network topology issues, not business: the principle to be followed is "do not unnecessarily proliferate local services". The Handle System provides interoperability across all handles, at least for anything that involves handle system resolution and administration: the consequences are performance, especially for administration, which is different than interoperability. So we would probably not want a LHS at every RA in the USA, but e.g. one in Asia might be justified (and might be used by any Asian DOI users). Whereas DOI-AP and NA issues are a matter for IDF and X, the LHS issue should involve the Global Handle Advisory Committee too (as it involves other, non-DOI, applications issues).

6.6 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.

6.6.1 Native resolver

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:10.123/456”, 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.

6.6.2 Proxy server

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:10.123/456) would be resolved from the address: “http://dx.doi.org/10.123/456 . 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.

The use of the DOI proxy server (the gateway between the Handle System and HTTP) does not interfere with the HTTP referrer field (that is, the source of the link is maintained, it does not appear as though the user is coming from dx.doi.org instead of from the source). Nothing goes 'through' that proxy server: it sends a redirect back to the original client with the current URL or other information relating to the handle resolution, and the final HTTP GET request comes from the user's client just as it otherwise would.

DOIs are used through a HTTP proxy server (in the "http://dx.doi.org" formulation as a URL) will continue to be persistent. As long as (1) the core DOI system is maintained, that is, as long as a given DOI (10.123/456) can be resolved using the handle system, and (2) as long as the proxy server named dx.doi.org is kept running, and (3) as long as the core network services that enable the http-based web to function remain in place, then a DOI (http://dx.doi.org/10.123/456) referenced through that proxy will remain persistent. The key to understanding why this is so is modularity. The core DOI resolution service is used by the proxy but is not constrained by the proxy. Additional gateways could be built and additional methods could be used to access the core DOI resolution system without interfering in any way with the ongoing operation of the dx.doi.org proxy.

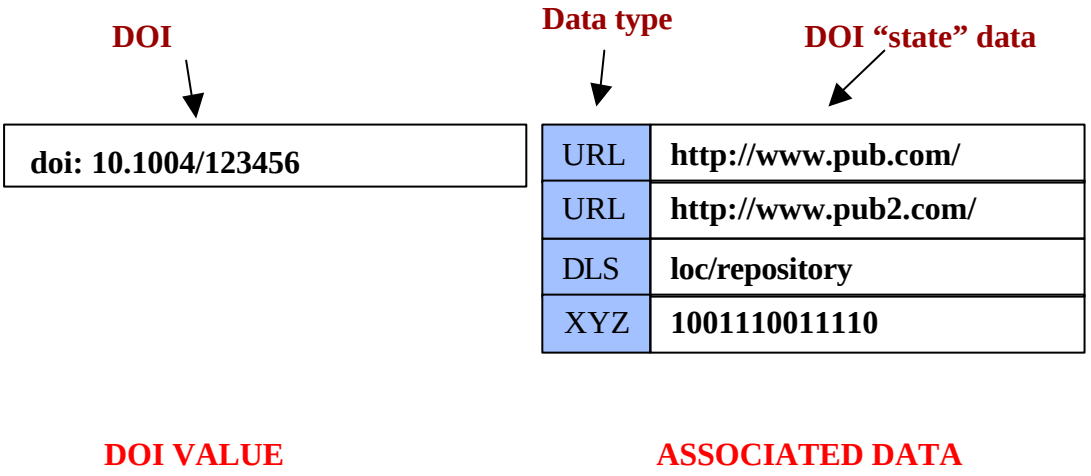
Having created and advocated the use of the proxy, CNRI and IDF are committed to maintaining it in perpetuity, as it will be an essential component to maintaining the integrity of the millions of instances of DOI-based web links. Maintaining the utility of those links over time will require maintaining both the core DOI system and the specific gateway service, dx.doi.org, that those links reference and so use to gain access to the core DOI system. This, of course, is not at all unique and is just another variation on the Internet theme of layering services on top of one another. dx.doi.org is itself dependent on the Domain Name System (DNS), which is itself dependent on IP addressing and routing, etc. This picture will probably grow more complex as time goes on (we hope it does), with the core DOI resolution facilities used in multiple ways and by multiple services. OpenURL resolvers, for example, will find DOIs in their 'raw' form, e.g., id=doi:10.123/456, and so could choose among using the dx.doi.org proxy, or setting up their own web-to-DOI proxy server(s), or using the handle protocol to query the DOI system directly.

6.6.3 Other mechanisms

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.

6.7 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.

For use with the data type URL (currently the most common application) we recommend that DOI data be entered as a full path? (e.g. http://www.somepublisher.com/photo/photo#1.gif), rather a relative reference . Whilst a relative link could be used as the DOI data, we cannot predict the context in which the DOI will be resolved, i.e. what the current base html reference will be.

A DOI could resolve to a Java applet or a CGI script or other dynamic mechanism.

6.8 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.

6.9 Relationship of DOI and OpenURL

Note: it is intended that this section will be amended to take account of any developments with the OpenURL specification arising from the current work of NISO committee AX, "Development of an OpenURL Standard"

6.9.1 Overview

OpenURL is a syntax for transporting descriptive information (metadata and identifiers) embedded within URLs between arbitrary services on the Internet. NISO has undertaken to standardize this syntax, as they earlier did with the DOI.

The Digital Object Identifier (DOI) is a system for resolution of identifiers to global services. OpenURL is syntax allowing the contextualization of requests to those services to local requirements. OpenURL can be used together with DOIs to provide a richer user experience that incorporates both the global and the local requirements of the user. The OpenURL format may become an important part of DOI Services as these are developed further. Both DOI and OpenURL represent maturing philosophies of identifying and linking to networked content by providing different but complementary, and potentially highly synergistic, methods of linking to multiple services associated with a single named entity. That is, both provide approaches for moving beyond the 'click it and go there' model for network content. To understand this, it is necessary to understand that both DOI and OpenURL are each much more than pointers following a certain syntax, that each also includes a network architecture for service provision and a set of assumptions or policies for the way in which that architecture is used.

In the case of OpenURL the object of that architecture, or framework, is to enable OpenURL-aware servers operating on behalf of some specific community or institution to take in identifiers and associated metadata, which are carried in an OpenURL, and return a collection of contextually appropriate pointers to relevant resources, e.g., a locally held copy of a given journal article accessible to members of a given university community plus references to other articles by the same author. A key issue in the OpenURL world is the transformation of a generic link, say to a publisher's online copy of a journal, into an OpenURL pointing to the right server for the given user, which must also carry the id and metadata needed to create the contextually appropriate extended service links as described above. In the current deployment this is only being done by the resource pointed at by the URL that the user initially encounters. So in the example of a link to the publishers copy of a journal, the publisher must 1) agree to redirect that http request to the user's local OpenURL-aware server, when appropriate, 2) must add information to the link as needed for the local server to do its job, and 3) must know the location of the local server.

In the case of DOI, the object of the architecture is to enable content producers to assign and maintain actionable identifiers to the content that they produce and to provide a network service which allows those identifiers, in the form of DOIs, to be resolved to multiple items of current state data, notably including location(s) and metadata. In the current deployment most DOIs resolve to only a single value, which is the publishers copy of, or web page for, the identified entity. The DOI community, however, is steadily moving in the direction of assigning and maintaining multiple values to most DOIs, specifically including core metadata. The retrieval and display of these multiple values is generally referred to as multiple resolution. A prominent issue in the DOI world, especially in connection with DOIs used in scholarly publications generally held in libraries, is the appropriate copy problem. Simply stated, the

logically centralized resolution service maintained by the content producers for their DOIs has no way to resolve a DOI to a locally held copy of the identified entity.

So the short-term synergy between DOI and OpenURL is clear: DOI needs a solution to the appropriate copy problem and OpenURL needs a source of identifiers and authoritative metadata. DOI also provides a (virtually) single point in the network for the creation and subsequent redirection of OpenURLs, which is more manageable than asking every content provider to enable this facility. Solving the appropriate copy problem would be a significant accomplishment in and of itself, but there are many opportunities for productive collaboration even beyond that. One particularly pressing issue is the problem of identifying users as part of a given community or environment and so identifying the appropriate local OpenURL-aware server. In initial prototypes involving DOI and OpenURL, this was done using cookies, an approach whose explanation is always prefaced with apologies and whose replacement is much desired. This same identity problem is, of course, also present in rights management and access control and so is of great interest to content producers as well.

6.9.2 OpenURL technical description

OpenURL is both a syntax and a system. The OpenURL system is not defined – anything that uses the OpenURL transferred data could be said to provide an OpenURL System. In practice the OpenURL transferred data is being used with information about the context of a user interested in a particular resource. This user-contextual information is not part of the OpenURL syntax but is instead supplied through other information supplied when the URL is activated (such as HTTP header information, a digital certificate, cookies or some other identification process).

OpenURL can provide a means of moving a description of an information entity to a service component that can offer context-sensitive resolution. It transports metadata or user information (as a by product) to a server that provides some response for the object for which the OpenURL is constructed.

The Open URL syntax consists of two components: the service request and the parameters. The target of the service request component of an OpenURL is a user's local service provider. The parameter component of an OpenURL transports the target object's metadata.

An OpenURL consists of a base URL followed by a query for one or more objects. So:

http://resolver.local.org/getlocal?author=Shelley

sends an OpenURL compliant request to a receiving service provided by getlocal at the location specified that a query with the parameter author and the value Shelley. What is not seen in the syntax is that the service will also receive any information about the user that may be sent along by default with the request as part of any authentication that has taken place between the users client and the server. The local service can then decide, based on the metadata sent and what the server knows of the user's credentials, how to respond to the request.

6.9.3 Using DOIs in OpenURL

DOIs can be used within the OpenURL syntax to query local services about availability of resources at a local level. As an example the following could be used to see if a local copy of a resource were available:

http://resolver.local.org/resolutionservice?id=doi:10.1045/1

The local service could have a list of DOIs that it has a local service for and offer the local service – potentially alongside the global information services obtained by resolving the DOI through the global handle system.

OpenURL also allows more complex constructs that those illustrated above. Multiple metadata items can be sent to a local service in a single request (potentially specifying more than one entity as in the example below):

http://resolver.local.org/resservice?id=doi:10.1045/1&&id=doi:10.1045/3

This OpenURL describes two objects referenced by DOIs separated by '&&'.

This can be further extended by adding information about where a user has been directed from (origin information). An example that illustrates both the use of metadata and origin insertion is:

http://resolver.local.org/res?sid=Dialog:DB&issn=12345678&volume=2&issue=2&spage=13

Here sid indicates the source and the database from which the link is made, and the metadata (ISSN, volume, issue and start page number) is used to determine what can be returned to the user (given the users rights that have been determined through an external authorization process). For information on NISO Committee AX (Development of an OpenURL Standard) see http://library.caltech.edu/openurl/Working_Documents.htm

OpenURL provides a syntax for encoding metadata about one or multiple resources. It provides a way to pack identifiers and metadata into a standard URL. This interface can be used to tie together otherwise disparate services such as centralized resolution systems and local knowledge of available resources.

In order to allow for the delivery of context-sensitive services information, recipients of an OpenURL must implement a technique to determine the difference between a user who has access to a service component that can deliver context-sensitive services; and a user that does not. The mechanism used to determine a user's membership of a particular group could be cookies, digital certificates, part of a user's stored profile in an information service, an IP address range, or something else. This user recognition is not a part of the OpenURL syntax and is separate to OpenURL.

If the user is a bona fide member of a group, the local resolution service will be available to that user. OpenURLs resolve to resources determined by context-sensitive service providers. Once an OpenURL is embedded in a resource it is fixed, and the service provider that it relies upon is explicitly specified by way of the pre-parameter part of the URL (hence an OpenURL has all the properties of any other URL). This provides an alternative resolution to the DOIs (provided by OpenURL-compliant service components) that can operate in a context-sensitive manner.

6.9.4 Comparison of OpenURL and DOI

OpenURL hard codes the location of the service in the URL – DOIs use the resolution system to provide access to a variety of services. As such the persistence of OpenURL is dependant on the availability of the service that is encoded in the OpenURL. The OpenURL once distributed cannot be modified except at each localised service. Where a DOI is used as the embedded metadata, it is possible that a user may be rejected from accessing local resources in which case deference to a global resolution system should be supported. The OpenURL intentionally embeds intelligence in a string that is supplied to a particular service – as a means to explicitly describe resources to attempt to provide a particular service based on that description.

The comparison of OpenURL enabled systems with the DOI system is straightforward. DOI is a global system. All information about a resource is the same in the global system wherever the DOI is resolved from (although multiple URL values or other data may be returned from the global system for subsequent selection and use by users with different requirements). The data that is associated with a DOI can be modified, and extended, and is not locked into embedded implementations. Even if an OpenURL carried all data that could be obtained from a DOI at a particular point in time, it would be static when used as a pointer in a document.

Thus the DOI system provides an authoritative centralized resolution system with careful control of the results of the resolution process. Additionally this identifier can be resolved to multiple pieces of information, including pointers to well-structured metadata.

The result of a DOI metadata resolution could be delivered in OpenURL format. A default resolution for items not available within a service providing local context information service would be provided by the DOI. DOI provides an authoritative namespace with unique identifiers and corresponding authoritative metadata and mechanisms to register identifiers – OpenURL does not support these functions.

6.9.5 Localization using DOI and OpenURL

OpenURL is premised on the idea that links should lead a user to appropriate resources. Systems that support the use of OpenURLs for localization will not resolve DOIs by the default DOI resolver. Rather, the DOI-link will be forwarded to an alternative resolver that will provide context-sensitive services. The user must have access to an alternative resolver and the authority that has registered the DOI must allow the alternative resolution. In this scenario the DOI becomes the key to the alternative, or complementary, resolution thus providing consistent delivery of content for both global and local use.

A project to use OpenURL to address the 'appropriate copy' problem was undertaken in 2001 with participation from CrossRef and organizations in the Digital Library Federation. This experimentation involved supporting the selection of the most appropriate copy or copies for a particular user where multiple legitimate copies of an article exist - typically that for which the user has access rights by virtue of their affiliation. In this initiative, various implementations of localization servers were used to read an OpenURL as input and take action upon it. Results were published as "Linking to the Appropriate Copy: Report of a DOI-Based Prototype"; D-Lib Magazine, September 2001 (http://dx.doi.org/10.1045/september2001-caplan).

6.10 Relationship of DOI and Persistent URLs (PURLs)

"A PURL is a Persistent Uniform Resource Locator. Functionally, a PURL is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. The client can then complete the URL transaction in the normal fashion. In Web parlance, this is a standard HTTP redirect.

The OCLC PURL Service has been strongly influenced by the active participation of OCLC's Office of Research in the Internet Engineering Task Force Uniform Resource Identifier working groups. There is nothing incompatible between PURLs and the ongoing URN (Uniform Resource Name) work. PURLs satisfy many of the requirements of URNs using currently deployed technologies and can be transitioned smoothly into a URN architecture once it is deployed" (from www.purl.org)

As noted in this statement, PURLs are all http. This is both their strength and their weakness. When you send a PURL to a PURL server, you are really sending a special URL to a web server via http, and the web server will send back a perfectly typical web server answer - all http. The difference is that there is a special PURL server attached to that web server (it may very well just be an Apache module) that inspects the URL, looks at a table to see what it means today, and returns that. It is one level of indirection, just like a single value DOI Handle, but it is all contained within a single server and that single server is permanently attached to a specific domain name. PURL servers don't know about each other.

So in that way it is no different from the way DOI is using a Handle proxy, dx.doi.org, which re-interprets DOI Handle queries into http. If DOI were never going to go beyond the proxy server approach and never make use of the multiple resolutions and data types PURL would be

a comparable and perhaps simpler and faster technological component then the DOI's chosen Handle protocol (www.handle.net).

There are ways in which one might imagine PURLs being developed to provide an approximation towards multiple resolutions and multiple data types. Content negotiation has always been in http, but like most of the www stuff it is oriented at attributes of the document in hand. The more you push this, from document centric things like "give this to me in German" to more 'attributE' things like "tell me about rights", the more tenuous the approach would become.

PURLs being wrapped up in http means they are designed to be used only in the web, this may not be an obvious problem at present, but the development of many mobile and other platform technologies means that not everything that happens on the internet from this point forward will necessarily be an extension of the www protocols.

PURLs have been widely available for several years but are not widely implemented in commercial settings and do not in our view provide a sufficiently sophisticated infrastructure for identification in relation to DRM.

In summary:

- PURLs are just single redirect servers (another level of indirection) whereas
- DOI is a direct protocol (but DOI implementation is currently using a proxy in most cases)
- DOI has multiple resolution and hence can build intelligent clients
- DOI is scalable
- DOI is a URN (or at least URN-like) hence useful in other contexts than the web as an Information identifier (hence its use with, and the importance of, metadata etc)
- DOI is a managed system whereas PURL is local and needs local technical support etc.
- DOI has additional policy rules and business implementation requirements.

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.

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 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, working groups, and/or interested parties for comment. In response to member comments, further drafts may be prepared reflecting the consensus view of the membership and published in the Handbook. Significant changes in policy are presented to the Board of the IDF for approval.

7.3 Current priorities for policy development

Current work on policy is mainly dealt with through the Registration Agency Working Group (RAWG), since Registration Agencies provide the social infrastructure for the implementation of the DOI system.

Some of the issues involved are discussed in Chapter 11 (The appointment and role of Registration Agencies). Policy development includes:

- The development of policies relating to the establishment and use of DOI Application Profiles (see Chapter 5). Documentation of metadata requirements and development

of tools for DOI application profile development will inform the appropriate policy decisions.

- Articulation of DOI services etc. The Technical Working Group is establishing mechanisms for the creation of DOI services, using multiple resolution and metadata techniques. Implementation of these mechanisms may raise issues of governance, access authorisation etc.

- The next iteration of an RA Agreement articulating the relationship of an RA to the IDF. An existing terms and governance document has outlined these issues (RAWG).

- The extension of DOI applications to new sectors and applications and the policies related to social governance. There is a natural desire for each major community to be self-determining, yet also a requirement for some common rules of the road (extreme granularity, complexity and eccentricity are commonplace when describing rights in a multimedia, multinational and persistent environment). In some cases natural areas of overlap (images embedded in text for example) may have policy implications; the RA concept allows for considerable autonomy in DOI applications. Mainly these decisions seem to rest on politics and market acceptability rather than any clear test of technical interoperability. What is needed is community involvement from non-text sectors, which the IDF has been tirelessly promoting and seeking.

7.4 Some major 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 standardized 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.

- Records of resolution activity by DOI are not made available. However the IDF may make data patterns and registration information available for research in the interests of e.g. Handle performance analysis and optimization. In such cases, all data will be scrambled using a one-way hash on each DOI value in resolution log data. This renders them completely un-interpretable and un-resolvable, but retains the patterns of usage for research analysis.

- 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.

7.5 Handle System policies and procedures

As the DOI uses the Handle System, it inherits the underlying policies and procedures of that System. These are summarized as follows (January 31, 2002; updated policies and procedures may be issued from time to time by CNRI and, where possible, this Handbook will incorporate such updates immediately).

7.5.1 Policies and procedures

Achieving overall integrity of the Handle System entails ensuring that each of the following conditions are maintained by participants in the system:

- Compatibility and smooth inter-working among system components;
- Consistency and reliability in service performance;
- Proper management and performance tracking of the overall system;
- Non-interrupted operation of GHR; and
- Overall system security.

7.5.2 Requirements for Third-Party organizations running LHS components shall include:

- Compatibility Test Suite Compliance (under development) covering:
    - Performance requirements;

        - Security requirements (proper updating, GHR service information, proper management of public key pair, proper implementation of security protocol and support of security algorithms); and
        - Operational requirements.

- Registration Requirements:
    - All Naming Authorities ("NAs") must be registered in the GHR, including the designation of an Administrator for each NA;
    - Secure maintenance of Private Keys by each Administrator; and
    - Timely report by the Administrator of service configuration changes, such as a list of NAs and the current type descriptions in accordance with the Handle System typing framework ("HTF").

Third-party organizations developing Handle System client components are encouraged to use the CNRI client software and the standard API supplied with it. In the event such organizations wish to use their own interfaces, these interfaces must remain compatible with the then current Handle System interface specification, which may evolve over time, as posted at CNRI's handle website: www.handle.net.

7.6 Data associated with a DOI

The simplest DOIs (such as those in the earliest implementations of DOI) are essentially redirection from a persistent name (the DOI) to a changeable URL. The information associated with the DOI in the DOI system is therefore simply the URL and relevant administrative information for managing the DOI.

However, in more sophisticated applications, a DOI may have additional associated data which help characterise the identified entity and which can be used to build various applications and services related to the identified entity. The Application Profile (AP) is a key example of such additional data. APs are used to group sets of DOIs which have similar characteristics, such as the same metadata schemas and business rules for DOI assignment. Thus, discovering that a given DOI is a member of a given AP is a shortcut to knowing what metadata elements can be found for the DOI, for knowing who is responsible for maintaining the DOI, and for any other characteristic that is common to the set of DOIs which are of that AP.

DOI data which is not common to all members of an AP will be associated with an individual DOI on a one-to-one basis. All Application Profiles contain a minimum of some publicly declared metadata (the kernel metadata) which is sufficient to provide users and applications with a basic description of the entity identified with a DOI

Application Profiles are described in more detail in Chapter 5. The association mechanism is described in Appendix 2; further development of documentation for APs is in progress.

7.6.1 Application Profiles

Every DOI is associated with one or more Application Profiles (APs). APs, which will themselves be identified by DOIs, are abstractions used to group DOIs into sets in which all DOIs of the given set, or AP, share a metadata schema, business rules for DOI assignment, and other common characteristics. An AP consists of at least a set of structured metadata elements, plus some rules (policy, business and procedural rules, not all necessarily automated). AP metadata, business rules, and other specifics will be determined by the community defining the AP; in practice this is likely to be, or to closely involve, the RA concerned. The formal representation of APs is still under development.

APs are an aid to using DOIs, enabling all DOIs assigned to e.g. journal articles to behave in a consistent and predictable fashion that would necessarily differ from the characteristics and behaviour of DOIs assigned to e.g. recorded music. For example, if one intended use of DOIs

is to lead to metadata for the identified entity and the metadata for journal articles and recorded music, outside of a small common kernel, will be quite different. This is only one example: in fact the data structures and potential services associated with DOIs by their assigners will not only depend upon the type of entity being identified but also by the intended usage of the DOI. An Application Profile groups together characteristics not only of the type of identified entity (roughly what has been called the “genre”) but also the intended usage, or application, of the DOI.

The core elements of an AP will be a metadata schema and various business and procedural rules. The business and procedural rules will cover such policies as "who can assign a DOI within this AP" and "what elements of metadata are public in this AP" and so on. The metadata elements common to all members of an AP will be defined through the use of the DOI data dictionary, which is an implementation of the indecs data dictionary developed as part of the ISO MPEG-21 process. Entities within this data dictionary will be assigned a unique iid (indecs identifier). In the DOI implementation of the data dictionary, each iid will also be a DOI (DOI.iid). This standardization of elements will allow developers, using a planned registry of APs, to know which elements are shared by which APs. Beyond metadata and business rules, APs may also include standard services, e.g., any DOI of AP X may be sent as an http query to location Y in order to request rights information. The use of DOIs to identify APs brings the standard benefits of indirection, that is, location Y in the above example can change without affecting the millions of DOI records that might reference AP X.

The DOI/AP relationship can be in one of three states:

- Zero AP: no AP is associated with this DOI. Most DOIs are currently in this state.
- Base AP: the only data associated with this AP is the kernel metadata (the minimum set of 6 elements, plus the DOI value and the DOI AP name).
- Full AP: the kernel metadata plus other metadata (which must be mapped to the DOI data dictionary) plus business rules and procedures plus any other common elements such as available services. We expect a number of different APs to evolve, roughly corresponding to communities of interest.

APs are intended, as are the other DOI mechanisms, to serve as infrastructure for the coherent management and use of intellectual property. While they will be defined and maintained by communities of interest, probably as represented by IDF Registration Agencies, they may also serve as convenient mechanisms for associating third party services with classes of DOIs. Registries will be established for this purpose. The specific rules and procedures for relating a given AP to third party services will be determined by the creators of that AP. To the extent that an AP is public, of course, anyone may operate a service applicable to that set of DOIs.

New APs must be approved by the IDF and centrally registered, to minimise duplication of effort and maximise interoperability. Defined APs will be made available to others who wish to construct new APs or re-use existing APs.

Registration Agencies add value at various levels by offering services to registrants. These services can include the definition of APs, and the one-off mappings needed in creating these from metadata sets already in use in the particular community concerned. There are two mandatory requirements for an RA using the DOI metadata system:

1. To declare the DOI kernel metadata in a standard format

2. To map the full to the DOI standard data dictionary

Once an AP is defined, RAs can offer services in allocating DOIs within this AP, ensuring the AP information is completed, populating the DOI system with allocated DOIs, maintaining up to

date records, etc. Consultancy on implementation, design of new applications, etc are other obvious areas for business development by RAs.

For more on this topic, see Chapter 5 and the associated appendices.

7.6.2 Data held in the resolution system

The core resolution system for all DOIs is the Handle System® [see Chapter 6 - Resolution]. Each DOI is registered as a handle in the Handle System and associated with a set of typed values. These values are returned in response to a resolution request for a given DOI. The values can be changed while the DOI remains constant, giving the DOI its basic qualities of being both actionable and persistent.

DOIs, with the exception of certain special cases, are registered with a minimum of one value, of the type “URL”. This can generally be thought of as 'location' but it really functions as the default value of the DOI in the context of the web and may not actually be the location of the identified entity. For anything beyond the simplest DOI, the declaration of an AP is an additional value within the Handle record, with its own data type. The DOI kernel metadata has as one element, "DOIApplicationProfile," which will reference this same data.

The association of an AP with a DOI may be sufficient, or may require additional data within the handle record. If services are associated with a given AP, for example, but the location of the service varies with DOI, then the declaration of the AP may need to be accompanied by the location of the service specific to that DOI. Similarly, two Registration Agencies (RAs) could share an AP but, in order to determine which RA had registered a given DOI, the AP declaration would have to be accompanied by an indication of RA. The precise mechanisms for accomplishing these tasks will be defined by the AP. At a certain level of variability across DOIs within an AP, of course, it may be better to create an additional AP rather than stretch one to cover too many different cases. Functional requirements will determine which is the case.

Additional data, beyond APs and any DOI-specific AP data, can be associated with a DOI as it is found useful. While the association of services and DOIs can be done through the AP mechanism, it may be that some services are best associated with each individual DOI and not through an already related AP. If this additional data is related using the Handle System, new data types can be created, as the Handle System typing mechanism is extensible. As with APs, data types must be approved and centrally registered, with the aim of minimising duplication and maximising interoperability.

Where data types require entities which are already defined within the DOI Data Dictionary, the DOI.iid will be referenced. Data types will also be identified by means of DOIs.

The combination of data typing through the resolution system, and interoperable metadata accessed through an Application Profile, provides a powerful set of tools for the creation of DOI services.

7.6.3 DOI services

DOI services describe the future vision of applications that exchange data, share tasks, and automate processes over the Internet by using the information associated with a DOI. The term was coined in analogy to “Web services”: for DOI applications on the Web, DOI services would be Web Services. As a new class of Internet-native applications, web services promise to increase interoperability and lower the costs of software integration and data interchange: these aims are clearly identical to those of DOI (and its underlying tools of resolution – Handle System – and metadata – indecs framework). Based on unambiguous rules, DOI services make it possible for computer programs to communicate directly with one another and exchange data about intellectual property entities regardless of location, operating systems, or languages.

For further information, an internal paper on “Web Services and the DOI” is available. Services definition in general is an evolving topic, and an active area of development for the IDF – please contact us for current information.

The combination of data typing through the resolution system, and interoperable metadata accessed through an Application Profile, provides a powerful set of tools for the creation of DOI services.

7.6.4 Information about a DOI which is publicly available

Once a DOI is assigned, anyone may resolve that DOI without charge. At least some information will always be available on resolution.

The information available on resolution depends on the Application Profile (AP) of the DOI. DOIs can be associated with one of three categories of AP Public availability of information is as follows:

- Zero AP: no data other than a URL is registered and therefore only that is available.
- Base AP: the kernel metadata set (the minimum set of 6 elements, plus the DOI value and the DOI AP name) is registered with each DOI within this AP. The values of each DOI’s 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 the DOI identifies can be accessed by any user and services built which can interpret DOIs.
- Full APs: these contain the kernel metadata set, plus other metadata values (which must be mapped to the DOI data dictionary). Whilst the AP scheme must be made available (so that users can determine which metadata fields are associated with the DOI), the actual values of any metadata for each DOI need not be; whether some or all of these are made available will be determined by the registrant or AP rules.

For more on this topic, see Chapter 5.

7.6.5 Private use of the DOI system

An exception to the rules on public information could be some uses of the DOI, which are not public (either permanently or temporarily). No such applications are currently in use; these would require special declarations and treatment. Private use of the DOI may have advantages either in conferring on a private scheme the benefits of interoperability, persistence, well-formed data structures, and governance structure; and in allowing the subsequent migration of private identifiers into the public realm without having to reassign identifiers with a policy or technical change which allows them to be private (and potentially switched to public) if desired.

The intention of the DOI system is the allocation of persistent, interoperable, well-formed, resolvable identifiers for public access and interoperation between parties without prior agreement: those DOIs are intended to be used by any discoverer of the DOI. Ideally, prospective applicants of DOIs would understand the benefits of a open, interoperable architecture. There may be instances where such public declaration is not appropriate for a users needs:

1. Where a wholly non-public use is envisaged, the concept of private use of DOIs may be a means of extending the adoption of the technology and principles that the DOI System is built upon and distributing costs and benefits widely. The features of the system also make it attractive for possible use in closed communities: interoperability, persistence, well-formed data structures, and governance structure persistence may be desirable within firms, or consortia, or between parties in an agreement. For example, use of DOIs for pharmaceutical research data may be very useful but public access to even basic information about what is being identified may be sensitive. The Handle System is available for private use (see

Chapter 6) but this alone does not confer the advantages of the DOIs structured metadata approach and rules-based governance.

2. Migration from Private to Public identifiers is a likely development which argues in favour of one underlying system. In addition, it is possible that some DOIs may be appropriate for private allocation but later made public: for example, some pharmaceutical research data may initially be private but then required to be made public as part of an FDA filing etc. Rather than migrate all the documents to a new identification scheme, for example. This implies that it would be advantageous if DOIs could be allocated on exactly the same technical basis as described for public DOIs, but with a policy or technical change which allows them to be private (and potentially switched to public) if desired.

This possible application could be very important to the success of DOI, but that private use will require private policies. There are currently no such applications of DOI so the following section should be considered to be an initial exploration of possible rules, not yet definitive.

Private DOIs could take different forms, but a common feature is that something about the data that the IDF states is normally public will be restricted to a particular community. One definition of “private use” could be anything that deviates from the information accessibility practices that we espouse through the DOI Handbook. This would cover:

- DOIs that were not known through the Global Handle System and so cannot be resolved by those who are not made aware of a particular local resolver;
- DOIs where the kernel metadata is not made available (so a user does not know what the DOI is associated with, but could nonetheless resolve it);
- DOIs where some of the resolution service data is not available without authorisation

It will be obvious that a spectrum of accessibility options could be devised in this way. This definition of “private use” of DOIs equates to 'not public use' of DOIs.

The concept of “private use” is something that we wish to explore further, to take advantage of DOI as a solution across diverse information asset management problems. Any organisation wishing to consider this area in more depth is invited to contact the IDF for an exploratory discussion.

The policy regarding registration agency agreements such as sub licensing aspects of the Handle System and the appropriate level of “franchise” fees will also need to be reviewed in such private applications.

7.7 Exclusivity

Exclusivity of DOI registration rights covering either a specific geographic territory or a particular wide area of application in general (e.g. a sector such as “audio”) will not be granted to a DOI Registration Agency.

IDF’s stance is that RAs will survive or perish as independent businesses on the basis of the added-value services and unique selling propositions they bring to offer to the market, not on the basis of DOI registration alone, or by being granted artificial exclusivity for a wide range of activities. DOI registration is "common infrastructure". We are aware however that this is new ground, and no one has a cast-iron demonstration of the best model in infrastructure development - the early days of telephones, railways, banks, bar codes etc. all offer different lessons, and the end position may differ from the start-up period. Investors in RA businesses are naturally looking for the lowest-risk proposition possible: this is understandable; however, no commonly accepted models for namespace management and resolution services in general are yet established. In the future, globally unambiguous names will be critical, and a ubiquitous infrastructure to "resolve" those names will be essential. We must allow service and technology providers to overlap into adjacent domains, and we must leave interfaces

(technical and business) open to allow experimentation with various models to happen and evolution to follow its course.

7.8 DOI error messages

It is inevitable that in adoption of DOI technology, some mistakes or misunderstandings will occur. This may include actions, which result in an attempted resolution not being successful; it is very important that such errors are detected and corrected. Automated procedures for common errors may be possible in some cases, but in the first instance a simple procedural agreement has been implemented. The following procedure is now in place: it is likely that for now the bulk of the error reports will go to RAs which represent the majority of deposits - the current wording therefore reflects typical message, but other RAs will probably use similar procedures:

(1) Resolving (via the proxy) a DOI that's not in the system returns a "DOI Not Found" error page to the user. You can view the page with the following: http://dx.doi.org/10.1000/8888. The page requests that users report errors to a special DOI address (doi-help@doi.org.) There will be several staff monitoring the address.

(2) CNRI will do a preliminary check to rule out a system problem. If the problem is with the handle system, CNRI responds to the sender appropriately. If the DOI is not found, receipt of the error message is acknowledged with the following:

"Thank you for reporting this error. It has been forwarded to the appropriate DOI registrant contact for action. Note that if the item the DOI identifies is new the DOI may have been unintentionally made public before being registered with the DOI system and may be available later".

(3) CNRI will forward the sender's message to the RA. If the DOI belongs to a "default RA" owner, CNRI will report the error to the owner and follow up.

(4) The RA will take appropriate action, and see that the sender and CNRI (via cc's to the doi-help@doi.org address) are kept informed of the action taken. This is meant to be informative and, at least for now, the IDF, via CNRI, will not try to put in place any kind of system to guarantee that the RA takes the appropriate action.

7.9 Ensuring persistence

7.9.1 Persistence of DOI

It may seem odd for "persistence" (permanence) to be discussed in a chapter on policy, rather than on of the sections on technology. There is a simple reason for this: persistence is ultimately guaranteed by social infrastructure (policy); persistence is fundamentally due to people, and technology can assist but not guarantee.

Identifiers must persist in the face of legitimate change. There are legitimate, desirable, and unavoidable reasons for changing organisation names, domains etc. One aim of naming entities/resources is to avoid tying an entity name to a domain name, or any other piece of variable metadata. Consider the domain names/trademarks issue. The entity can be persistently named as a first class object irrespective of its location, owner, licensee, etc. Distinguishing names from locations is essential for E-commerce. It is trivially true that "all names are locations" (in a namespace), but practically, most people worry about spaces like URLs, and that's the wrong level. Naming entities as first class objects, rather than locations, enables better management of multiple instances of an object, for example.

Persistence is again something we are familiar with in the physical world: ISBNs for out of print books can still be useful. Persistent identification alone is a good enough reason to adopt identifiers such as DOI which provide a means by which potential customers can find your digital offering even if a “broken link” URL of a retailer or other intermediary intervenes.

Technology can help with persistence. Using DOIs, only one central record, which is under the control of the assigner, needs to be changed in order to ensure that all existing DOIs which are “out there” in other documents can still resolve correctly: a redirection resolution step enables management in the redirection directory, thereby ensuring that one change can be picked up by many users, even if they are unaware of the change. But to manage the data in the directory takes effort, time, incentive, etc. - either you do that locally (using tools such as purl, managing a service yourself) or as a global service (DOI). In the case of DOI management of data is a service role (and hence also business activity) for registration agencies. We can learn from other activities like bar codes, ISBNs, and other data systems. People aren't free, so there's a cost to this, and just like the physical bar code system, the DOI aims to be a self-funding operation. DOIs won't be appropriate for many things, and some people won't feel this people cost merits the reward. But we do think DOIs are a viable solution for content management of intellectual property on a large scale.

DOI is an implementation of URN (Uniform Resource Names) and URI (Universal Resource Identifier) concepts, and can be formalized within these frameworks. The aim of each is to allow persistence of naming irrespective of other characteristics.

The central DOI resolution system is managed to ensure that persistent names can be resolved to non-persistent attributes such as location. One of the problems with the World Wide Web today is that once an object is moved, anyone searching for that object may encounter an error message. This is because URLs identify a location, not an object. The DOI, by contrast, specifies an object, not a location. Each DOI is registered in the Handle System and can be resolved to at least one location somewhere on the Internet. When an object is moved, all a rights holder has to do is repoint the DOI to the new location and the object can be found once more; any external party accessing the DOI does not need to know of the change and will be taken transparently to the object. The DOI system is designed to enable rightsholders to make up-to-date changes easily and consistently, and to monitor errors. Additional tools (such as workflow implementations) are already being developed by outside parties, and more will follow.

7.9.2 Persistence of the resolution technology

One of the key issues for the IDF in implementing Handle (HDL) as the technology for DOI was: how do we know that HDL itself is going to be persistent; will HDL be around in 5 years/50 years? There are both social and technical measures, which are relevant.

The HDL system is an open standard, so anyone can build/use one; but it relies of course on the top level Global Handle Registry to be in place somewhere (just as e.g. the internet Domain Name System assumes there will always be a root server and directory around somewhere). CNRI has a commitment to funding and maintaining these; were that to fail, there are enough large scale implementers of handles to ensure that it will be "picked up" by someone (e.g. Library of Congress, the US Dept of Defense, IDF, etc.). The Global Handle Advisory Committee, containing representatives of major handle users and stakeholders, exists to enable the fair and open evolution of the Handle System in the public interest and to promote its widespread adoption. IDF has a seat on the GHAC.

At the technical level we can take steps in improving resilience of the infrastructure, mirroring machines to insure against power outages etc.: the normal things one would do to improve technical system reliability. Key Handle infrastructure is placed with a professional hosting company with massive resources to ensure 24x7 cover. Further steps to make the system more persistent from an organizational point of view are under discussion, largely influenced by IDF requirements.

7.9.3 Persistence of the identified object

Just as there are legitimate, desirable, and unavoidable reasons for changing organisation names, there may be equally legitimate, desirable, and unavoidable reasons for declaring that an entity identified as a DOI is “no longer available”. For example, a major publisher’s policy on article withdrawal of electronic products states: “under exceptional circumstances, an article must be removed from an electronic product due to legal obligations on behalf of the Publisher, owner, copyright holder or author(s); or on moral or ethical grounds if an article with an error, or with results/statements has been found inaccurate and could be potentially damaging”.

The DOI system provides a mechanism for managing this process. At minimum, a DOI registrant is free to have the DOI resolve to a response screen indicating that the identified entity is no longer available. This is in itself will be very useful (consider for example that ISBNs for books which are out-of-print are still useful). A response such as this is certainly more useful than “404 not found”. Beyond this, a publisher or RA is free to define its own policy; for example, the entity may be made available in an archive form, with a reason for the withdrawal noted.

This is closely related to issues of archiving and preservation, in which IDF has an active interest: it may be useful to develop a default or fall-back mechanism for certain Application profiles or DOIs, whereby DOIs which are no longer available through the original distribution channel of their registration are re-directed to an archival source, or to a standard source of data.

There may be specific rules for this developed within a DOI Application profile; or there may be some generic rules, which can be devised for all DOIs. It is clear that there are many different reasons for such “out of print” digital objects: for example, an old version replaced by a new: the publisher response to a query about an old edition DOI could be a creative marketing approach (i.e., "you have requested information on an early edition which is no longer available.." or, .." it has been superseded by the new edition but you can still obtain the older version from xyz [second hand dealer or old source]...")

IDF has concluded that it would be premature to determine a one-size-fits all mechanism; it is likely to be a result of functional requirements for the particular DOI, including commercial issues. However this is an issue where we welcome active contributions and suggestions.

7.9.4 Stability and invariance of the associated metadata

A principle policy is that kernel metadata be stable and persistent (see Appendix 2). This needs to be considered in relation to the very common situation where, when a commissioned work, or a planned publication, is first registered in a database, it has only a working title and possibly even a proposed author who may then turn the project down. There are three logical ways of handling this:

1. Make it an absolute rule that if the metadata changes in any way at all, a new DOI must be registered (it may be questioned whether this is enforceable);
2. Allow some kernel elements (title, primary agent) to have one or more superseded values as well as their current value. "Stable” would then mean that the content of an element, once entered, could not be changed, but its status as "current" or "superseded" could; and "persistent" would mean that superseded values would never be deleted.
3. Adhere strictly to the literal interpretation of "stable” as “invariant" and "persistent", so that if (say) a title changes after DOI registration, the registered title in kernel metadata cannot be changed, but the extended metadata managed by the RA would carry the definitive title and link it to the registered title.

Each demands certain disciplines from registrants and RAs; and each has different implications for the possible use of the kernel metadata. This may be an issue which will be dealt with at the Application Profile level; working policy at present is that the precise interpretation of this is something that registration agencies may want to agree, but the *principle* is that the intention upon registration is that the *kernel metadata is not likely to change*. While we acknowledge that mistakes and updates occur in the real world we are setting a high standard to encourage registrants to get it right. All registration agencies will want to allow errors to be corrected, so we should not assume a difference between theory and practice. For this reason we adopt the term “stable” rather than “invariant”.

7.10 Ensuring uniqueness

The system will not accept duplicate DOIs. As each prefix is unique to a publisher, no two DOIs from different publishers can ever share the same prefix.

As far as is reasonably possible each DOI should be associated with a unique item of intellectual property. IDF policy is that there should be a method for registration agencies to check if an item of intellectual property already has a DOI assigned to it; initially this may be through arrangements between RAs operating with the same or related Application Profiles. The action to be taken if a match is found is to be determined by the registration agency. There will be no requirement for a centralized database maintained by the IDF that can be queried by any registration agency.

7.11 Separation of metadata from identifier

The DOI identifier (number) does not carry in its syntax any information about e.g. who assigned the DOI; the DOI is a dumb number (an "unintelligent" or "non-significant" identifier). Any such intelligence is to be found in the accompanying metadata (see Chapter 5).

The DOI identifier string is an opaque string. 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.

The logic behind this is inescapable and applies to any identifier, which claims to be persistent: the identifier (DOI in this case) must be persistent; therefore, if an entity changes ownership, its identifier cannot change; therefore, the identifier itself cannot indicate anything about ownership, like "publisher". The same is true for any other aspect, which is changeable. Assigning a prefix to a publisher (or anything else; a journal, an imprint, a record label) is a one way function. It enables unique numbers to be generated, but you cannot do the reverse: "this is 10.12345 so it must belong to...." is not valid.

The suffix of a DOI may, if the registrant wishes, be an existing identifier (e.g. an ISBN, or a SICI) which contains some intelligence in that other system (see Chapter 4). A general point here is that if an identifier allows the rules of construction to include the use of intelligent components (e.g. the registrant or an ISBN in the suffix) there are going to be stakeholders/users who will believe that they can interpret the identifier. Such uses may be legitimate in the namespace of that previous identifier, though not within the DOI system itself. The dangers of this must also be recognised (and this is true of any identifier system, not just DOI): intelligent numbers are "hardwired" with metadata rather than having the ability to have the metadata re-wired (i.e. updated). It is preferable to use well-formed metadata linked to the identifier but capable of being updated, for information concerning rights, ownership, etc. that is inherently changeable. Intelligence in the number is less of a problem if the intelligence in there is inherently unlikely to change (e.g. page number of an

identified publication -- hence identifiers like SICI which are intelligent). But it is problematic if you try to put in changeable information like publisher: nearly all the interesting metadata may need to be changeable ultimately (for persistent future-proofing/ archiving -- even (especially) things like file formats etc.) These points are fundamental to efforts such as <indecs>, which the DOI activity has used as a basis for its metadata principles.

7.12 Separation of identifier from object

In certain digital objects, an identifier such as DOI may be part of the binary bit stream, which makes up the object. However the DOI could be stripped from the bitstream and the intellectual property could be used without the identifier. DOI itself does not provide a mechanism to prevent this, but third party developers are free to offer added-value features such as authenticity checks and copyright management systems that will block access if a DOI is missing or tampered with. The widespread use of a standard DOI system will encourage the development of such systems by offering a very large potential application market.

Assigning an identifier does not of itself change any aspect of a digital object's use; it is a prerequisite for copyright management - necessary but not sufficient - hence copyright infringement could occur. However, removing such an identifier from material constitutes "modification or removal of copyright management information" and is prohibited under legislation implementing the WIPO Copyright Treaty, such as the US Digital Millennium Copyright Act.

The point of the DOI is to identify materials for legal management purposes. DOIs could be used with proprietary copyright management technology in mechanisms that enforce copyright technically. We view this as an added value use of the DOI.

There are application possibilities with Public Key Infrastructure and the Handle System protocol, which offers trusted resolution, as a possible infrastructure, which may be usable for DOIs. Of particular interest for commercial uses, the Handle technology includes two features suitable for trusted transactions:

- The system has been built to allow ad hoc trusted resolution across the entire distributed system. There is some overhead involved in utilizing this feature.
- The system supports distributed trusted administration. The permissions are at the handle (DOI) level, not at the server level. Again, this is especially useful for a distributed system. If you have administrative privileges on a given DOI, you don't have to worry about where or how many or even who owns or runs the servers. For further information refer to http://www.handle.net/papers.html

We have no intention of replacing existing value-added copyright protection measures (either proprietary or open standard); rather we wish to encourage their use with DOI entities by providing a readily interoperable platform and set of policies.

7.13 Use of DOIs by services other than the original publisher

Just as booksellers use ISBNs but don't assign them, any party in the information chain must be able to use "actionable identifiers" - DOI is not a "publisher-only" system. Identification is logically a separate function from rights assignment, hence the rules of an application profile need not specify that only the rightsholder may allocate an identifier. (Having a DOI assigned to an entity cannot, of itself, certify anything about the copyright status of that entity). DOIs may be *assigned* by third parties; and most certainly DOIs may be *used* by third parties. Rules about who may and may not assign an identifier within an application profile may be part of an Application Profile specification.

The digital world does not imply a single worldwide accessible copy, disintermediating all but the publisher. In fact, the opposite is true: information on a network is widely distributed, and whilst a rightsholder (DOI assigner) could in theory maintain a central database, the rightsholder cannot control every mention of his intellectual property, or enforce a single gateway to the material. There are many occasions where a rightsholder will want to encourage widespread access to his material by others independently of his control (but under specified conditions), to maximise value whilst limiting misuse. A useful comparison is the standard book number, the ISBN, which is used by millions of booksellers and libraries every day; those uses are not linked in any way to publishers' databases; they do not go through a publishers "gateway". Yet many added value services use the ISBN as a key to other databases. Customers and users need this consistent identification scheme to enable them to go to material with confidence through a variety of services, without having to check back to the original publisher every time they want to navigate on the network. DOIs can be used just as physical bar codes are in the non-digital world: as a means of automating the supply chain.

Even with a simple implementation such as single URL resolution, DOI may be useful in the same way as this ISBN example. But there are many more interesting possibilities, which can be envisaged too:

- Using multiple resolution, one or more of the resolution options could be to third part organisations, either for provision of services or for provision of content on agreed terms.

- The use of existing DOIs with local resolution technologies (e.g. OpenURL) allow for automating links to an "appropriate copy", a service which might be offered or facilitated by an intermediary or third party service. Having identified a resource generically with a DOI, I need to locate an appropriate source to which I have access rights, or on some other basis (this too could have a DOI).

- DOIs could be assigned within an application profile by service organisations, either on behalf of rightsholders or for entities where the original rightsholder is inappropriate (as a hypothetical example, out-of-print books lacking ISBNs).

- A record created "about" some other entity may itself have value (for example, a text synopsis or abstract). That record might itself be assigned a DOI, which has a relationship to the original work. The key to describing this relationship between two entities is metadata (The generic indecs metadata model we are using to develop further implementations offers formal mechanisms for expressing these relationships as part of application profiles); and that relationship may be automated using DOIs.

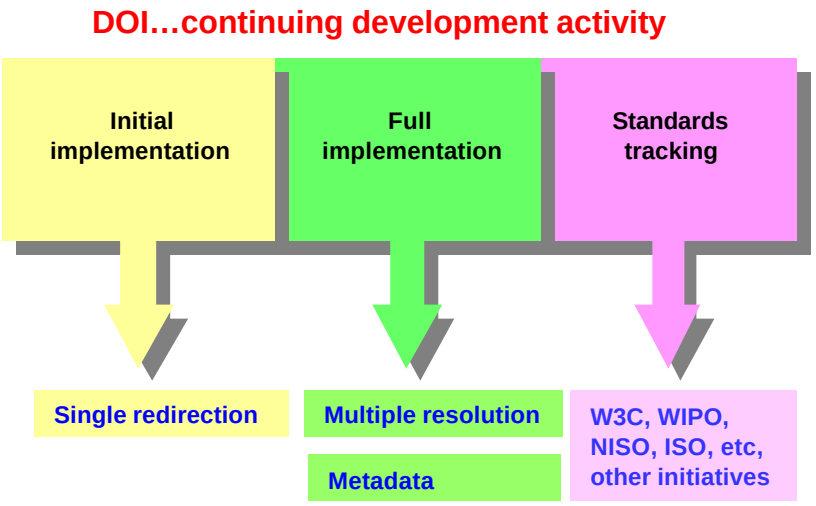
- Many organisations prefer to rely on proprietary identifiers assigned by their supplier (for example, a wholesaler or distributor) who they know will have managed this task consistently, rather than rely on the identifiers issued by the original providers which may not be comprehensive. A similar situation may well exist between service organisations using DOIs and their customers (DOIs can of course resolve to other DOIs or to an internal service organisation identifier).

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. It describes IDF's development activities and processes and its involvement in the development of underlying technologies, and with related standards development organizations.

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. Several 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 Chapter 5) 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.

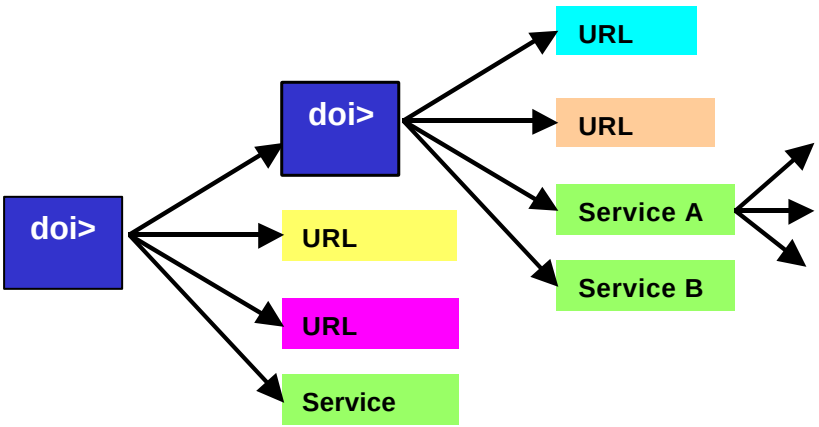
8.3 Full implementation

Work is currently 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 *Resolution*) 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 *Metadata*). This full implementation will eventually be the dominant model of use. In addition, the DOI Foundation is working very closely with all other relevant standards efforts to ensure total compatibility between various efforts from other organisations, and to take advantage of technology developments.

The principle of multiple resolution of a single DOI using Handle System technology has already been demonstrated. Initial demonstrations present the user with a pop-up window offering the alternative resolutions, but establish the necessary platform for the development of automated services. The IDF's Technical Working Group is constructing a detailed technical framework/architecture for articulating multiple resolution as defined services: one DOI may resolve to many different data types, which could represent many services. The working group is to define principles, rules, and guidance framework for consistent development of multiple resolution across applications, and as a means of building functions and services into DOI resolution. This will provide a bridge from "arbitrary" implementations (the list of possible things one could do with a DOI, the list of implemented things in demonstrations such as DOI-for eBooks) to an extensible, structured, framework where new features can be added by anyone but in a controlled way that ensures interoperability with other sectors, applications, and apparently (as yet) unrelated services. Likely next topics will include the use of these frameworks in conjunction with structured metadata, incorporating relevant development efforts now being carried out in the further development of the <indecs> project and further development of the DOI "Namespace" data dictionary.

The IDF is currently 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 DOI and Handle System development

CNRI continues to improve the Handle System. Feedback from users, including the IDF, as well as developments in related network protocols, have resulted in 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.

Much of CNRI's Handle System development work looks beyond the basic resolution facilities (rapid, scalable, and reliable resolution of "handles" to multiple, typed values representing current state data) to improvements in both server and handle administration (creating and maintaining handles and handle data) and security. The implementation of "sessions" for handle administration, which reduces the number of messages exchanged between client and server during handle administration processes, has significantly reduced the time required to process requests. In addition, support for encrypted communication between clients and servers has been added.

The protocol provides a "trust model" for the Handle System, based on public/private key encryption. Every handle server has its own public/private key pair that can be used to authenticate server-to-server transactions, and each individual handle has one or more defined "handle administrators". Administrators are themselves identified by handles, and each administrator may have his own public/private key pair (or secret key) for authenticating administrative permissions.

The Handle System infrastructure forms the basis, among other things, for a completely distributed administrative system. Any individual or process, with valid credentials, can administer handles in the relevant handle service. A reference may be included 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, the 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.

IDF participates in the Global Handle Advisory Committee, set up by CNRI to enable the fair and open evolution of the Handle System in the public interest; the committee provides advice and guidance to CNRI and the individual sector constituencies of the Handle System user community on matters of strategic direction, finance, technology and standards, infrastructure, administration, etc.

At the end of the project (June 2000) <indecs>™ Framework Ltd (a not-for-profit company, limited by guarantee) was established by the partners in the <indecs> project (Kopiosto, CAL, EDItEUR, IFPI, MUZE Inc. and IDF) to fulfil the sole task of owning the valuable intellectual property rights created during the project, and to encourage well-formed metadata initiatives based on the <indecs> analysis. The company was not constituted to limit the application of the lessons learned during the project; however, the partners agreed that the establishment of the company was essential both to own the IPRs and to avoid their potential misuse. Subsequently rights in the <indecs> Analysis were assigned by the company jointly to the International DOI Foundation (developing the DOI) and EDItEUR (developing ONIX International, a dictionary implementation of the <indecs> framework which is particularly useful as the working basis for deriving DOI Application Profiles in the text sector.) This was done on the basis that IDF and EDItEUR were active in public implementations of <indecs>, and each was a non-commercial entity; IDF and EDItEUR already collaborate closely.

8.5.2 DOI-Namespace development (2001)

The <indecs> framework was a reference model. In order to implement the reference model, a practical reference implementation was required. DOI-Namespace was an initial effort to implement the <indecs> metadata framework in a practical way, as a tool to build DOI Application Profiles.

The IDF took the <indecs> analysis and aimed for integration with CNRI's digital object architecture and technology in order to provide a coherent and consistent means of expressing interoperable metadata and identifiers in the DOI system. In order to develop tools for practical implementation such as Application Profiles (chapter 5), IDF conceived the DOI-Namespace (DOI-NS) as the necessary toolbox for the development of Application Profile metadata schemas, via a functioning "DOI namespace" metadata dictionary to support interoperability between DOI Application Profiles, and potentially with other metadata schemes (e.g. Dublin Core, ONIX, MARC, PRISM etc). DOI-NS was developed as a data dictionary of defined metadata "keywords" with strictly defined internal relationships, with structure is based on the <indecs> metadata framework and initial terms are drawn principally from <indecs>, EPICS/ONIX and Crossref metadata elements. DOI-NS is not a repository of metadata records, but a registry of mapped and defined terms. DOI-NS is not a metadata scheme or metadata set to be imposed on anyone. It is a tool for defining and mapping

metadata relationships to support interoperability. Keywords from other namespaces can be mapped to any other keywords, as synonyms or with other equivalence relationships.

Functions for which DOI-NS was (and is) intended are:

- To support the development of metadata sets (see below) for DOI Application Profiles (DOI-APs), including providing robust vocabulary sets for important terms (e.g. contributor types, version types) which currently lack.
- To register all terms in all DOI-AP metadata sets (e.g. providing common XML-tags).
- Provide mappings between DOI-APs.
- Provide mappings between DOI-APs and other metadata sets (eg ONIX, MARC, Dublin Core, SMPTE, SMEF, RIAA/IFPI, PRISM).
- To support a "transformation" service to enable data from one scheme to be rendered in another.

During the course of 2001, interest in the development of such schemes widened, and an opportunity arose to combine further DOI-NS work with a new phase of development of the <indecs> analysis with other organizations; this was <indecs>rdd. To this end, the work done on DOI-NS was used as a major input to <indecs>rdd.

8.5.3 Contecs:DD; the next phase in <indecs> development (2001-2003)

Contecs:DD is a consortium-based initiative to develop a rights data dictionary, originally known as the <indecs>rdd consortium and renamed in July 2002 (the renaming was merely for technical administrative purposes).

The <indecs> framework was a reference model. In order to implement the reference model, a practical reference implementation was required. DOI-Namespace was an implementation for IDF alone. <indecs>rdd is a generic Rights Data Dictionary (RDD), a common dictionary or vocabulary for intellectual property rights based on the <indecs> framework. It is the next phase in <indecs> development, which provides an improved basis for DOI Namespace with benefits of easier and widespread interoperability.

Because rights metadata is inseparable from other metadata, and because the <indecs> framework specifies a general metadata framework, the work done in developing <indecs>rdd also deepened and expanded the original <indecs> framework, building on it whilst providing a practical reference implementation. The dictionary resulting from this activity was adopted (in Dec 2001) as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard.

In April 2001 the International DOI Foundation (IDF) funded a feasibility study for a consortium which would fund standard rights terms to enable the exchange of key information between content industries for ecommerce trading of intellectual property rights. Recognition of the urgent need for a content industry wide RDD from activities in, amongst others, IDF, the Open eBook Forum (OeBF), the World Wide Web Consortium (W3C), the Internet Engineering Task Force (IETF) and the International Standards Organization (ISO) Motion Picture Experts Group (MPEG) prompted the IDF's funding for the study. Founding sponsors EDItEUR and the International DOI Foundation (IDF) were joined by the Motion Picture Association of America (MPA), the Recording Industry Association of America (RIAA), the International Federation of the Phonographic Industry (IFPI), Accenture, ContentGuard, Enpia Systems, and Melodies and Memories Global (a subsidiary of Dentsu) in a consortium to develop <indecs>rdd. IDF and EDItEUR donated a license to the intellectual property of the original <indecs> project, and IDF donated the DOI-Namespace work done to date, to the consortium. Rightscom, the digital rights strategy consultancy, was project manager of the consortium; the original <indecs> technical consultants were key to the work.

The dictionary resulting from this activity was adopted (in Dec 2001) as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard, to provide a key part of the architecture required to deliver interoperability between digital rights management (DRM) standard systems (see Appendix 4). The adoption of the <indecs>rdd proposal marks the beginning of the collaborative phase of this very important work in MPEG, which is scheduled

for International Standard in March 2003. The MPEG Rights Data Dictionary and the accompanying Rights Expression Language specifications will allow interoperable large-scale exchange of digital media, including subscription services, trusted peer-to-peer services and a wide variety of other distribution models.

The <indecs>rdd submission was selected as MPEG baseline technology, but further work is necessary to ensure compliance with the rest of the MPEG range of standards, and enable mapping to and from schemas outside MPEG so that interoperability is enhanced. Clearly the functionality of such a tool is very similar to that required for DOI Application profile development. IDF is now working with the <indecs>rdd consortium to further deepen and develop <indecs>rdd as part of the MPEG process; in parallel, we are also developing tools and procedures for DOI Application profile development which can be implemented in the short term (and not need to await completion of the MPEG 21 process final MPEG-21 standard, currently planned for publication at the end of the first quarter of 2003).

Developing DOI with <indecs> in this way has provided two very significant benefits:

- It provides a more cost effective and widely supported means of developing the appropriate tools than developing DOI-Namespace alone;
- It will ensure that DOIs assigned using the DOI Application profile metadata approach will be compliant with the wider ISO MPEG21 Multimedia Framework.

We see this as a highly significant step in the development of DOI.

8.6 Standards tracking; strategic alliances and liaisons with other organizations

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)
- OEBF (the Open eBook Forum)
- MPEG 21 (the ISO Multimedia Framework; see preceding section)

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).

The International DOI Foundation is a member of some standards organizations, and maintains a number of liaisons or alliances through memberships and/or exchange of information with others, which allow us to act as a collaborative interface in discussions on standards and infrastructure development across the spectrum of intellectual property and technology communities. This provides advantages both to members of the Foundation (who may otherwise not be able to participate in all of these discussions) and to the strategic partners (who deal with IDF as a common voice for the intellectual property community in this area).

Alliance organisations normally receive a copy of the monthly IDF Director’s Report, in exchange for their equivalent report. Many of our alliances are member organisations in their

own right; in this case the IDF report is not for distribution to their members, but for the information of the alliance representative and their immediate colleagues. The aim is to alert partners of issues of relevance, and provide a mechanism for the alliance partner to alert IDF to any activity they may already be working on which is relevant to an IDF activity.

In addition to the major alliances noted here, the IDF has a number of other relationships with significant development and standards activities in many areas of intellectual property and technology. Some of these are specific to particular application areas, and are undertaken in order to seed activities and outreach from the DOI to potential implementations. This list is expanding and we welcome expressions of interest from organizations who wish to establish such a relationship.

Auto-ID Center (http://www.autoidcenter.org/main.asp)
The Auto-ID Center, headquartered at the Massachusetts Institute of Technology with a sister lab at the University of Cambridge UK, is developing an Electronic Product Code (ePC) which will be embedded onto individual products and physical objects on memory chips ("smart tags") that connect objects to the Internet. The activity is an evolution of the EAN/UCC Universal Product Code (bar code). The intention of the project, funded by manufacturing industry, is to "allow the Internet to extend to everyday objects", connected in a dynamic, automated supply chain that joins businesses and consumers together in a mutually beneficial relationship.

The Auto-ID Center ePC Alliance is made up of non-profit trade body organizations with 10 members or more with an interest in this and related technologies; IDF is a member of this committee. The Alliance works as a committee, looking at the applications of the ePC for their specific constituency. Members include such diverse groups as retail, food marketing, grocery manufacturers, chain drug stores, convenience stores, textile services and hologram manufacturers.

CENDI
CENDI is an interagency working group of senior Scientific and Technical Information Managers from ten major programs in nine U.S. Federal Agencies from the sectors Commerce, Energy, Environmental Protection, National Aeronautics and Space Administration, National Libraries, Defense and Interior. A liaison program has been instituted to share information and consider content management opportunities; some of these agencies are also users of the CNRI Handle System.

cIDf
The IDF is collaborating with the Content ID Forum with a view to seeking harmonization of their work on identifiers for content in the digital environment. The Content ID Forum was established by Professor Dr. Hiroshi Yasuda at the University of Tokyo for the purpose of providing a mechanism for copyright management. cIDf has a special interest in embedding identifiers within digital objects. cIDF's initial application focus was on video and images.

CNRI
The Corporation for National Research Initiatives (CNRI) undertakes, fosters and promotes research in the public interest. The activities centre on the strategic development of network-based information technologies. DOI uses CNRI's Handle System®, a distributed computer system which stores names, or handles, of digital items and which can quickly resolve those names into the information necessary to locate and access the items, and has a collaborative agreement with CNRI.

COUNTER
(Counting Online Usage of Networked Electronic Resources; http://www.projectcounter.org) is an initiative, driven by librarians and publishers, to develop tools that will provide both with greater insights into online information usage. Project COUNTER, which is governed by a fully international Steering Group and has a dedicated project director, is a development from the Usage Statistics Working Group of the Publishers and Libraries Solutions Group (PALS), a joint

initiative of the Publishers Association (PA), the Association of Learned and Professional Society Publishers (ALPSP) and the Joint Information Systems Committee (JISC).

IDF recognises that in order to count usage it is necessary to precisely identify and differentiate the entities being counted: the DOI offers a tool for this purpose and we have advocated its use in this project. At this stage it is premature to make a concrete proposal for any organization that would ultimately be responsible for the implementation of COUNTER.

EDItEUR
EDItEUR is the International Group for E-Commerce Standards for the Books and Serials Sectors. These standards include the ONIX International dictionary and expressions in XML, which provide a basis for the construction of metadata sets for multiple applications.

ICE
The Information and Content Exchange (ICE) - (http://www.icestandard.org). This protocol specification provides businesses with an XML-based common language and architecture that facilitates automatic exchanging, updating, supplying and controlling of assets in a trusted fashion without manual packaging or knowledge of remote Web site structures. ICE facilitates the controlled exchange and management of electronic assets between networked partners and affiliates. Applications based on ICE allow companies to easily construct syndicated publishing networks, Web superstores, and online reseller channels by establishing Web site-to-Web site information networks.

IETF
The Internet Engineering Task Force (IETF) is the protocol engineering and development arm of the Internet. IDF participates in appropriate IETF discussions and meetings.

<indecs>
The IDF was a partner in <indecs> (Interoperability of Data in E-Commerce Systems), an international collaborative project to develop a framework of metadata standards to support network commerce in intellectual property. The work of the <indecs> initiative is now being continued as Contecs:DD and further developed jointly by IDF and EDItEUR (see Chapter 8).

IRTF
The Internet Research Task Force is a sister organization of IETF that promotes research of importance to the evolution of the future of the Internet by creating focused, long-term and small Research Groups working on topics related to Internet protocols, applications, architecture and technology. IDRM is an IRTF Research Group formed to research issue and technologies relating to Digital Rights Management on the Internet, in which IDF is active.

ISO
The Foundation has a Category A liaison with International Standards Organization (ISO) Technical Committee 46, which includes an exchange of all relevant official documentation and participation in the meetings and Advisory Group ISO TC46.

Within ISO TC46 we are particularly involved in the work of SC9, the TC 46 Subcommittee that develops and maintains ISO standards on the presentation, identification and description of documents (ISBN, ISSN, ISRC, ISRN, ISMN, etc.). Current work of SC9 includes the development of an International Standard Textual Work Code (ISTC); the development of this standard has touched on many issues common to the IDF's work. We have also been involved in discussions on the metadata associated with such identifiers.

MPEG
IDF is a formal liaison body with The Moving Picture Experts Group (MPEG). MPEG is a working group of ISO/IEC for the development of standards for coded representation of digital audio and video. IDF is particularly involved in the MPEG-21 "Multimedia Framework" activity because of the convergence of media types in a digital world, and the commonality of concerns and issues across sectors, where identifiers such as DOI can offer a key role in such standardization activities.

NISO
The Foundation is a member of the US-based National Information Standards Organization (NISO). NISO develops and promotes technical standards used in a wide variety of information services. NISO took an early interest in DOI development; the DOI syntax is a NISO standard (ANSI/NISO Z39.84-2000) which was published in May 2000.

WIPO
The Foundation has permanent NGO Observer status at the Assemblies of the Member States of WIPO (The World Intellectual Property Organization). WIPO is an intergovernmental organization with headquarters in Geneva, Switzerland. It is one of the 16 specialized agencies of the United Nations. WIPO is responsible for the promotion of the protection of intellectual property throughout the world through cooperation among States, and for the administration of various multilateral treaties dealing with the legal and administrative aspects of intellectual property.

The WIPO Digital Agenda, adopted by the Assemblies in September 1999, includes as one of its aims the facilitation of "interoperability and interconnection of electronic copyright management systems and the metadata of such systems" (Digital Agenda, item 6).

XBRL
IDF is a member of the XBRL (Extensible Business Reporting Language, International Liaison Committee. The committee's mission is "to grow XBRL to ensure interoperability, adoption and member success", consisting of representatives from geographic regions, various perspectives and industries. IDF was invited to represent a focus on the intellectual property communities.

8.7 IDF Development process

In addition to working with developers of the underlying technologies (Handle, <indecs>) and related standards (W3C, IETF, ISO, etc), IDF runs its own working groups and e-mail discussion lists.

E-mail discussion lists are of several types; some are public, some are restricted; some lists are run in conjunction with other organisations.

Working groups require a commitment from participants, normally in the form of IDF membership. An "Affiliate" status membership was introduced in 2001: this allows participation in a working group at a reduced fee, without a commitment to full membership in IDF. Other organizations may see advantage in full membership, allowing participation in multiple and generic DOI work. On occasion, IDF may invite non-member experts to join a working group.

Two working groups currently operate which are generic in application and which have a special status: the Technical Working Group (TWG) and the Registration Agency Working Group (RAWG). The RAWG and TWG groups clearly interact, as some policies require technical articulation, and some technical aspects may influence policy formulation. The WG chairs will exchange all relevant information, but at present there are sufficient differences between the membership and scope of the two groups to consider these as separate activities.

8.7.1 TWG (Technical Working Group)

The TWG is the primary focus for discussions on the implementation of the DOI technical infrastructure. This group evolved from earlier efforts in a number of areas, which overlapped considerably in coverage and functionality. The core of the group activity is currently the construction of a detailed technical framework/architecture for articulating multiple resolution as defined services: one DOI may resolve to many different data types, which could represent many services. The working group is to define principles, rules, and guidance framework for

consistent development of multiple resolution across applications, and as a means of building functions and services into DOI resolution. This will provide a bridge from "arbitrary" implementations (the list of possible things one could do with a DOI, the list of implemented things in DOI-EB) to an extensible, structured, framework where new features can be added by anyone but in a controlled way that ensures interoperability with other sectors, applications, and apparently (as yet) unrelated services. The TWG also considers other aspects of the technical framework for resolution and metadata interaction. These include the preparation of a DOI URI specification (the representation specification for DOIs in the web) and parameter passing (specific parameter issues assuming the dx.doi.org style of DOIs and proxy URL). Likely next topics will include the use of these frameworks in conjunction with structured metadata, incorporating relevant development efforts now being carried out in the further development of the <indecs> project and further development of the DOI "Namespace" data dictionary. TWG is chaired by Larry Lannom of CNRI.

8.7.2 RAWG (Registration Agency Working Group).

The RAWG is the forum for development of the rules governing Registration Agencies, including the ongoing evolution of formal Terms and Conditions as specified in the Letter of Intent which all RAs sign. The RAWG includes all approved RAs plus any member with an interest in contributing to this topic. The RAWG is the primary focus for discussions on the development and implementation of the DOI policy infrastructure. RAWG is chaired by Norman Paskin of IDF.

8.7.3 Application Working Groups

The IDF wishes to develop more working groups to encourage development of applications and outreach to other sectors and communities. The RAWG and the TWG serve as the primary policy and technical 'clearinghouses' with the TWG as the place to discuss technical issues and try out experiments and prototypes as needed. These form a foundation layer on which application-specific WGs can build.

Interaction between the RAWG/TWG WGs could be considered "horizontal" and mission-specific/genre/application WGs "vertical". Cross-membership of working groups is, for now, considered sufficient to avoid duplication of effort and exchange of information.

The widespread implementation of DOIs will be facilitated by practical demonstrations. An early example of this was the DOI-X working group (now closed) which developed a prototype for cross-linking, subsequently taken to full-scale implementation by CrossRef.

Successful working groups:

- Must involve active participation by the relevant communities or business sectors
- Must involve a technical facilitator who is familiar with the generic DOI infrastructure and the use made of this in other applications
- Must involve a designated project co-ordinator (may be the same as technical facilitator) who will provide management support but who has no vested interest in one specific commercial outcome.

A successful working group therefore involves a marriage of active involvement by member organizations plus a commitment by IDF to support one or more individuals working specifically on the DOI application development. In many areas we do not yet have sufficient members to form a self-sufficient and adequately funded working group; and there are more possibilities for application working groups than can be currently funded. This presents an opportunity to develop such a group, by active recruitment of interested parties.

The criteria for creation of a working group will necessarily remain flexible; some may require a commitment of a few existing or new members; others may be better run with a larger group of Affiliate status participants; and yet others may work collaboratively with related efforts already under way in other organizations. The driver for the creation of such a group

must come not just from IDF but from at least one active participant in the relevant application sector. The IDF will seed and support the development of such groups over the next year. Some initial suggestions where we already have active interest are:

- Books (with special reference to Electronic Books)
- Learning Objects
- Software
- Business Financial information
- Museum community
- Patent information community
- Datasets
- Music
- Images
- Software
- News information
- Sport results information
- Information Community Relations (to work with the user community to insure that the DOI system is responsive to the needs of scholars, libraries, researchers and other information users)

If you are interested in seeing one of these working groups formed, or have suggestions for other topics, please contact us.

8.7.4 Governance of Working Groups

The Working Group has been envisioned as a key component in the development of the IDF and the Digital Object Identifier ("DOI") since the very inception of the organization. The IDF By-Laws defines the scope of "Work Groups" as follows: "Working Groups shall be established by the Executive Committee to work for a specific period of time on specific technical or organizational issues that come before the Executive Committee. Each Working Group shall have such duties and responsibilities, as the Executive Committee shall specify in the charter creating such Working Group".

Each member of the IDF agrees to abide by the principles expressed in the IDF By-Laws, including the establishment of Working Groups for limited purposes over limited periods of time. If there is a question of the subject matter to be addressed in these Working Groups, that will be addressed by the IDF governing body. Working Groups report, via the co-ordinator, to the IDF Director who will take appropriate action including implementation of recommendations, or any necessary referral of recommendations to the IDF Board.

Only IDF members can become members of a Working Group, and Working Groups are designed to advance the interests of the IDF as a collective organization, to the ultimate benefit of all members. While intellectual property rights are proposed to be retained by the IDF as an organization, they can in appropriate circumstances be subsequently licensed on a non-exclusive and royalty-free basis to each and every member. Therefore, when a Working Group advances the cause of the IDF, it is for the purpose of potentially benefiting every member of the IDF.

8.7.5 Intellectual Property Rights and Working Groups

The IDF is in the best position to ensure that its membership is able to enjoy the full benefits of any prototype application developed within a Working Group. Simply stated, the intellectual property rights that will inevitably arise out of a given Working Group should eventually rest with one entity. If the rights initially rest with everyone (through non-assignment or lack of protection of those rights), there is a possibility that one of the Working Group members (or some other interested party upon learning of the development) will take the steps necessary to obtain the intellectual property rights to that advancement. Should this occur, the other members of the IDF, as well as IDF itself, could be compromised in enjoying the benefits of

their own advancement. The owner could withhold licensing the technology indefinitely or demand payment of significant royalties from IDF and/or the very members that initially assisted in the development of the advancement. No one involved with the IDF wants to see these types of events come to pass.

In order to pre-empt any occurrence of this type, the basis of Working Group activities is that any and all intellectual property rights that arise out of the Working Group sessions will reside with the IDF alone. As a neutral body, dedicated to the non-discriminatory advancement of the DOI, the IDF is in the best possible position to ensure that each and every member gains comparable use and other benefits arising from any intellectual property developed within its Working Groups. This policy provides a defensive measure designed to protect the interests of IDF members and potential members from being injured by a rogue party, rather than an offensive attempt to profit from the retention of intellectual property rights.

It is important to remember that the very purpose of the IDF is to promote the distribution and use of the DOI; the protection and licensing of relevant intellectual property rights may be seen as an indispensable element in facilitating and achieving those objectives. Indeed, the Certificate of Incorporation of the IDF establishes that it will "arrange for the licensing of any trademarks, copyrights or other intellectual property related to DOIs" for a membership "open to all corporations, other business entities, governmental agencies, not-for-profit organizations, academic institutions and other interested parties who support the goals and subscribe to the purposes of the [IDF] and who commit to pay the annual dues of the [IDF]." Thus, the IDF is open to any and all interested parties, and it is obligated to facilitate the licensing of intellectual property between and among those parties and itself. It should also be remembered that IDF is a not-for-profit corporation. Any intellectual property rights retained by the IDF in the products of its Working Groups can be made available to its members and would not be withheld based upon some profit motive inconsistent with its primary objective or charter.

Moreover, the IDF is controlled by the very members that comprise it. The IDF By-Laws establish that the Board of Directors shall "approve licensing arrangements and other contracts entered into by the Corporation in furtherance of its objectives…." As such, the licensing policy of the IDF is controlled by the very members that comprise it, and any "profit" or advantage that is to be gained from ownership or licensing of intellectual property by the IDF is enjoyed by its members generally. Therefore, any concern that the IDF is attempting to "profit" from its retention of intellectual property rights at the expense of its member organizations is misplaced.

8.7.6 Disclosure of information in Working Groups

Members may be concerned about the disclosure of their own proprietary information in the context of a Working Group, especially where Working Groups may be comprised of actual or potential competitors involved in the same or similar industries. For that reason, the IDF does not require full disclosure of its membership in Working Groups or elsewhere; members are free to disclose or not to disclose relevant information (proprietary or non-proprietary) as they see fit. Individual members, therefore, are in the unique position to prevent competitors or potential competitors from being able to profit from disclosure of their own proprietary information by merely limiting its disclosure. However, IDF will also consider the legitimate expectation that a member's disclosure or participation in a Working Group for the benefit of IDF will not ultimately be usable or made available to that member's competitors in any manner that could be directly detrimental to the disclosing member.

Because of the importance and sensitivity of issues of disclosure and intellectual property rights in relation to potential patentability etc., IDF will develop further procedures and agreements at the suggestion of its members to cover any issues not dealt with adequately in the previous paragraphs.

8.7.7 Working Group Agreements

A Working Group Agreement designed to address member disclosure of proprietary information and ultimate ownership of the intellectual work product resulting from the Working Group's deliberations is to be developed. Non-Disclosure Agreements are available between IDF and any parties, and will be incorporated as part of a generic Working Group agreement. Because the Working Groups (and IDF generally) serve as quasi-standards setting bodies in regards to the DOI infrastructure, IDF must remain careful that its intellectual property is not inadvertently based upon existing patents or other protected rights held by individual members. Accordingly, this Working Group Agreement should contemplate patent and other relevant disclosures by participating members (with appropriate non-disclosure safeguards) as a prerequisite for participation. Where use of aspects of existing protected intellectual property may be necessary, the Working Group Agreement should seek to address the terms and conditions upon which those IP rights will be made available to IDF (and its present and future members) and any implications for universal and non-discriminatory availability of the work product of the Working Group (that should be addressed before development proceeds). Members of the Working Group should be required to prospectively disclaim any right or interest in any development by the Working Group at any stage of its progress.

8.8 IDF Email discussion lists

The IDF has established a series of email discussion lists. These lists use a web-based interface and maintain an archive of list messages. For detailed information, see the DOI Mailing List information on the http://www.doi.org/announce.html web page. Some of these lists are relatively active whilst others may be intermittent depending on participation.

8.8.1 IDF sponsored lists

DOI News
Subscribers to DOI News receive monthly announcements from the IDF covering the latest activities and developments. It is not a discussion list. To join, send an email to contact@doi.org with "Subscribe Me To DOI News" in the subject line.

Discuss DOI
General DOI Discussion Forum

DOI-EB
Moderated discussion list for the DOI-EB prototype working group. For IDF members. Requires approval to join. Contact Steve Mooney (s.mooney@doi.org).

DOI-TWG
Mailing list for the DOI Technical Working Group. For IDF members. Requires approval to join. Contact Larry Lannom (llannom@cnri.reston.va.us)

RAWG
Closed discussion list for the committee on the appointment of DOI Registration Agencies. Requires approval to join. Contact Norman Paskin (n.paskin@doi.org)

Members
Private list for IDF Members.

Prefix-Holders
Private list for holders of DOI prefixes.

8.8.2 Related Non-IDF sponsored lists

<indecs>Rights Data Dictionary Advisory Board
Moderated list for communication among the Advisory Board Members of the <indecs>rdd activity. Requires approval to join. Contact n.paskin@doi.org. (now known as Contecs:DD)

ONIX_IMPLEMENT
This related mailing list is maintained by EDItEUR for those who are implementing or planning to implement ONIX International standards (a dictionary implementation of the <indecs> framework, which is particularly useful as the working basis for deriving DOI Application Profiles in the text sector.) To get information, please go to http://www.editeur.org/onix.html or send email to ONIX_IMPLEMENT-subscribe@groups.com.

handle-dev
Private list for Handle System administrators and developers. Requires approval to join. Contact Larry Lannom (llannom@cnri.reston.va.us)

8.9 Third party development

DOI application development does not necessarily require a working group to be formed. We envisage that an increasing number of applications will be developed by individual companies, as value added services which they wish to use for their own purposes or to offer to others.

We suggest that prospective third party developers may wish to contact IDF to discuss their plans, to determine areas of possible common interest, or existing relevant work, which may be underway. Non-disclosure and other confidentiality agreements can be entered into.

IDF can also recommend developers and consultancies with specific knowledge of the DOI system and its components, which may be available to work with third-party developers. At this stage IDF does not have a formal list of accredited developers, but may develop this if there is a demand for such services.

9. Application

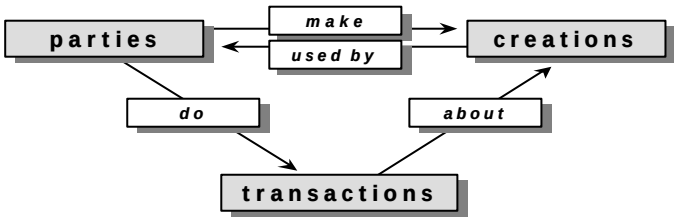
This chapter sets out some guidelines for applying the DOI. What types of entities can a DOI be used to identify? At what level of "granularity" should it be applied? What impact does the DOI Application Profile have on the application of the DOI? Who can apply a DOI? It also describes the actual and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

9.1 What entities a DOI can be used to identify

9.1.1 DOI scope: intellectual property entities

DOI is an implementation of URI (Uniform Resource Identifier, sometimes-called Universal Resource Identifier, IETF RFC2396). It uses the Handle system for resolution of the identifier, and the indecs framework (as implemented in ISO MPEG21 part6) for metadata description, each of which also define their scope as equivalent to that of URI. The specific syntax of the DOI identifier is specified by a NISO standard (ANSI/NISO Z39.84).

While a DOI can therefore be used like any other URI to identify "anything that has identity", the DOI system is a combination of components (identification, resolution, metadata and policies) devised with the specific primary aim of identifying any *"intellectual property entity".* The initial focus of DOI applications was "Creations" - that is, resources made by human beings, rather than other types of resource (natural objects, people, places, events, etc). However these other types of resource are also necessarily involved in intellectual property transactions, and so may be identified by DOIs where appropriate (see the high level indecs model in the diagram below).



As an example, the initial aim of DOI was not to be used to identify natural objects (e.g. specimens in a natural history museum, or natural substances used in pharmaceutical research): but if these were involved in intellectual property interactions there may be an application of DOI to museum artefacts or pharmaceutical components which would be appropriate. Similarly, DOI was not initially an identifier for agreements or licences (which in the indecs framework are types of events), but implementers may find it useful to identify these with DOIs alongside the intellectual property that they govern.

Informally, the 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.

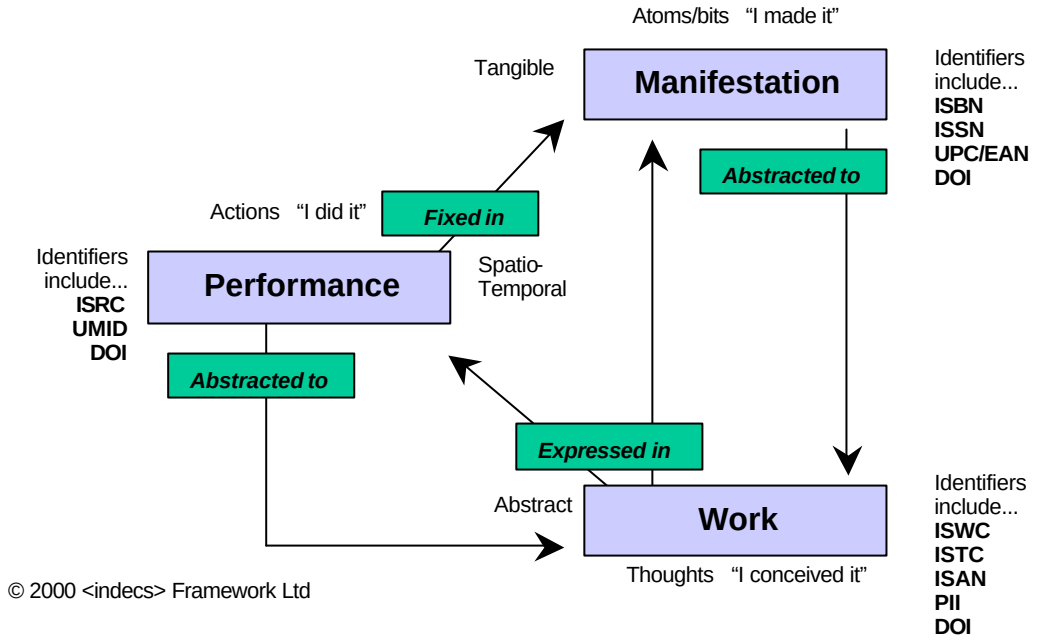
Formally, DOI scope is defined in terms of a data model, the model underlying the indecs/MPEG21RDD work: a DOI can be assigned to any entity which is a Resource within the indecs (MPEG21 RDD) model of e-commerce. This means the type of entity must be described in terms of attributes in the dictionary (e.g., media, mode, content, subject), and become an entry in the indecs Data Dictionary (MPEG RDD) used by the DOI system. The practical outcome of this is important and provides a pragmatic *functional specification:* a DOI can identify any Resource, but the DOI system requires that the Resource is defined (technically and hence precisely) in terms of agreed public (RDD) attributes. This is one role of the DOI metadata.

Within the world of intellectual property entities as resources, the primary focus of DOI has been on the identification of a Creation. The metadata component of the DOI uses the concept of a Kernel set of metadata. The kernel metadata as currently defined relates only to Creations, and a different kernel will need to be defined for fundamentally different Resources or entities such as parties, places or agreements. There is no problem in principle in doing this as the concepts are analogous; it may be a logical and necessary step (e.g. if a DOI Registration Agency wishes to use DOIs to identify individual licence agreements, authors, consumers, etc).

9.1.2 DOI applied to Creations

A DOI is a general purpose numbering scheme for any type of intellectual property, in any medium. As noted above, the initial focus of DOI Applications was to those intellectual property resources that are Creations: resources made by human beings. Creations may be in both tangible and intangible forms/ DOI can be assigned not only to manifestations of intellectual property (books, recordings, electronic files) but also to performances and to "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. In fact there is nothing new in using abstractions or representations in trading – we do it all the time with physical property: representations such as deeds and mortgages are what alters (not the physical bricks etc.) when a house changes hands. Similarly with intellectual property, representations such as licences and files are traded. Digital trading of these pieces of property requires that each entity be uniquely identified. The relationships between these different classifications of intellectual property are summarized in the following diagram developed by the <in*d*ecs> 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), and the associated metadata (see Chapter 5) will provide a definition of the entity being identified. For example, the DOI for an abstract work can (using the DOI System) be resolved to URLs or (better) 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.

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 <in*d*ecs> 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.

Functional granularity should be considered in addressing any question as to application. For example, if a journal publication were to exist in English and Spanish, how many DOIs would

there be per article? There is no simple yes/no answer. This is a "functional granularity" issue, and hence ultimately a decision for the publisher. A publisher could consider the English (E) and Spanish (S) to be different "versions" of the same underlying "work" or "creation" (similar to having both a pdf and html version) in which case one DOI. Or a publisher could consider them two separate underlying works, hence two DOIs. These could perhaps be related in one or more applications using the <indecs> entities and relationships or they could be grouped together under a third DOI for the work. This latter approach is envisioned as a possible future evolution of the DOI system involving multiple resolution, in which a single DOI for the work could be resolved to multiple additional DOIs for versions of the work, e.g., language, and each of those DOIs could further be resolved to multiple locations. Functionally the decision comes down to this; does the publisher wish to distinguish between E and S for any purpose, e.g., to enable certain mirror sites to carry only the Spanish or English versions and not have to carry both. The safe option is always to take granularity down as low as possible (two DOIs), retaining the flexibility to aggregate them in one or more ways at a later date.

9.3 The role of the DOI Application Profile

In practice, decisions as to the appropriate level of granularity, or policies as to rights ownership requirements of registration, may be decided at an Application Profile level (rather than as a global rule applicable to all DOIs), as part of a community effort; and hence individual registrants may not need to worry unduly about these aspects once they have committed to a Registration Agency implementation.

The reason for this is simple: since the scope of DOI is so wide, there can be no one ruling about e.g. “who can register a DOI” which would meet the needs of e.g. the music industry, the image industry, the archive community, at the same time. Rather, each may have legitimate rules applicable to that particular sector and application. Decisions on these issues should be made at the level most appropriate, (the principle of subsidiarity) The DOI system does not aim to impose restrictions on individual communities using DOIs in the way which most suits their business model.

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 Chapter 5). 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 Chapter 5) will develop their own – different – rules relating to the "legitimacy" of eligibility of registrants.

The Digital Object Identifier system is designed for use by all rights holders and users, irrespective of media type. As with any identifier system, once the identifier is publicly available, anyone can use it to ensure accurate and unique specification of the property within their own environment or systems. However only the registrant (who may be a rights holder or an authorized representative), with a formal agreement through a DOI Registration Agency, has access to change the linkage of the DOI.

As an E-commerce building block, the DOI system is capable of being customized to suit specific implementations. For instance, the system's unique multiple resolution function, whereby a single DOI can be linked to multiple objects, such as a music file, the metadata describing the file and perhaps an E-commerce site where the file can be purchased, could be the basis of a comprehensive E-commerce process for sampling and purchasing intellectual property objects. The DOI Foundation is keen to see the DOI taken up by all rights holding communities and can offer assistance and advice on implementation.

10. The role of the International DOI Foundation

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

10.1 DOI business and organizational principles

10.1.1 Cost recovery as the basis of DOI deployment

The implementation of the DOI system will add value, but necessarily incur some costs. The three principle areas of cost are:

- Number registration; maintenance of resolution destination(s); registration of metadata; validation of number syntax and of metadata; liaison with Handle registry; customer guidance and outreach; marketing; administration. Whilst DOI number registration can be automated almost entirely, the control of metadata and maintenance of updated records may require human intervention, as must the other tasks listed here, which must be paid for.
- Infrastructure: resolution service maintenance, scaling and further development. These costs can to some extent be shared across all DOI applications.
- Governance: common "rules of the road"; development of the generic system

Once assigned, DOIs are free in use; whenever a DOI is encountered, it can be "clicked on" and used. The costs of running the system are borne by the registrant.

Our aim is that the system should become self-financing; costs for running the system should be recouped from those who benefit from the system. Cost recovery from registrants, not end-users, is a practical measure: the tasks should be delegated to appropriate organizations who can offer appropriate expertise, economies of scale, synergy with existing operations, marketing presence, etc. (in the case of number registration, these organizations are referred to as "Registration Agencies" (RAs) – see Chapter 11).

10.1.2 The creation of Registration Agencies

When DOIs were launched, no third party DOI Registration Agencies existed. Early users of DOIs interacted directly with the International DOI Foundation; the IDF was therefore the only Registration Agency. It was recognized that in order to be widely deployed, the system needed to be widely distributed, and the model chosen (discussed in more detail below) was to appoint third party registration agencies – essentially holders of a franchise of the DOI system. The IDF then becomes more like a wholesaler and the Registration Agencies become retailers. Registration Agencies, rather then IDF, will become the face of the DOI system as far as the end customer id concerned.

Registration Agencies are now being appointed, in a gradual process of migration to this franchise model. It is therefore necessary to appreciate that the policies and procedures related to DOI allocation are moving from IDF to this network or federation of Registration Agencies. An aim of this section of the Handbook is to provide more information on the consequences of this transition for users of the system.

When a critical mass of registration agencies have been appointed the fees charged by the IDF to existing "direct" prefix holders will be reviewed, as will the rules and procedures applicable to such prefix holders. The IDF will offer limited pricing flexibility and DOI functionality if prefixes and DOIs are registered directly with IDF, and will recommend that DOIs are obtained from approved Registration Agencies (rather than direct from the Foundation) to take advantage of pricing schemes and related applications designed to suit the needs of the particular intellectual property community concerned.

10.2 The DOI business model: DOI costs

At the outset of the DOI development, a very simple model was introduced whereby a prefix assignment is purchased for a one-off fee. A fee was introduced not to cover actual costs, but to recognize the fact that some charging for DOIs would be the intention. IDF used a simple initial economic model: a charge of $1000 for allocation of a prefix (a one-off charge) allowing unlimited number of DOIs to be constructed using that prefix. The current charge is one-off and entitles the registrant to an infinite number of suffixes; there is no annual fee, though we reserve the right to vary this at a future date; there is no limitation placed on the number of DOI prefixes that any organization may choose to apply for. It was recognized at the outset that this fee structure was a starting point but would be insufficiently flexible for the long term.

DOIs allocated using these prefixes purchased directly from IDF are registered without structured metadata: they are “zero Application profile” (see Chapter 5). The disadvantage of using the direct $1000 route is that there is no metadata support and no social infrastructure support of the type, which can be given by a Registration Agency such as CrossRef.

We are now in a process of migration to a wide variety of potential business models, using third part registration agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. The direct prefix purchase route is still an option, but our intention is that eventually all future DOIs will be registered through one of many Registration Agencies, each of which will use one or more defined DOI Application Profiles (see Chapter 5), and each of which is empowered to offer much more flexible pricing structures. The pricing structures and business models of the Registration Agencies will not be determined by the IDF; each RA will be autonomous as to its business model. Business models for these agencies could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc. The IDF will place minimal constraints on the business models offered by RAs, and enter into discussion on practical implementation of any of these (see also Chapter 11 and 13).

The customer should be interested in "what does the retailer charge". An RA will provide a service - e.g. CrossRef. One of the things they will do for that service is allocate a DOI - and the metadata (or help with it, or specify it, or...). But it’s not the ONLY thing they do. So you can't look at the charges of an RA and say "that’s what a DOI costs". Crossref don't charge a single fee but it works out at anything up to 60 cents per DOI allocated.

DOI RAs may find it beneficial to develop new DOI applications for their customers, or to the same market segment, in order to widen the potential for use and income stream from their DOI activities. In some other sectors, products created as a spin off from basic registration activities provide the funding to cross-subsidize and create a low price for registration itself - a "positive feedback loop".

The one-prefix/one fee model is insufficiently flexible not only for business reasons but also for technical reasons. We want to encourage the appropriate use of DOI prefixes without undue financial penalty. For example, to encourage multiple prefixes within a single organisation (this may prove administratively convenient especially in large organizations); or at a different level of granularity (e.g. prefixes allocated to imprints, record labels, image libraries, magazines, journals, etc). A fixed fee per prefix does not do this.

While the migration to these more sophisticated business models is under way, and market development is being undertaken by Registration Agencies, the IDF will continue to offer only a basic and relatively inflexible prefix-based fee structure. We receive many requests for flexibility in prefix allocation or costs. We will attempt to deal with these requests sympathetically but must point out that we have limited options in designing an equitable pricing scheme to suit every need before the appointment of specific registration agencies.

Can DOIs be made available at no charge? Yes.

(a) IDF is willing to allocate a DOI prefix free of charge to organizations for limited experimental non-commercial uses. Please contact us if you wish to apply for this.

(b) The business model includes two separate steps: a business relationship between IDF and an RA (the "franchise fee"); and a business relationship between an RA and a DOI registrant (the "registration fee"). The two are not directly connected; this enables the RA to offer to registrants any business model whatever, which suits its needs. This could include assigning DOIs without charge. Hence DOIs can be used in both commercial and non-commercial settings, interoperably. However, the franchise fee in such an example cannot be zero; this would immediately undercut any commercial use, and it would not provide any financial support for the operation of the system itself. Like any other piece of infrastructure, an identifier system (especially one which adds much value like metadata and resolution) must be paid for eventually by someone. So an organization could, if it wished, assign DOIs freely (registration fee zero to registrants) and subsidize this added-value service by paying the franchise fee to IDF.

10.3 The DOI organisation model

10.3.1 The starting point

The IDF organisation was set up on a similar model to the World Wide Web Consortium (W3C). The funding is from Members, and there is no direct relation to the operational running costs of the DOI system. The member-based IDF currently subsidizes all the operational running of the system specific to the DOI implementation of Handle technology. The members of IDF, as with members of W3C, pay a membership fee to support development of the system as a pre-competitive standards activity, which when widely implemented will enable costs savings or new business opportunities in the community (an analogy: allowing the tide to enter the yacht harbor, enabling each yacht to float free but with no advantage to any one over the others). The IDF is run by its Members, via an elected Board and appointed Director.

The Foundation was created in 1998 and supports the needs of the intellectual property community in the digital environment, by the development and promotion of the Digital Object Identifier system as a common infrastructure for content management. The Foundation is controlled by a Board elected by the members of the Foundation, with an appointed full-time Director who is responsible for coordinating and planning its activities. Through the elected Board, the activities of the Foundation are ultimately controlled by its members. Membership is open to all organizations with an interest in electronic publishing and related enabling technologies. The aim in the long term is to reduce reliance of the DOI system on this membership fee, enabling the fee to be reduced, abolished, or redirected (as the Membership decrees), and seek income directly related to DOI system usage in order to cover costs.

The IDF currently contracts with various technical providers e.g. CNRI and will similarly contract with any other organizations to which operational tasks are delegated. IDF will continue to control the relationship with the global resolution provider, on behalf of all RAs. RAs are free to subcontract or partner with others to deliver part of their services.

The International DOI Foundation, Inc, is a non-stock membership corporation organized and existing under and by virtue of the General Corporation Law of the State of Delaware, USA. The Corporation is a “not-for-profit” organization, i.e. prohibited from activities not permitted to be carried on by a corporation exempt from US federal income tax under Section 501(c)(6) of the Internal Revenue Code of 1986 et seq.

10.3.2 The future Operating Federation organisation model

A different sort of organisation typifies a mature or established standards deployment activity, e.g. ISBN, EAN, VISA. The EAN model is perhaps the closest to likely mature DOI system yet identified. In such a model, the operating entities which deploy the standard (in the case of DOI, we call these Registration Agencies (RAs)) find it advantageous to form a federation structure. The Operating Federation is run by the Agencies via an agreed structure and with appointed central staff.

The "Federal principle" specifies that users of the system make decisions at the lowest appropriate level; a governance layer ensures interoperability of lower levels; the governance layer implies "minimal constraint". From this it follows that each RA is free to determine its own business activities, constrained only by the agreed level of federal governance. This is a market economy model: in a market economy, anybody can trade with anybody, and they don't have to go to a market square to do it. What they do need, however, are a few practices everyone has to agree to, such as the currency used for trade, and the rules of fair trading.

In order to consider the development of such a structure for the DOI system, it's useful to project forward to a stable mature economic model, and then to work back to how this might be achieved.

10.3.3. Economic model for Operating Federation

This is analogous to e.g. the EAN (physical bar code) business model. The purpose of the Operating Federation is to maintain and control the few practices everyone has to agree to ensure the health of the system.

The business model of each RA is determined by the RA itself. The RA enters into agreements with its customers (DOI assigners); these may be radically different between RAs, depending on the needs of its customer communities, its own other business, the value-added services it offers, competitive pressures, etc. The agreements will however require conformance to the overall federal standards.

The RA enters into a contractual agreement with the Operating Federation, guaranteeing the conformance to minimal criteria covering conformance to technical, information management, and economic criteria. The "economic criteria" for being a member of the Federation is a payment to support central Federation governance: a "participation fee", or "franchise fee".

In the mature model, all RAs agree that certain high level or "central" functions are to be carried out not by one RA but by the central Operating Federation organisation on behalf of all. The costs of carrying out these agreed central functions result in an annual operating budget. The budget costs are apportioned across all RAs in an agreed fashion, resulting in the "participation fee" paid by each RA into the central Federation. The sum of the participation fees matches the budget costs. If the proposed central budget is agreed by the RAs (who govern the Operating Federation) to be too high, RAs must either agree to decrease the central functions, or modify their business model to generate the necessary increased participation fee. The participation fee can be viewed as the minimum cost necessary to participate in the system and gain access to the infrastructure, technology, existing brand value, franchise materials, etc.

The percentage of total central Operating Federation costs borne by RA can be determined by a fixed criterion (e.g. in the EAN system, by a formula based on potential market size), or dynamically by agreeing on a pre-set formula (e.g. for DOI it could be a function of number of prefixes, number of DOIs, and number of resolutions).

For illustrative purposes only, let us assume that central Operating Federation costs in a mature (long-term) model are $2M p.a. If there were 50 RAs, then average costs per RA =

$40,000 (which seems not unreasonable - each RA is bearing costs equivalent to employing a middle-order administrative person). In the early days of setting up RAs, this model alone cannot support the whole system. If there were for example only 2 RAs, it would not be feasible for them to apportion central costs of $2M between them (= $1M each). A more realistic course is for the Operating Federation to come into existence but for its costs to be derived from two sources:

- the Operating Federation model, fixing an arbitrary but affordable participation fee per RA (for illustration., say $50K); and
- the (existing) membership model.

As the number of RAs increases, the proportion of costs contributed from Operating Federation increases and from Membership declines. Consider a possible migration path (this is necessarily a simplification for illustrative purposes, assuming simply one fixed participation fee, but the same principles can apply to a more dynamic cost apportioning model and multiple year budgets)

- the total IDF costs are $2M (illustrative figure only);
- 14 agencies agree to a participation fee of minimum $50K each;
- the remaining sum ($1.3M) is met from the "membership" model

These figures ignore the effect of volume charges for number of DOIs beyond the minimum; these charges will substantially increase the revenue derived from RAs and decrease the need for member-based subsidy. A more detailed financial model is used in practice to allow for these factors.

In practice, it would seem likely that a variation on this model would be feasible. Some areas of development costs may not be essential to the existing Operating Federation, but some potential RAs or members may want to see these developed: e.g. future functionality beyond that already provided. A structure can be conceived which supports both and where governance is via two "chambers", embracing both the current organisation model (for "Development") and the Operating Federation model. An organisation which is both a Member of the Foundation and an RA (participant in the Operating Federation) would receive some benefit (e.g. votes in both "chambers" but probably also some reduction in one or other fee).

We have begun the transition to this form of model by the creation in 2001 of a new category of membership, that of “Registration Agency”. Registration Agencies currently have 25% of the seats on the controlling board of the Foundation. It is the intention that this percentage be increased in proportion to the percentage of revenue derived from RA operating activities. We have evidence that this transition is under way. In 2000, we had no income from Registration Agencies (as of course we had no formal RAs). In 2001, we budgeted (and achieved) 4 RAs, and RA fees produced 14% of the Foundation’s income. In 2002, our budget is for a further 4 RAs and the current forecast is that RA fees will provide 22% of the Foundation’s income. Of course, the timing of when this figure will reach 100% is dependent not only on numbers but on volume: to take a simplified example, if our fixed costs were $2M and each RA produced the current minimum of $50K, we’d obviously need 40 RAs to recoup costs totally: but in fact, our first RA (CrossRef) is already producing about twice that amount of income due to the large volume of DOIs they assign (i.e. “20 CrossRefs” would mean a balanced budget). Undoubtedly there is a chicken-and-egg, or threshold, effect, as the success of DOIs in one area encourages new applications.

10.4 IDF and relationships between Registration Agencies

The organizational model outlined here provides a clear basis for the relationship between end customers (registrants) and Registration Agencies; and between RAs and the Operating Federation.

There is however another set of relationships, which needs to be considered, between the various Registration Agencies themselves. In the Operating Federation model as implemented in e.g. EAN or ISBN, each RA has a geographical basis. Although customers are free to choose which RA to use, in practice most may go to one familiar to them, a local language agency. In the digital world, it is not clear whether such a basis is appropriate. In favour of such an arrangement is the need for language-specific related materials and support (e.g. local language guideline materials, helpdesk systems, and potential specialized consultancy staff). Arguing against such an arrangement is the fact that in a digital world, geographical barriers are less important, and an arrangement focussed on content sector or content type may be more effective. In some major markets (e.g. the English language markets) it could be possible that the intellectual property sector approach will be favored; whereas in smaller language markets, a geographical (or at least linguistic) basis may be more appropriate. Initial RA appointments made by IDF include examples of each.

The IDF Registration Agency Working Group has been set up to deal with issues such as this, and agree on common principles which foster a climate in which working as a registration agency is attractive, yet any long term monopoly is avoided. For further information on the work of the RAWG and more detailed documentation, please refer to the IDF. RAWG participation is limited to members of the IDF.

10.5 The IDF as the DOI Maintenance Agency

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). 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. This Handbook is an exemplar of the role of the IDF in promulgating a common and consistent approach across all Registration Agencies and users.

IDF 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. IDF maintains formal and informal alliances and strategic relationships with a number of standards bodies and other organizations; for a current list refer to the DOI web site. With respect to internal DOI standards, the IDF acts as final arbiter.

11 The appointment & role of Registration Agencies

This chapter explains the role of Registration Agencies (RA) 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.

For information on choosing a RA, see Chapter 13. For further information on organizational and business principles, see Chapter 10.

11.1 The role of DOI Registration Agencies

11.1.1 Overview

At the outset of the DOI development, a very simple model was introduced whereby a prefix assignment is purchased for a one-off fee. A fee was introduced not to cover actual costs, but to recognize the fact that some charging for DOIs would be the intention. IDF used a simple initial economic model: a charge of $1000 for allocation of a prefix (a one-off charge) allowing unlimited number of DOIs to be constructed using that prefix. The current charge is one-off and entitles the registrant to an infinite number of suffixes; there is no annual fee, though we reserve the right to vary this at a future date; there is no limitation placed on the number of DOI prefixes that any organization may choose to apply for. It was recognized at the outset that this fee structure was a starting point but would be insufficiently flexible for the long term.

DOIs allocated using these prefixes purchased directly from IDF are registered without structured metadata: they are “zero Application profile” (see Chapter 5). The disadvantage of using the direct $1000 route is that there is no metadata support and no social infrastructure support of the type which can be given by a Registration Agency such as CrossRef. DOI prefixes obtained directly from IDF may however be useful if you wish to experiment or consider developing your own applications

We are now in a process of migration to a wide variety of potential business models, using third part registration agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. The direct prefix purchase route is still an option, but our intention is that eventually all future DOIs will be registered through one of many Registration Agencies, each of which will use one or more defined DOI Application Profiles (see Chapter 5), and each of which is empowered to offer much more flexible pricing structures. The pricing structures and business models of the Registration Agencies will not be determined by the IDF; each RA will be autonomous as to its business model. Business models for these agencies could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc. The IDF will place minimal constraints on the business models offered by RA's, and enter into discussion on practical implementation of any of these (see also Chapter 10 and 13).

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 3).

11.1.2 Development of Registration Agency role

The IDF is moving to full deployment of the DOI in co-operation with Registration Agencies and potential RAs (primarily through its Registration Agency Working Group (RAWG)), 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. The formal agreement between the IDF and RA is the remit of the RAWG and a detailed Terms and Conditions document is available: this is work in progress and subject to modification. The followings sections provide a summary; note however that the more detailed Terms and Conditions document takes precedence in the event of any apparent conflict.

11.1.3 RA technical requirements

- 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 (e.g. CNRI's Handle System) 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

11.1.4 RA information management requirements

- 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.

11.1.5 RA general requirements

- 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.

11.1.6 IDF responsibilities

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

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 (see Chapter 10). 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 are detailed above.

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 3).

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. Criteria for Registration Agencies (RA) acceptance are not yet comprehensively defined, as it is inappropriate at this stage to lay down over-rigid specifications. Some major criteria, which will be taken into account when considering applications, are:

- Commitment to membership of IDF and a role with the Registration Agency Working Group (RAWG);
- Acceptance of non-exclusive terms;
- Submission of a detailed written proposal addressing the following areas:
- The 'community of interest' that is to be represented should be identified and delineated.
- A description of the proposed application for the DOIs to be registered: this should be at minimum a text description with some explanation. Ideally, but not compulsory, it should be accompanied by a detailed technical specification on the form of a DOI Application Profile; the latter step will be easier when the DOI Namespace (data dictionary and registry) is available but such profiles can certainly be generated at present especially where the field is covered by existing well-defined metadata schemes such as ONIX;
- An outline description of the business model, describing in particular how registrants will be engaged with the RA and on what terms in relation to DOI application and use, along with a revenue model and plans to establish the appropriate consultative framework within the community of interest: it is not necessary to provide a detailed view of business aspects not related to DOI registration and use, but it is important that the IDF have a perspective on the likely way in which the DOI system will be perceived by the RAs users;
- Evidence of the financial viability of the applicant organization;
- Evidence of the ability of the organization to enforce policies such as persistence of identification;
- Evidence of appropriate technical ability and understanding of the DOI system (both Handle and metadata aspects). 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. In consultation with the Director of 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.

Although some of these criteria are subjective, they provide enough substance to engage in serious dialogue, and the fact that a completely comprehensive list is not available is not an impediment to progress. At this stage it is essential to retain a degree of judgement and flexibility: there is no doubt that more RAs are welcome, but these must be chosen on so as not to prejudice future development. If only some of the criteria can be met by a candidate, the IDF can (at its discretion) offer a limited acceptance of e.g. 60 days during which all the remaining criteria are to be met.

11.4 Formalizing the relationship between an RA and the IDF

11.4.1 Current procedures

If an organization wishes to be involved in determining RA status or applying to be an RA, the way forward is to make an initial commitment by joining IDF; this allows participation in the Registration Agency Working Group and access to all IDF materials, working with IDF members in supporting and developing the system. The IDF is in a period of great expansion and development, and members enjoy early access to development information and determine the course of our work.

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 (for more information contact the IDF, info@doi.org). The letter of intent also commits the Registration Agency to working with, and developing, the document on "Terms and Conditions for DOI Registration Agencies" which is part of the IDF's Registration Agency Working Group.

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. Each RA letter of intent contemplates that a definitive agreement will be developed between the parties to delineate their respective rights and obligations and the initial letter of intent provided that the parties would endeavor to develop that definitive agreement in 2001; this was subsequently extended to 2002 to take account of the need for further deliberations by the RAWG.

11.4.2 Future developments

IDF aims to fashion RA member arrangements that will continue to encourage maximum participation and contributions to/development of IDF projects, while allowing IDF to retain broad flexibility for universal availability of its developments and translation of those developments into commercial products (which the Contecs:DD project presages). RAs will be particularly involved in such agreements as they are likely to participate in working groups, to be a formal RA, and to have considerations of intellectual property rights. Each of these areas is to be the subject of a detailed agreement in the foreseeable future:

1. A Working Group Agreement designed to address member disclosure of proprietary information and ultimate ownership of the intellectual work product resulting from the Working Group's deliberations is to be developed (see Chapter 8 for further information).

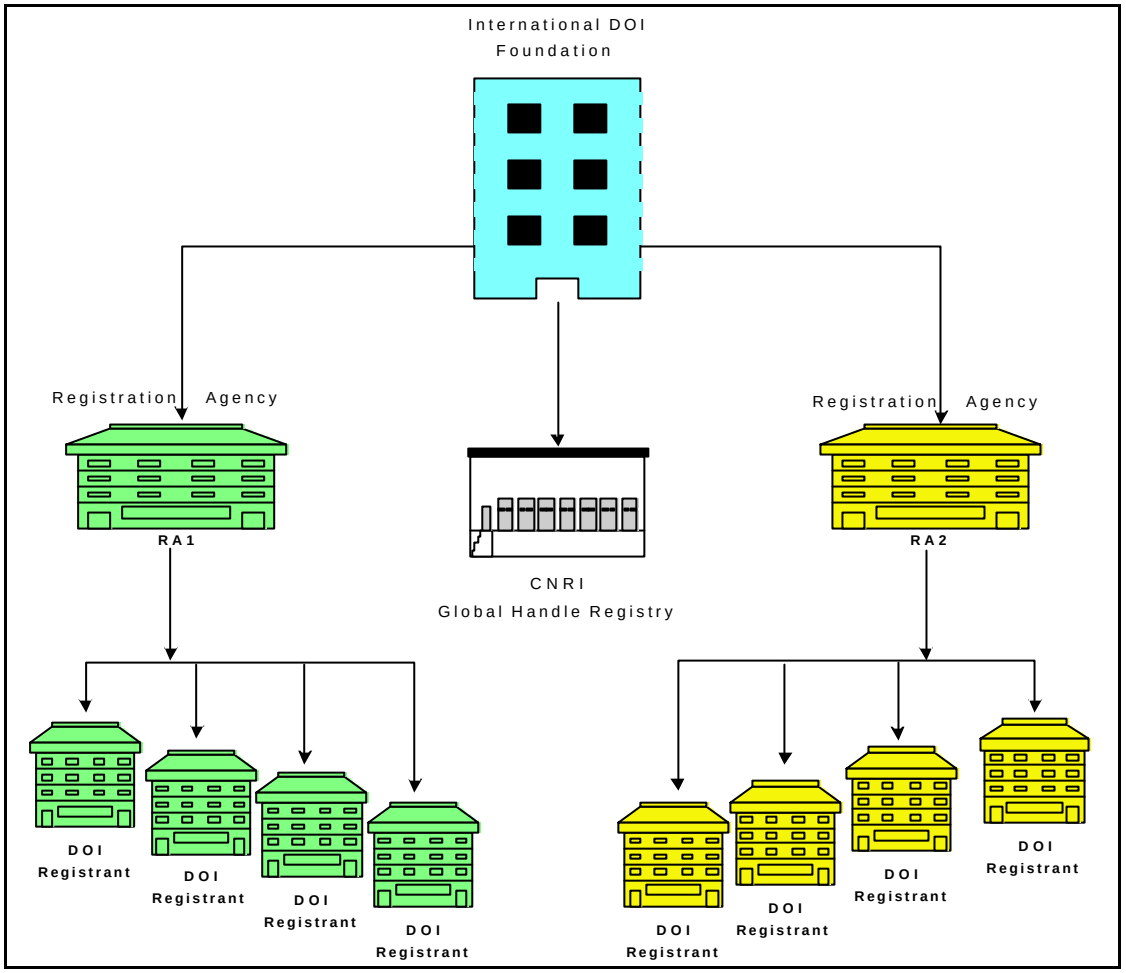
2. A definitive Registration Agency Agreement implementing and expanding upon the terms of existing letters of intent and incorporating IDF's policies (see Chapter 7) is to be developed. As noted in chapter 7, there are competing interests involved in the issue of exclusivity, including that (a) grant of exclusivity is in tension with IDF's objectives of constructing and maintaining a DOI infrastructure that provides availability on a universal and non-discriminatory basis, and (b) a prohibition on granting any form of exclusivity rights may have the effect of significantly limiting IDF's ability to attract participation by desirable Registration Agencies or may render that participation more problematic. Because of these factors and the evolving nature of the organization, a flexible approach is necessary. However these competing interests are reconciled by IDF with respect to each RA's special circumstances, any individual rights granted to a particular RA should be reflected in the definitive Registration Agency Agreement executed with that RA. Any exclusivity arrangements will be incorporated into the definitive Registration Agency Agreement (as well as any corresponding reciprocal ancillary arrangements that will be granted to IDF by the RA).

3. A template Member Licensing Agreement will be developed that can be employed whenever a member (including a Registration Agency) desires to utilize intellectual property owned or developed by IDF.

We also continue to favour development of a document detailing the rights and responsibilities of IDF members that wish to utilize IDF intellectual property in business (particularly in light of the intellectual property advances being made on Contecs:DD and the Resource Metadata Set template). This agreement would also be of assistance in developing IDF members positions on licensing of technology.

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. 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 Use of DOI Trademark in Domain names

"DOI" and "doi.org" are registered trademarks of the International DOI Foundation. Trademark rights are granted through registration, not use. Thus, trademark registration is required for each individual country in which the mark is to be protected; the only exception being the European Union, which allows a single registration to cover all 15 member nations. The IDF currently owns the "DOI" trademark in the United States and the European Union.

Ideally to protect these trademarks worldwide, the IDF would register for trademark protection and the corresponding "DOI" related domain names in each national jurisdiction in which it does significant business; however, it is impractical to do so, since the time and expense associated with registration in those national jurisdictions is prohibitive. Also the IDF cannot "pre-empt" "DOI" domains in every country, since the number of variants is endless (doiarticle.com, doiregister.com, doigovernment.com etc). The IDF registers instead only generic domains such as doi.info (in this particular case, someone else, who did not have the right to use the "DOI" trademark, tried to register it; this was challenged and won by IDF under ICANN rules). It is also worth noting that in some countries, local rules mean that businesses in that country are the only bodies who can register appropriate domain names.

The trademark "DOI" may, therefore, form part of a domain name used by a Registration Agency under defined conditions. The granting of exclusive use of "DOI" related domain names to any RA in their own country however, would make the entrance of future competitors in that country less likely and more difficult.

Therefore the IDF may permit an RA to register a "DOI" related domain name in their geographical region, so long as an enforceable agreement is executed between the RA and the IDF, permitting the IDF to reclaim the domain at any time (including the termination of its status as an RA or the entrance of an RA competitor in that region) along with any intellectual property rights related to the "DOI". This will promote shared use of "DOI" related domain names in the event that competitor, or competitors, should arise in the relevant geographical region.

The Registration Agency Agreement, which is currently being prepared, will be the correct place to include such protections; until that document is executed however, any domain name registrations should be preceded by the signing of an interim agreement.

11.7 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 (see Chapter 10).

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. An initial fee structure agreed by the IDF Board was in place until 31 December 2002. From 1st January 2003, this has been revised, following consultation with the IDF Board and the RAs. The fee structure will be subject to future review, however we will ensure that changes are predictable and fully discussed with the IDF Board, on which the Registration Agencies are 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 $35,000. 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. For DOIs registered in 2003, the franchise fee will be calculated as US 3 cents per DOI registered and for DOIs registered in 2004, the franchise fee will be calculated as US 4 cents per DOI registered, with a minimum and maximum as noted below. Accounting will be made on 1st January and 1st July each year (i.e. on 1st July, payment is due for all DOIs newly registered in the last six months). Each Registration Agency will guarantee a *minimum* payment of $20,000 each year irrespective of the number of DOIs allocated.

- Each Registration Agency will have a *maximum* franchise fee ceiling of $100,000. In a situation where a Registration Agency generates a franchise fee of more than $100,000 in one year, the fee will be renegotiated.

- A maintenance fee will also be charged. The fee will be on the following basis:

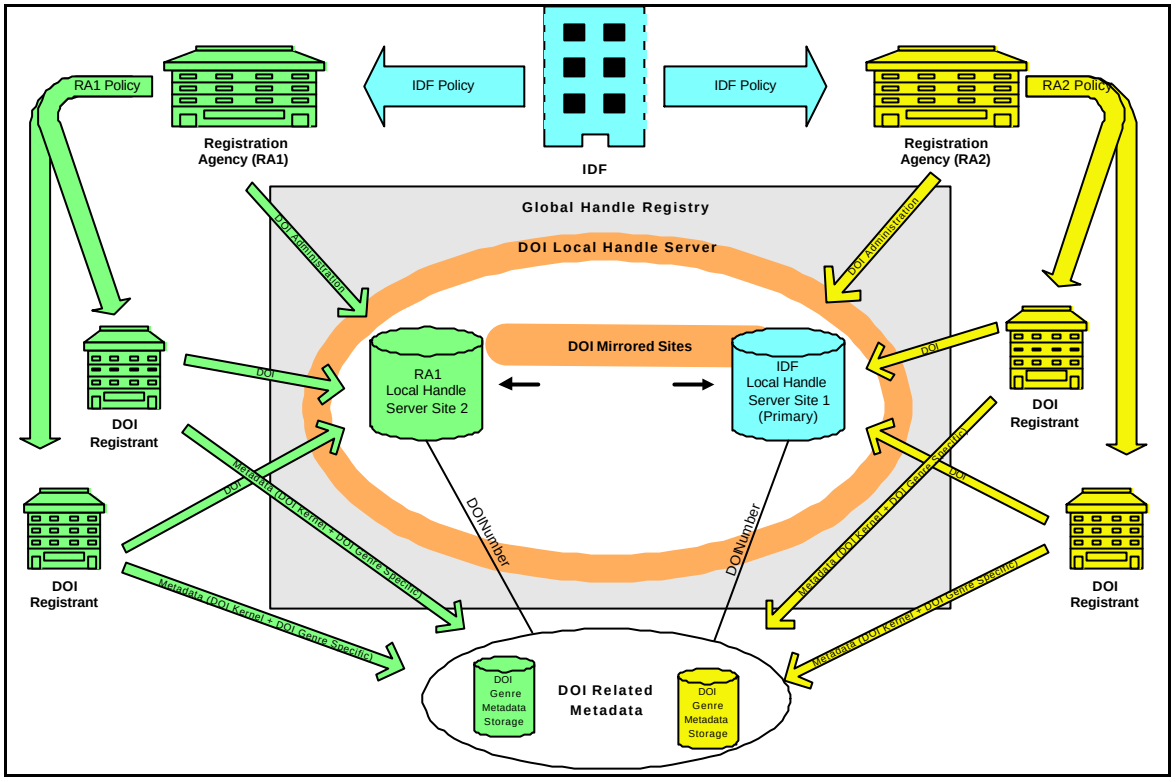
    - $0.01 for the first 5 million DOIs previously registered
    - $0.005 for DOIs beyond the first 5 million

  At the beginning of each year, the number of DOIs previously registered will be calculated to the end of the previous year. The annual maintenance fee will be calculated based on the above levels and will be invoiced in two parts, in January and July of that year along with the relevant franchise fees.

11.8 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 3).



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.

11.9 An overview of the roles of RAs and IDF

This is not an exhaustive inventory of roles but intended to indicate the outlines of who does what and how the roles may interact. A more formal specification is in development by the IDF's Registration Agency Working group.

IP registrants

- Register with one or more registration agencies
- Ensure appropriate content management of own material (maintenance of own URLs and metadata), either directly or by contract (e.g. with RA)

Registration Agencies

- agreements with IP owners
- registration of prefix and individual DOIs with DOI system
- definition of appropriate Genres and mapping to other appropriate metadata sets
- metadata collection /added value
- provision of data to Value Added Services by agreement with IP registrants, etc (may include services run by the same organisation)
- marketing of DOI and related services
- training - guideline development, etc.
- (optional) maintenance of Handle mirror site
- (optional) subcontract service provision

These areas may be specific or customised to one Registration Agency, provided only that top level minimal common specifications are followed.

IDF

- minimal common agreements for registration agencies
- provide DOI resolution service
- ensure resolution service integrity and performance
- maintain Data Type Registries (fields within handle specifying different multiple resolution results)
- provide DOI-specific guidelines on scope, metadata implementation, etc., sufficient to enable RAs to prepare appropriate marketing, guidelines, and training materials for their own use.
- maintain DOI Genre (application profile) inventory
- provide basic information necessary for marketing materials etc
- policies e.g. archiving, testing, etc
- interaction with other related activities, standards bodies, etc.
- promotion and further development of the generic DOI system
- governance administration and agreement with RAs on division of tasks
- training/advice for RAs on content management issues relating to DOI (e.g. use of metadata model)
- role in governing body (operating federation)

11.10 Registration Agency Local Handle Service (LHS) Operation

IDF operates a Central DOI Directory as a master directory of all assigned DOIs, which may also be used as a Local Handle System by IDF Registration Agencies. However in execution of their DOI related services, IDF Registration Agencies (RAs) may elect to run their own local handle service (LHS) rather than depositing DOIs into the handle service run by the IDF. In order to do this, the RA will need to comply with certain requirements set forth by the IDF to ensure the integrity of the DOI System and provide uniformity across RA resolution services. This section outlines the issues involved in setting up a RA managed local handle service.

11.10.1 Why use a Local Handle Service at an RA?

A number of issues have been highlighted in discussion with existing or potential RAs, which have led us to introduce this possibility. These include:

- Control: RAs ideally wish to have immediate control of business-critical infrastructure components such as DOI registration.
- Performance: RAs wish to have high performance standards for administration by depositors and resolution by users, with the ability to choose levels of performance standard appropriate to their application.
- Escrow: IDF needs to have an escrow copy of the DOI deposit data; RAs are willing to provide this but doing so should not impede their business.
- Business model transition: IDF and RAs wish to move more rapidly to an "operating federation" structure, where the bulk of costs and operational income are borne by the RAs rather than by IDF. The fundamental assumption that "membership fees support development until operating federation can take over" is assumed to be still valid, but we seek to encourage the migration.
- Ability to fund growth: as RAs grow, they wish to ensure appropriate growth of both LHS server capability. IDF wishes to ensure sufficient funding for the underlying Global system to ensure high standards for all LHS use.

These issues can be addressed by allowing (but not mandating) that RAs themselves operate a Local Handle Service for their DOIs, mirrored to an IDF-operated DOI LHS. The advantages of this are:

- RAs are in a better position with regard to control, performance etc.
- IDF receives escrow data, but this process is no longer on the critical path of business performance.
- Funding of infrastructure growth to accommodate growing use becomes more of an issue for each application and less a bottleneck and funding problem for IDF (though load-intensive applications, increased numbers of DOIs, increased amount of information per DOI or increased frequency of change of DOI would necessitate some corresponding increase in mirror resources).

11.10.2 DOI Handle Service configuration

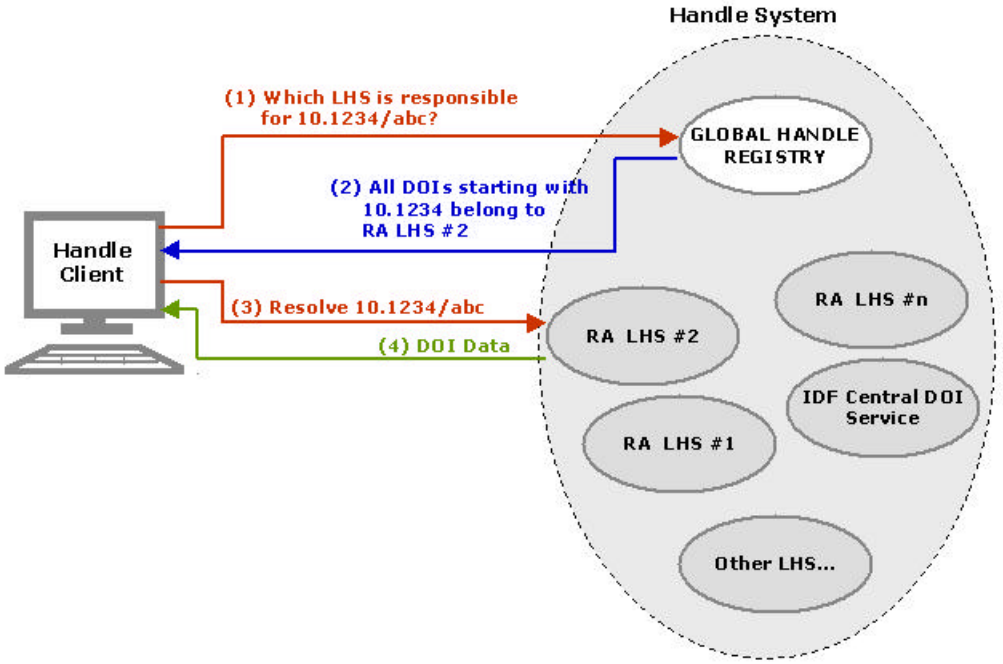
Currently, the IDF manages one local handle service known as the DOI Service or DOI Directory. This Service consists of multiple servers and contains the DOIs for the several hundred existing DOI prefixes, also known as naming authorities. Each prefix within the handle system is associated with service information, which indicates where handle clients should go to resolve a DOI beginning with that prefix. (see Figure 1). All DOI prefixes currently point to the handle service, which is a collection of servers, run by the IDF. Allowing multiple handle services, under the management of multiple RAs, will require re-pointing the prefixes to the new handle services as well as the introduction of new policies and procedures to maintain the integrity of the DOI system.

Handle System prefixes are not, for most purposes, hierarchical; they work as peers. Thus, there is no need for DOIs that start with, for example, 10.1234/ to necessarily be co-located

on the same handle service as DOIs, which start with 10.1234.1/. It is, however, important for all DOIs that start with any single prefix to all be co-located on the same handle service. While it would be theoretically possible to have DOIs starting with any given prefix split across multiple handle services run by multiple RAs, it would be very inefficient since it is the prefix, or naming authority, which is used by handle clients to locate the correct handle service for the DOI in hand. There is no limit to the number of prefixes for which a single handle service may be responsible. (For detailed information on the Handle System see Appendix 4 and www.handle.net.)

In the case of a prefix holder needing to transfer DOIs in their entirety to another RA's handle service the service information in the prefix record would simply be updated by the administrator of the Global Handle Service (currently CNRI) and all subsequent resolution requests for DOIs starting with that prefix would be directed to the new service. It would be the case of a prefix holder wishing to move only some DOIs starting with a given prefix to another RA handle service that would be difficult. This could be the case, for example, if a publication were sold to another publisher and the old and new publishers used different RAs, each of which ran their own handle service. One approach would be an agreement between the two RAs. Each DOI would be modified to give administrative permission to the new owner but the original owner's RA would have to keep their service running. The new owner could elect to redirect each DOI to a new DOI (in the new service) or modify the DOI data, e.g., the URLs, directly. The best approach, however, would be to avoid the problem by increasing the granularity of naming authorities, e.g., one per imprint or journal rather than one per publisher, decreasing the likelihood of splitting a prefix across services.

Figure 1

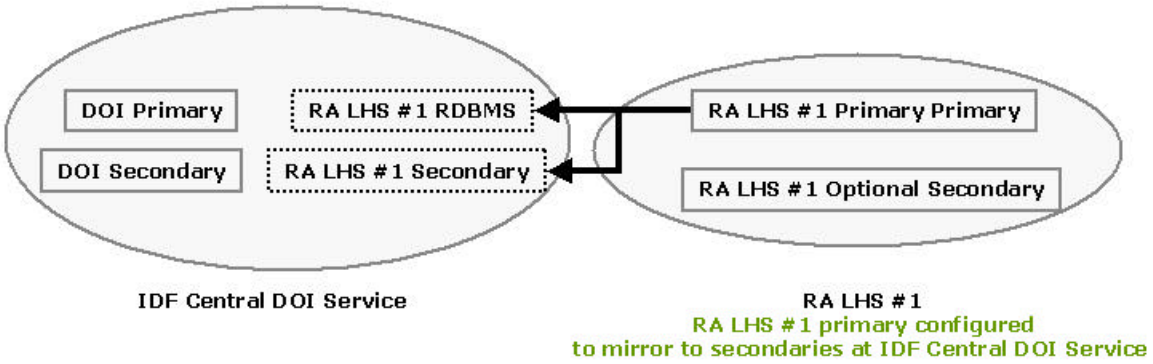


11.10.3 RA responsibilities

In order to guarantee a high level of service quality across the entire DOI system, the IDF will impose certain technical requirements on any RA running a separate handle service for the resolution of DOIs. Each separate RA handle service will have to mirror its DOI data back to the IDF's central DOI service, currently managed by CNRI. Every handle service is comprised of one or more sites and each of those sites can be made up of one or more servers. One of the sites is always designated Primary and all administration must go through that site, with

the data being mirrored to other sites, called secondaries, within the service. Each RA running their own service will, at a minimum, have to maintain the following configuration: one primary site, a secondary site at the IDF facility, and a third site in the form of a relational database, also at the IDF facility. The second site can be used for resolution but will also be considered escrow data. The third site (the relational database) will not be used for resolution but rather for housekeeping. Using basic SQL queries, the database will be used for generating reports for statistical analysis and billing. Each RA will have to work with IDF staff to make sure their configuration includes these two sites. The computers for these servers will be set up and managed by the IDF. The RA responsibility will primarily be to configure their local handle service to mirror back to the IDF facility appropriately. Please see figure 2 for an illustration of the configuration. Each RA will also have to appoint a chief technical contact and provide full contact details to the IDF. It will be the duty of the RA to update this information should the contact change.

Figure 2



11.10.4 Prefix allocation and service configuration

Each RA will be allotted a certain number of prefixes. These prefixes will be a block of sequential numbers with no special meaning as mentioned earlier. Each RA will need to understand the basic set up of their prefixes. Each prefix has associated with it service information. This service information is the map/layout of the handle service. The service information is incorporated in a service handle (another level of indirection for ease of administration). Each RA LHS will have its own service handle. The RA will need to know and understand the service handle as it uniquely describes their service. The service handle will be used as the RA 'account' number for all correspondence with the IDF and CNRI.

A prefix will have the following values: (Example)

Prefix: 0.NA/10.1201

Data Value:100: HS_ADMIN 0.NA/10:200(CNRI admin that points to another handle)

Data Value:101: HS_ADMIN 0.NA/10.1201:200(RA admin that points to a group within the prefix record - see next in the list)
Data Value:200:HS_VLIST (Group/list of administrators)
0.NA/10.1201:300
10.1.admin/user1:300
Data Value:300:HS_PUBKEY(Public key for local server administration)
Data Value: 1:HS_SERV 0.SERV/10.1(Service handle)
Service Handle: 0.SERV/10.1 (includes information for each site in a service such as ip address and port numbers). CNRI will create the service handle for each RA LHS.

CNRI will maintain administrative permission for the prefix as well as the RA. This is intended as a backup for administration.

Administration of DOIs will require the use of an admin DOI. Each admin DOI will begin with 10.1.admin/ or similar, differing slightly for each RA's LHS.

It is the responsibility of the RA to inform CNRI and the IDF in the event of any major transaction that could possibly interrupt the mirroring mechanism. It is also the responsibility of the RA to inform CNRI of any configuration changes in their LHS.

The RA is required to enter into a Handle System Technology Commercial Sublicense Agreement with IDF to use a LHS. IDF has licensed rights from the Corporation for National Research Initiatives to technology and software, including copyright in Local Handle System software, U.S. Patent No. 6,135,646, the Handle System® and Global Handle Registry® trademarks, and certain know-how relating to the Handle System Technology, and pursuant to its Handle System Technology Commercial License Agreement with CNRI, IDF is authorized to sublicense limited rights to the Handle System Technology. Please contact IDF for further information.

11.10.5 Issues arising

A number of issues relating to the implementation of LHS by RAs are currently under discussion within the IDF's working groups. Included in this are these considerations:

- RAs would bear more of the infrastructure costs, and should therefore receive a corresponding benefit.
- RAs gain benefits noted above and may also be able to develop additional benefits from having local access to the Handle System.
- The emerging business model for Global Handle funding (under development by the Global Handle System Advisory Committee) needs to accommodate this DOI model.

The precise balance of costs and benefits is under discussion. A starting point for consideration is that fixed costs such as RA compulsory membership in IDF be payable to IDF, whereas some variable costs (such as LHS infrastructure provision) and income (such as DOI registration fees) be retained by the RA or a cost billed by the global HS operator (like a utility model). This is a refinement of the basic business model discussed in 11.2.

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. 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 in Chapter 1.

The IDF has a Board of fourteen 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

The expenses of managing and operating the DOI System are supported by the IDF. The IDF itself is supported by (a) membership subscriptions, to enable the IDF to be funded through its development phase; (b) funding from Registration Agencies fees. The IDF business model is an operational federation (of Registration Agencies) gradually assuming responsibility from a membership-based organisation paralleled by funding gradually moving largely to funding from Registration Agencies' fees.

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 direct from IDF (i.e. other then via a Registration Agency). Unlimited DOIs can be constructed using that prefix; however no additional support is offered. This fee was not intended to provide continuing or sufficient revenue to support the DOI in the long term, but to partly defray basic costs. This fee structure is insufficiently flexible; a more flexible structure is offered by Registration Agencies, who also offer additional infrastructure support and guidance in the use of DOIs. Our aim is to encourage the appropriate use of DOI prefixes without undue financial penalty: for example, to encourage multiple prefixes within a single organisation, or at a different level of granularity. Registration Agencies provide DOI prefixes as part of their overall package of services to Registrants. The charge and process for obtaining a DOI Prefix as part of this service is a matter for individual Registration Agencies: it is highly unlikely that a single business model for DOI allocation costs will be uniformly applicable. The IDF wishes to encourage a migration to a wide variety of potential business models, in recognition of the fact that there will be no "one size fits all" solution. Business models could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc.

Once assigned, DOIs are free in use; whenever a DOI is encountered, it can be "clicked on" and used.

12.4 Future funding of the DOI System

The aim is 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 - The appointment and role of Registration Agencies)

Ultimately, the DOI System will be funded entirely through its commercial applications (the operating federation of Registration Agencies). Funding of the total system in this way is not possible until a critical mass of agencies has been built up, and therefore membership funding is continuing to defray part of the costs until that point is reached.

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).

However we still have a small number of RAs to carry any additional burden of HDL costs. We don’t see RA funding as a solution to cost recovery in the short term.

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 4 of 14 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).

13. Operating procedures

The procedures and the rules for the registration and maintenance of DOIs and associated metadata will differ between different Registration Agencies. This section provides an outline of current procedures.

13.1 How to obtain a DOI Prefix

At the outset of the DOI development, a very simple model was introduced whereby a prefix assignment is purchased for a one-off fee. A fee was introduced not to cover actual costs, but to recognize the fact that some charging for DOIs would be the intention. IDF used a simple initial economic model: a charge of $1000 for allocation of a prefix (a one-off charge) allowing unlimited number of DOIs to be constructed using that prefix. The current charge is one-off and entitles the registrant to an infinite number of suffixes; there is no annual fee, though we reserve the right to vary this at a future date; there is no limitation placed on the number of DOI prefixes that any organization may choose to apply for. It was recognized at the outset that this fee structure was a starting point but would be insufficiently flexible for the long term.

DOIs allocated using these prefixes purchased directly from IDF are registered without structured metadata: they are “zero Application profile” (see DOI Handbook Chapter 5). The disadvantage of using the direct $1000 route is that there is no metadata support and no social infrastructure support of the type, which can be given by a Registration Agency such as CrossRef.

We are now in a process of migration to a wide variety of potential business models, using third part registration agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. The direct prefix purchase route is still an option, but our intention is that eventually all future DOIs will be registered through one of many Registration Agencies, each of which will use one or more defined DOI Application Profiles (see Chapter 5), and each of which is empowered to offer much more flexible pricing structures. The pricing structures and business models of the Registration Agencies will not be determined by the IDF; each RA will be autonomous as to its business model. Business models for these agencies could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc. The IDF will place minimal constraints on the business models offered by RA's, and enter into discussion on practical implementation of any of these (see also Chapter 10 and 11).

To obtain a DOI Prefix, you need to apply to a DOI Registration Agency. In January 2003, the DOI Registration Agencies operating are CrossRef (www.crossref.org), Enpia Systems (www.enpia.com), Content Directions Inc (www.contentdirections.com), Learning Objects Network (http://www.learningobjectsnetwork.com/), Copyright Agency Ltd (http://www.copyright.com.au/), The Stationery Office (www.tso.co.uk) and the IDF itself. Additional DOI Registration Agencies will be coming on stream during 2003, and up-to-date details can be found on the DOI web site.

13.1.1 Selecting a Registration Agency

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.

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.

Registration Agency	Application Profile[1] / Area of Coverage	Contact for more info
IDF	• Zero AP (DOIs with no metadata) • All areas for prototype and experimental work	info@doi.org
CrossRef	• Scientific journal articles and related areas	www.crossref.org
Content Directions Inc	• Consulting on all areas of DOI application • Partnerships with content mgmt vendors, DRM vendors, etc. • Books (print/electronic), magazines, newspapers • Photographs/Images (stills, movies, video) • Music and sound • eLearning (degree/non-degree education: K-12, college, graduate, vocational, ongoing) • Physical product information, catalogs, B2B information • Medical records/other database records •	www.contentdirections.com
Enpia Systems	• Korean language	www.enpia.com
Learning Objects Network	• Learning objects (digital items for re-use in training and education)	www.learningobjectsnetwork.com
Copyright Agency Ltd	• Licensing of material from authors, journalists, visual artists, photographers and newspaper, magazine and book publishers	www.copyright.com.au
The Stationery Office	• Collecting and distributing information to and from government departments and agencies in the UK	www.tso.co.uk

13.2 Registering a DOI with the IDF

The IDF itself, as a Registration Agent, currently registers DOIs in the "Zero-AP" (see Chapter 5) 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.

[1] Application profiles are in the process of being formally defined.

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 http://dx.doi.org/10.1000/181. 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 (see Chapter 5).

Quality control of resolution is a task of Registration Agencies. IDF undertakes to use best endeavours to maintain resolution of DOIs registered directly through IDF, and any errors should be brought to the attention of IDF (info@doi.org).

13.3 Registering a DOI with associated metadata

Registration Agencies support registration of DOIs with associated metadata declaration, i.e. using a DOI Application Profile. 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, or collaboration, between Registration Agencies. Registration Agencies will provide their own information to their community of registrants.

DOI Prefix Requests: DOI prefix applications made by RA members are not handled directly by IDF (any such requests received from known RA customers are re-routed by IDF to the RA). The RA fills out the Prefix request form for its members, with the RA administrator as the technical contact. IDF then creates the prefix and sends the prefix and password to the RA, which is then communicated by the RA to the new prefix holder. The RA becomes the principle contact for the registrant.

Registering DOIs: The following procedure is as an example of the current 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.

The development and implementation of quality control measures used to ensure the validity of the metadata content are the responsibility of the Registrant. Quality control and data checking can be assisted by processes put in place in the RA's metadata collection process.

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), but adding timestamp data value. 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.

13.4 Block prefix allocation

Alternatives to the procedure above under *DOI Prefix Requests* include (a) use of a Local Handle Service by RAs – see chapter 11: procedures for this option are currently in development; and/or (b) handing out a block of prefixes to Registration Agencies, which obviates the need for a Registration Agency to fill out the administration forms to register a prefix (but necessitates some more detailed interaction with the Handle tools directly).

This section describes the process of giving Registration Agencies a block of prefixes to distribute and the steps they will need to take to create the administration handles, and add the administration handles to the prefix handle records.

- Each RA will have an admin handle under 10.admin/. Example:10.admin/JUser
- Each RA will run and understand the HandleTool client; this is part of the standard handle server distribution (www.handle.net).

CNRI, on behalf of IDF, creates a block of prefixes (the number is agreed to with the RA). For each prefix the RA admin handle is added as a value with permission to 'Add Value'. A list of

these prefixes is sent to the RA. When the RA receives a request for a prefix they allocate the requestor's admin handle using the Handle tool:

1. Select 'Create Handle'. Type in the Handle Name: 10.admin/new_login(username).
2. Select 'Add Admin'. Type in the 10.admin/new_login for the Admin ID Handle and type in 300 for the Admin ID Index. The permissions should be: Add Handle, Modify Admin, Add Admin, Read Value. Press OK.
3. You will be prompted for a Key. Select Secret Key. Fill in the requestor's password (either allocated by RA or informed by registrant). The index needs to be 300.
4. Create a 2nd Admin value which will be the admin handle of the RA. (Safeguard in case the password is forgotten and allows the RA permission to change it). Press 'Add Admin' again. Change the 'Index' to 101. The Admin ID Handle will be the RA's admin handle (10.admin/RA_login). The Admin ID index should be 300 and check ALL permissions.
5. Select 'Create'. If not already authenticated, you will be prompted for your authentication. The Authentication Type will be Secret Key (default). Your ID Handle will be your admin handle (10.admin/RA_login). The ID Index will be 300 (default). The Secret Key is the RAs password assigned when RAs admin handle was created. Press OK. You should receive a message that the handle was created successfully.
6. Now add the admin handle to the prefix record: select 'Modify Handle'. Type in the prefix which is being assigned (0.NA/10.something). Press Return. The prefix values will appear in the Data View box. Highlight 'index=200 type-HS_VLIST' then press the Modify button. Another box will appear. Press the Add button. Type in the Handle ID Name, which is the admin handle, just created (10.admin/new_login); the index should be 300. Press OK. You will now see the new value in the list, press OK. You should then get a message that the Handle value was successfully updated.

The admin handle created will now work with the specific prefix just added it to. This information is communicated to the new prefix holder [prefix, username (which they choose) and password].

13.5 Transferring DOIs from one Registrant to another

If a compilation of multiple assigned DOIs (for example, a journal containing a collection of articles; an imprint; a recording catalogue; etc) is transferred from one registrant to another the DOIs within that compilation are transferred as well. For DOIs registered via the IDF the following guidelines apply. Each RA will develop appropriate procedures for this as well. For clarity, in the following we assume the transfer is a sale and refer to "seller" for the original registrant and "purchaser" for the new registrant; however transfer can cover any form of transfer, commercial or otherwise. If the purchaser is not already a DOI registrant, special arrangements may have to be made appropriate to the case; consult the IDF for guidance if necessary. The following assumes that the purchaser is already a DOI registrant.

The individual DOIs stay the same i.e. what the DOI identifies is not changed: *this is a fundamental requirement.* The DOI prefix does NOT change (recall that a prefix is not meaningful, but is initially assigned to a registrant for convenience in generating DOIs only; no reverse look-up can be inferred to a prefix). The administrative value is changed in order for the new owner to modify its data values (most likely the URL value). Both registrants (seller and purchaser) involved in the transfer need to email a message to doi-admin@cnri.reston.va.us giving permission for the transfer. As a follow-up, a letter from the registrant making the sale needs to be sent directly to the IDF as well. The email and letter should state the two registrants, with a contact person and email address, the entity (e.g. journal titles) and both registrants prefixes.

The DOI Directory Manager will request a list of the DOIs (text format, one DOI on each line) from the selling registrant. The Directory Manager will modify the DOIs in batch so that the purchasing registrant can update them.

Each DOI has its own administrative value and that value refers to a list of administrators (a group) within the prefix record in the Global Handle service.

Here is an example of a transfer between DOIs that begin with 10.1000 to the administrator of prefix 10.1200:

> The Prefix record on global looks like this:
> 100 HS_ADMIN 0.NA/10.1000
> 200 HS_VLIST 0.NA/10.1000 (admin users listed here such as ‘jUser)
>
> Each DOI that is created has as its admin value a *reference* to the HS_VLIST record in the prefix. In order to change administrators, the batch file will have to ADD VALUE (for the HS_VLIST reference to the new administrator) and then DELETE VALUE(to delete the existing HS_VLIST reference).
>
> The values of a DOI look like this:
> 1 URL http://www.doi.org
> 100 HS_ADMIN 0.NA/10.1000 index 200
>
> The *NEW* values will look like this:
> 1 URL http://www.doi.org
> 101 HS_ADMIN 0.NA/10.1200 index 200

The DOI Directory Manager will notify the purchasing registrant when the DOIs are modified. The new registrant owner can then use the UPDATE batch function from the java client or administration forms to modify the URL data of their newly acquired DOIs.

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

*ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier is reprinted here with permission of NISO Press. Copyright © 2000 by the National Information Standards Organization. Some text has been omitted from the front and back matter of the standard as originally published. Omissions from the complete document are indicated by the use of asterisks: *****
All NISO standards and technical reports can be downloaded as PDF files for no charge from the NISO website by clicking on the NISO Press icon.

* * * *

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 (this is an illustrative fictitious DOI)

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 metadata system technical information

Describes what metadata related to a DOI should be made available on request. Consists of an expression of the metadata and a style sheet to render the expression using an XSLT processor. Also included are recommendations about how to manage the associated namespace and promote the availability of the metadata.

***NOTE**: The term schema as used in this document does not refer to any particular schema language but to a model, which can be represented in a variety of schema languages. In particular 'schema' should not be confused with the W3C XML Schema as it is intended to allow representation in any appropriate language, including DTD, XML Schema, TREX, RELAX NG, Schematron, and others that may emerge.*

2A. Properties of DMS basic attributes

The following tables specify properties of the basic Attributes of the DMS as defined within the iDD.

In this draft, five of the ten are specified. Specification of the remainder is awaiting the conclusion of contextual analysis supporting the RDD Standard.

1. Identifier	
Definition	A unique label which makes the entity referable.
Structure	1. Alphanumeric string. 2. Each Identifier belongs to a Category which is an IdentifierType, for which AllowedValues exist in the DOI TermSet. New AllowedValues of IdentifierType may be mapped by an AP and included in its TermSet. 3. An Identifier is itself a (unique) Type of Name, and may therefore be substituted for Name. The distinction between an Identifier and a Name is one of functional granularity dependent upon context, as all Names are unique in some context. 4. Identifiers may contain some embedded meaning or metadata according to their Type (for example, a prefix which designates the issuing authority, as with a DOI). 5. Identifiers may have rules or syntax governing their construction. The DOI is an example of an Identifier with some syntax. 6. Identifiers may be broken up into components (eg prefix, core, suffix) for purposes of indexing, or for interrogating their inherent meanings. 7. An Identifier may be a composite with several attributes including Types and components.
Application	1. A Type (eg *DOI, JournalArticleNumber*), will normally be used as a direct substitute for *Identifier.* 2. Identifiers apply to the Resource, but may also apply to Types of any of the ten base attributes (including *Identifier* itself). “Allowed Values” and “Codes” are usually Identifiers used as Attribute Values. 3. A DMS element may have any number of Identifiers.
DMS Composite	*Element (Format, Occurrences)* Value (String, 1) (eg *076450049X*) IdentifierType (AllowedValue, 1) (eg *ISBN*)

	Component group (optional, 0-n) IdentifierPart (String, 1) (eg *7645*) IdentifierPartType (AllowedValue, 1) (eg *IdentifierIssuerCode*)

2. Name	
Definition	A non-unique label which makes the entity referable.
Structure	1. Alphanumeric string. 2. Each Name belongs to one or more Categories (eg *PrimaryName, AbbreviatedName, TranslatedTitle*) which are NameTypes, for which AllowedValues exist in the DOI TermSet. New AllowedValues of NameType may be mapped by an AP and included in its TermSet. 3. Each Name may belong to a Category of Language. 4. Names may or may not contain some embedded meaning or metadata according to their Type. 5. Names may have rules or syntax governing their construction (for example, a formal structure as with Titles of Classical Compositions).. 6. Names may be broken up into components (eg *Prefix, KeyName, Suffix*) for purposes of indexing or re-presentation. 7. A Name may be a simple value (“John Smith”) or may be a composite including several attributes including Types.
Application	1. A Type (eg *Title, PrimaryName*), will normally be used as a direct substitute for *Name.* 2. Names apply to the Resource, but may also apply to Types of any of the ten basic Attributes (including *Name* itself). 3. A DMS element may have any number of Names.
DMS Composite	*Element (Format, Occurrences) (Example)* Name (String, 1) (eg *The Adventures Of Tom Sawyer*) NameType (AllowedValue, 1-n) (eg *OriginalTitle*) AlternativePresentation (String, 0-n) (eg *Tom Sawyer, The Adventures Of*) AlternativePresentationNameType (AllowedValue, 0-n) (eg *IndexedTitle)* Component group (optional, 0-n) NamePart (String, 1) (eg *Melville*) NamePartType (AllowedValue, 1) (eg *KeyName*)

3. Measure	
Definition	A numeric quantification of the entity.
Structure	1. Number. 2. Each Measure belongs to a Category (eg *Height, Speed*, *NumberOfPixels*) which is a MeasureType, for which AllowedValues exist in the DOI TermSet, based on established international standards. New AllowedValues of NameType may be mapped by an AP and included in its TermSet. 3. Each Measure belongs to a Category (eg *MegaBytes, Centimetres, Percent*) which is a Type of *UnitOfMeasure.* The default UnitOfMeasure is *Count.* 4. All Measures may belong to a Category of *Precision (*eg *Approximately, LessThan)*.

	5. Measures may have multiple values with prepositional relationships (eg *From/To, Between*) which are also shown using the *Precision* attribute. 6. A Measure may be a simple value of a compound Attribute (eg *NumberOfPages* = "10") or may be a composite with several attributes including Type and UnitOfMeasure.
Application	1. A Type (eg *Height, Duration*) will normally be used as a direct substitute for *Measure.* 2. Measures apply to the Resource, but may also apply to Types of any of the ten basic Attributes (including *Measure* itself). 3. A DMS element may have any number of Measures.
DMS Composite	*Element (Format, Occurrences) (Example)* Measure (Number, 1) (eg *14)* MeasureType (AllowedValue, 1) (eg *Length*) UnitOfMeasure (AllowedValue, 1) (eg *Millimetres*) Precision (AllowedValue, 0-n) (eg *Approximately*)

4. Annotation	
Definition	A textual description of, or note or comment about the entity.
Structure	1. Text. 2. Each Annotation belongs to one or more Category (eg *Comment, Description, Note*) which is an AnnotationType, for which AllowedValues exist in the DOI TermSet. New AllowedValues of AnnotationType may be mapped by an AP and included in its TermSet. 3. Because Annotations are Creations in their own right, they may have or acquire all the attributes of another Creation, including a DOI.
Application	*1.* A Type (eg *Comment, Description, Note, Review*) may be used as a direct substitute for *Annotation.* 2. Annotations apply to the Resource, but may also apply to Types of any of the ten basic Attributes (including *Annotations* itself). 3. A DMS element may have any number of Annotations. 4. Because Annotations are Creations in their own right, they may be treated as Relatives with the RelatingTerm *HasAnnotation* linked to an Annotation Identifier or Name.
DMS Composite	*Element (Format, Occurrences)* AnnotationText (Text, 1) AnnotationType (AllowedValue, 1) AnnotationLanguage (AllowedValue, 0-n) AnnotationAuthor (0-n) (see Agent Composite) AnnotationDate (0-n) (see Date Composite) AnnotationPlace (0-n) (see Place Composite)

5. Category	
Definition	A classification to which the entity belongs.
Structure	1. String or AllowedValue.

	2. Each Category belongs to one or more Category of its own (eg *Colour, CreationType, Subject*) which is a CategoryType, for which AllowedValues exist in the DOI TermSet. New AllowedValues of CategoryType may be mapped by an AP and included in its TermSet. 3. If a Category Value is an AllowedValue, it will belong to a further Category which identifies the AllowedValueTermSet from which the Value was drawn. 4. If a Category Value is a String, the distinction between a Category and an Annotation is one of functional granularity.
Application	1. A Type (eg *Comment, Description, Note, Review*) will normally be used as a direct substitute for Category. 2. Categories apply to the Resource, but may also apply to Types of any of the ten basic Attributes (including *Categories* themselves). 3. A DMS element may belong to any number of Categories.
DMS Composite	*Element (Format, Occurrences) (Example)* CategoryValue (String or AllowedValue, 1) (eg *DigitalObject*) CategoryType (AllowedValue, 1) (eg *ResourceType*)

2B. DMS Declarations: principal features

This Appendix explains and demonstrates six standard features of the DMS, using the DOI Metadata Kernel for illustration.

1. *Types* of basic attributes

Few of the ten basic attributes (*see Table 2*) appear in their basic form in any particular Declaration. For any particular Resource, its attributes are likely to appear as ***Types*** of the basic ten.

The word “Type” is used in the DMS and in <indecs> in the sense of “subtype”: that is, a more “specialized” or limited kind of a basic attribute. For example, a ***DOI*** is a Type of Identifier, a ***Title*** is a Type of Name, ***Height*** is a Type of Measure, ***Description*** is a Type of Annotation, ***Creator*** is a Type of Agent, and so on.

Types of attributes may in turn have their own Types: for example, a Description might have a Type called a ***SummaryDescription***, a Creator may have a Type called an ***Author***, and so on. This “specialization” of Types may go on through any number of levels. In principle there is no limit to the number of Types of any attribute.

Some Types, such as the examples given above, are common to many domains and so are included as elements in the DMS. Others may be highly specific to a particular domains or applications. These can be mapped to the DMS elements, but might belong only to the extension for a particular Application Profile.

The use of Types can be demonstrated in the DOI Kernel. Six of the eight Kernel elements are Types of one of the basic attributes (Kernel elements in **bold**):

Resource

Has**Identifier**=[Value]
HasIdentifier=[Value]
HasType=**DOI**
HasName=[Value]
HasType=**Title**

HasCategory=[Value]
HasType=**DOIApplicationProfile**
HasCategory=[Value]
HasType=**StructuralType**
HasCategory=[Value]
HasType=**Mode**
HasAgent=[Value]
HasType=**PrimaryAgent**

2. Simplifying Declarations by substituting Types

A DMS Declaration can be simplified by routinely substituting Types for their parents. The illustration below shows the Kernel elements as direct attributes of the Resource. The original basic attributes are given in italics to show how they fit into the now-hidden underlying structure:

Resource

[HasIdentifier]
Has**Identifier**=[Value]
Has**DOI**=[Value]
[HasName]
Has**Title**=[Value]
[HasCategory]
Has**DOIApplicationProfile**=[Value]
Has**StructuralType**=[Value]
Has**Mode**=[Value]
[HasAgent]
Has**PrimaryAgent**=[Value]

This logical structure will allow the Declaration to be presented in the clearest way to the human eye while preserving the underlying structure in the XML DTD/Schema.

3. Attributes of Attributes

Attributes are commonly more complex than a single value. Any of the ten basic attributes (or their Types) may have its own attributes, drawn from the same list of ten.

For example, a Category may have a Name; an Agent may have a Category; an Annotation may have an Agent; or a Measure may have a related Time.

The eighth Kernel element (***AgentRole***) is an example of an attribute of an attribute, added below (red font):

Resource

[HasIdentifier]
Has**Identifier**=[Value]
Has**DOI**=[Value]
[HasName]
Has**Title**=[Value]
[HasCategory]
Has**DOIApplicationProfile**=[Value]
Has**StructuralType**=[Value]
Has**Mode**=[Value]
[HasAgent]
Has**PrimaryAgent**=[Value]
HasCategory=[Value]

HasType=**AgentRole**

Some of the other Kernel attributes will need further attributes, depending on how they are being referenced. The ***DOIApplicationProfile*** and ***PrimaryAgent*** will be identified either by a ***Name*** or ***Identifier:***

Resource

[HasIdentifier]
Has**Identifier**=[Value]
Has**DOI**=[Value]
[HasName]
Has**Title**=[Value]
[HasCategory]
Has**DOIApplicationProfile**
HasIdentifier=[Value]
Has**StructuralType**=[Value]
Has**Mode**=[Value]
[HasAgent]
Has**PrimaryAgent**
HasName=[Value]*
HasIdentifier=[Value]*
HasCategory=[Value]
HasType=**AgentRole**

*The DMS XML DTD/Schema will mandate one but not necessarily both of these attributes.

These attributes in turn may have their own Types:

Resource

[HasIdentifier]
Has**Identifier**=[Value]*
Has**IdentifierType**=[Value]
Has**DOI**=[Value]
[HasName]
Has**Title**=[Value]*
[HasCategory]
Has**DOIApplicationProfile**
Has**Identifier**=[Value]
Has**IdentifierType**=**DOI**
Has**StructuralType**=[Value]
Has**Mode**=[Value]
[HasAgent]
Has**PrimaryAgent**
Has**Name**=[Value]
Has**NameType**=[Value]
Has**Identifier**=[Value]
Has**IdentifierType**=[Value]
HasCategory=[Value]
HasType=**AgentRole**

Note: this example and the Schema assumes that APs will be identified with DOIs.

4. Allowed Values

Some attribute values are drawn from sets of ***allowed*** or ***controlled*** values or codes. This is principally true for Categories, but is not uncommon for Places and Agents. Allowed values (often referred to as ***code lists***) are valuable for interoperability as they may be drawn from

widely used standards (such as ISO Currency Codes), or they may be mappable to other code lists through the iDD. As far as possible, Category values used in DOI AP Declarations should be controlled.

Allowed values are shown in the DMS using the RelatingTerm ***HasAllowedValue***. In turn these Values will be drawn from a named set of codes which is known as an ***AllowedValueTermSet.*** Again this can be illustrated in the Kernel, which can now be shown in full, with Types substituted as appropriate:

Resource

[HasIdentifier]
Has**Identifier**=[Value]*
Has**IdentifierType**=[Value]
HasAllowedValueTermSet=[Value]
Has**DOI**=[Value]
[HasName]
Has**Title**=[Value]*
[HasCategory]
Has**DOIApplicationProfile**
Has**Identifier**=[Value]
Has**IdentifierType**=**DOI**
Has**StructuralType**=[AllowedValue]
HasAllowedValue=**PhysicalFixation**
HasAllowedValue=**DigitalFixation**
HasAllowedValue=**Performance**
HasAllowedValue=**Abstraction**
Has**Mode**=[Value]
HasAllowedValue=**Audio**
HasAllowedValue=**Visual**
HasAllowedValue=**Audiovisual**
HasAllowedValue=**Abstract**
[HasAgent]
Has**PrimaryAgent**
Has**Name**=[Value]
Has**NameType**=[AllowedValue]
HasAllowedValueTermSet=[Value]
Has**Identifier**=[Value]
Has**IdentifierType**=[AllowedValue]
HasAllowedValueTermSet=[Value]
Has**AgentRole**=[ControlledValue]
HasAllowedValueTermSet=[Value]

5. Composite Attributes

Sometimes attributes are represented by two or more elements in a structured group, known in DMS (as in ONIX) as a ***Composite***.

For example, a Measure may be described in a Composite comprised of these basic attributes and Types:

Measure=[Value]
HasType=[Value]
HasCategory=[AllowedValue]
HasType=UnitOfMeasure
HasCategory=[AllowedValue]
HasType=Precision

which after substituting types looks like this:

MeasureType=[Value]
HasUnitOfMeasure=[AllowedValue]
HasPrecision=[AllowedValue]

For example:

Height=12
HasUnitOfMeasure=Centimetres
HasPrecision=LessThan

Standard composites for each of the ten basic attributes are provided in the DMS (See Appendix A). RAs may define their own Composites which can be mapped to DMS like any other Term (for example, the CrossRef AP has Composites for ***doi_data***, ***article***, ***journal*** and ***author*** among others). CompositeAttributes may also contain other CompositeAttributes.

6. Contexts as Attributes

The last five of the ten basic attributes shown in table 2 have a special relationship: the last four (***Agent, Relative, Time, Place***) are attributes of the first (***Context***) according to the <indecs> Context Model:

Context
HasAgent
HasResource
HasTime
HasPlace

The Context is at the heart of the <indecs> Data Dictionary model where it is the key to defining and mapping most Terms. This section explains and illustrates how Contextual attributes can be used when needed in the DMS.

A Context is either an ***Event*** in which something changes, or a ***Situation*** in which something remains the same. For example, if someone translates a text, then there is or was a Context in which an Agent (the ***Translator***) created the new Resource (the ***Translation***) from the original Resource (the ***Source***) at a certain ***Time*** and ***Place***. It is through this ***TranslatingEvent*** that the Resource acquires its attribute relationship with the other elements. There are many circumstances where it is useful or essential to identify a Context explicitly (for example, in workflow management or rights management).

However, in Resource Metadata it is common for the Context to be ignored as an entity, and the activity (the "verb" – in this example, ***Translate***) at the heart of a Context to be "bundled" along with one or more of its attributes into a single Term. For example, the Event of Publishing is often shown through a number of different attributes such as "Publisher" (Agent), "DateOfPublication" (Time) and "PlaceOfPublication" (Place), rather than through an explicit description of the Event itself. This is described as a ***Resource-based view*** rather than a ***Context-based view***.

Both views are widely used and valid. Each has advantages and disadvantages. The two different approaches are illustrated below using a fictitious example of book publication. For simplicity the full attribute Type hierarchy is *not* shown but is shown in footnotes. First, an example where the publication Event is made explicit[2]:

[2] Showing the Type hierarchy, this would look as follows:
Resource "076450049X"
Has**Context** "Publication"
HasType "Event"
[HasAgent]
Has**Publisher** "IDG Books Worldwide, Inc"

Resource "076450049X"
Has**Event** "Publication"
Has**Publisher** "IDG Books Worldwide, Inc"
Has**Date** "1996"
Has**Place** "California, USA"

Secondly, where the Event is implicit[3]:

Resource "076450049X"
Has**Publisher** "IDG Books Worldwide, Inc"
Has**DateOfPublication** "1996"
Has**PlaceOfPublication** "California, USA"

Either of these approaches is supported by the DMS.

In this example the second approach is more compact, as all three elements are direct attributes of the Resource, whereas the first requires the extra step of indirection and some means of identifying or naming a Context. For this reason the second approach is the most common in simple Resource description schemes.

However, in cases where the metadata becomes more complicated, the "Context" approach can become more efficient. If, for example, we need to show that the book had been published on different dates in two different countries, a level of indirection would have to be introduced, perhaps like this:

Resource "076450049X"
Has**Publisher** "IDG Books Worldwide, Inc"
Has**DateOfPublication** "1996"
Has**PlaceOfPublication** "California, USA"
Has**DateOfPublication** "1998"
Has**PlaceOfPublication** "Paris, France"
Has**DateOfPublication** "1999"
Has**PlaceOfPublication** "Moscow, Russia"

(Note that in this example we have made the **PlaceOfPublication** an attribute of the **DataOfPublication**. It could just as easily have been done the other way round: the DMSstructure will support either).

Our example now contains a level of indirect attribution. Then suppose that the book has different publishers in one of the countries (assuming in this example that the resource itself and therefore its ISBN remains the same):

Resource "076450049X"
Has**Publisher** "IDG Books Worldwide, Inc"
Has**DateOfPublication** "1996"

[HasTime]
Has**Date** "1996"
Has**Place** "California, USA"

[3] Showing the Type hierarchy, this would look as follows:
Resource "076450049X"
[HasAgent]
Has**Publisher** "IDG Books Worldwide, Inc"
[HasTime]
Has**DateOfPublication** "1996"
[HasPlace]
Has**PlaceOfPublication** "California, USA"

Has**PlaceOfPublication** “California, USA”
Has**DateOfPublication** “1998”
Has**PlaceOfPublication** “Paris, France”
Has**Publisher** “XYZ Books”
Has**DateOfPublication** “1999”
Has**PlaceOfPublication** “Moscow, Russia”

We must now have two levels of indirect attribution in order to ensure that the metadata is unambiguous. However, if we say the same thing with explicit Contexts, we need only a single level of indirection:

Resource “076450049X”
Has**Event** “Publication”
Has**Publisher** “IDG Books Worldwide, Inc”
Has**DateOfPublication** “1996”
Has**PlaceOfPublication** “California, USA”
Has**Event** “Publication”
Has**Publisher** “IDG Books Worldwide, Inc”
Has**DateOfPublication** “1998”
Has**PlaceOfPublication** “Paris, France”
Has**Event** “Publication”
Has**Publisher** “XYZ Books”
Has**DateOfPublication** “1999”
Has**PlaceOfPublication** “Moscow, Russia”

Where Contexts have multiple Agents, Times or Places, or involve multiple Resources, it is often more efficient to use the explicit Context approach to organize Resource metadata. In the iDD, the Context structure is one of the keys to interoperability.

The DMS supports either approach, and importantly provides (through the iDD) an internal mapping between them, so that it can be possible to transform metadata from one basis to the other automatically without loss of semantic value.

2C. The DOI TermSet [placeholder]

The initial TermSet of the DMS will be drawn from the foundation Terms in the <indecs> Data Dictionary which will be first published following the MPEG meeting in Klagenfurt, Austria in July 2002. The Dictionary forms a part of the proposed RDD in the MPEG-21 Standard and will form part of the Committee Draft of the ISO 21000 Standard in October 2002.

The first version of the DOI TermSet will follow on the publication of the iDD.

2D. Mapping example

CrossRef metadata expressed in DMS Terms

This Appendix shows an expression of the CrossRef AP metadata using DMS Terms. *Note: this carries two provisos: that the mappings require agreement from CrossRef AP, and that the DMS Terms are provisional as noted in the Foreword to this paper.*

This expression is based, for simplicity, on the CrossRef DTD. CrossRef now also uses a more extensive Schema, which can be mapped and expressed using the same techniques.

Figure 1 shows how all the CrossRef Terms can be declared using Terms from the DOI TermSet (this illustration shows Terms only, and does not show characteristics such as optionality or dependence which will be enabled by an XML DTD or Schema).

This Declaration is highly simplified, using most of the techniques explained in Appendix B.

The assumption made here is that all CrossRef Terms will be members of the DOI TermSet, as they are all sufficiently general to be of likely value to other APs. A Declaration using these Terms could therefore be parsed by any application primed to deal with a general DTD or Schema for an Extended Kernel.

Because the same attributes can be expressed in different ways within DMS, there is no "definitive" model for a Declaration. For example, this example uses compressed forms of expression for most attributes (such as "**HostJournalVolumeIssueEndPageNumber")** where another expression might show this as an attribute of a related Creation (the *Page Number* of the *Host Journal Volume Issue*). The mappings within iDD support both these views, and others in between, so the decision is a functional one according to expected user requirements. Different forms of Declaration could be produced from the same mappings to meet different requirements. This mapping does not show the grouping of all Composite elements, of which CrossRef has several in its proprietary DTD: again, that is a presentation choice which would be supported by the DTD or Schema.

*Note: one CrossRef term (**specialized_numbering**) is not yet mapped as it requires consultation to clarify its meaning.*

Figure 1: A simplified expression of the CrossRef Metadata Declaration in the DMS TermSet
Kernel Elements are shown in red.

DMS Term	***Source CrossRef DTD element***
Resource	
Has**DOI** = [Value]	doi
Has**PII** = [Value]	pii
Has**SICI** = [Value]	sici
Has**PublisherProprietaryNumber** = [Value]	article_number
Has**HostJournalVolumeNumber** = [Value]	volume
Has**HostJournalVolumeIssueNumber** = [Value]	issue
Has**HostJournalVolumeIssueStartPageNumber** = [Value]	first_page
Has**HostJournalVolumeIssueEndPageNumber** = [Value]	last_page
Has**Title** = [Value]	title
Has**DOIApplicationProfile=Crossref**	(from AP)
Has**StructuralType** =**Abstraction**	(from AP)
Has**CreationType=JournalArticle**	(from AP)
Has**Mode** = **Visual**	(from AP)
Has**Author**	author
Has**GivenName** = [Value]	given_name
Has**Surname** = [Value]	surname
Has**AuthorPriority** = [Value]	author sequence
HasAllowedValue=**FirstInSequence**	first
HasAllowedValue=**AdditionalInSequence**	additional
Has**HostCreation**	journal
HasCategory=**Journal**	journal
Has**PrintISSN** = [Value]	issn + issn type=print
Has**ElectronicISSN** = [Value]	issn + issn type=electronic
Has**CODEN** = [Value]	coden
Has**FullTitle** = [Value]	full_title
Has**AbbreviatedTitle** = [Value]	abbreviated_title
Has**Part**	volume
HasCategory=**JournalVolume**	volume
Has**JournalVolumeNumber** = [Value]	volume

Has**Fixation** = [Value]	issue
Has**Category** = JournalVolumeIssue	issue
Has**Category** = JournalVolumeIssueNumber	issue
Has**ElectronicPublicationDate**	date + date type=online
Has**ElectronicPublicationYea**r = [Value]	year
Has**ElectronicPublicationMonth** = [Value]	month
Has**ElectronicPublicationDay** = [Value]	day
Has**PrintPublicationDate**	date + date type=print
Has**PrintPublicationYear** = [Value]	year
Has**PrintPublicationMonth** = [Value]	month
Has**PrintPublicationDay** = [Value]	day
Has**OtherPublicationDate**	date + date type=other
Has**OtherPublicationYear** = [Value]	year
Has**OtherPublicationMonth** = [Value]	month
Has**OtherPublicationDay** = [Value]	day
Has**URL** = [Value]	url

Note that the above example includes alternative ways of showing the Resource's relationship to its host journal: once using a series of Identifiers (**HostJournalVolumeNumber** etc) and once through a series of Relatives and their attributes and Relatives (**HostCreation,** Has**Part,** Has**Fixation**). Either or both approaches can be used. Other elements in the above example might be presented in different structures and remain consistent with the DMS and iDD mappings.

Figure 2 shows an expanded form of the same Declaration in which "bridging terms" are shown (in grey font), and in which Types have not been substituted. This is not intended as a serious proposal for a published declaration, but is included to illustrate something of the infrastructure that iDD puts behind the relating of terms in an apparently simple expression. In this example each step shows a single specialization from its parent type, which is the optimum level of granularity.

Note that where an attribute has several "bundled" qualifications or categories (such as HostJournalVolumeIssueStartPageNumber) the iDD has several different "routes" down which navigation of its database would be possible, and therefore several possible sequences of "bridging terms". Because this is a flat presentation it is only possible to show one of those routes in each case.

Figure 2: An unsimplified expression of the Crossref Metadata Declaration in the DMS TermSet

Resource
Has**Identifier** = [Value]
HasType=**DOI**
Has**Identifier** = [Value]
HasType=**PII**
Has**Identifier** = [Value]
HasType=**SICI**
Has**Identifier** = [Value]
HasType=**PublisherProprietaryNumber**
Has**Identifier** = [Value]
HasType=**IdentifierOfRelatedCreation**
HasType=**IdentifierOfHostCreation**
HasType=**IdentifierOfHostSerial**
HasType=**HostJournalVolumeNumber**
Has**Identifier** = [Value]
HasType=**IdentifierOfRelatedCreation**
HasType=**IdentifierOfHostCreation**
HasType=**IdentifierOfHostPrintedVolume**
HasType=**IdentifierOfHostPrintedSerialIssue**

HasType=**IdentifierOfHostPrintedJournalVolumeIssue**
HasHostJournalVolumeIssueNumber=[Value]
HasIdentifier=[Value]
HasType=**IdentifierOfRelatedCreation**
HasType=**IdentifierOfHostCreation**
HasType=**IdentifierOfHostPrintedPage**
HasType=**HostPrintedPageNumber**
HasType=**HostSerialIssuePageNumber**
HasType=**HostJournalIssuePageNumber**
Has**HostJournalVolumeIssue**
StartPageNumber
HasIdentifier=[Value]
HasType=**IdentifierOfRelatedCreation**
HasType=**IdentifierOfHostCreation**
HasType=**IdentifierOfHostPrintedPage**
HasType=**HostPrintedPageNumber**
HasType=**HostSerialIssuePageNumber**
HasType=**HostJournalIssuePageNumber**
Has**HostJournalVolumeIssue**
EndPageNumber
HasName=[Value]
HasType=**Title**
HasCategory
HasType=**DOIApplicationProfile**
HasAllowedValue=**Crossref**
HasCategory=Abstraction
HasType=**StructuralType**
HasCategory=**JournalArticle**
HasType=**CreationType**
HasCategory=**Visual**
HasType=**Mode**
HasAgent
HasType=**Creator**
HasType=**Author**
HasName=[Value]
HasType=**GivenName**
HasName=[Value]
HasType=**Surname**
HasCategory=[Value]
HasType **AgentTypePriority**
HasType **AuthorPriority**
HasAllowedValue=**FirstInSequence**
HasAllowedValue=**AdditionalInSequence**
HasRelative
Type=**HostCreation**
HasCategory=**Journal**
HasIdentifier=[Value]
HasType=**ISSN**
HasType=**PrintISSN**
HasIdentifier=[Value]
HasType=**ISSN**
HasType=**ElectronicISSN**
HasIdentifier=[Value]
HasType=**ISSN**
HasType=**CODEN**
HasName=[Value]
HasType=**Title**
HasType=**FullTitle**
HasName=[Value]
HasType=**Title**
HasType=**AbbreviatedTitle**
Has**Part**
HasType=**JournalVolume**
HasIdentifier=[Value]
HasType=**JournalVolumeNumber**

HasFixation
HasType=PrintedVolume
HasType=JournalVolumeIssue
HasIdentifier=[Value]
HasType=JournalVolumeIssueNumber
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=ElectronicPublicationDate
HasUnitOfMeasure=Year
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=ElectronicPublicationDate
HasUnitOfMeasure=Month
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=ElectronicPublicationDate
HasUnitOfMeasure=Day
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=PrintPublicationDate
HasUnitOfMeasure=Year
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=PrintPublicationDate
HasUnitOfMeasure=Month
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=PrintPublicationDate
HasUnitOfMeasure=Day

HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=OtherPublicationDate
HasUnitOfMeasure=Year
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=OtherPublicationDate
HasUnitOfMeasure=Month
HasTime=[Value]
HasType TimeOfEvent
HasType EventDate
HasType PublicationDate
HasType=OtherPublicationDate
HasUnitOfMeasure=Day
HasPlace
HasName
HasType VirtualAddress
HasType URL

2E. DOI Kernel Metadata

1. Introduction

This Appendix defines how to express the public metadata applicable to general intellectual property entities that have been determined to be a mandatory component of the DOI (the kernel) in XML. This expression ensures compliance with the minimal requirements based on the <indecs> principles for resource description. However, the coverage provided by the kernel is unlikely to be sufficient for the requirements of particular application domains so it should be considered a starting point for registration agencies wishing to construct a metadata profile (as part of an application profile) that incorporates the requirements of the IDF. Registration agencies can allow or mandate additional 'public' metadata elements, while maintaining the integrity of the overall framework, and further can allow 'private' metadata elements to be added that only a particular systems may be able to understand and use. A data dictionary is available from the IDF to support the creation of these metadata profiles, in the form of the DOI-namespace database.

This metadata expression is concerned solely with the kernel metadata and does not mandate any administrative metadata (such as when the metadata was created). The metadata supplied at the time of registration of a DOI must facilitate the derivation of the kernel declaration.

The DOI kernel metadata schema is expressed as an XML Schema. Other schema languages can be used to express the schema (such as DTDs or RELAX NG). The metadata expression should support different language encoding. The expression must always include the encoding used within the elements in the XML declaration.

2. Building the XMLSchema

The following principles underlie the formulation of the DOI kernel metadata expression:

1. A valid metadata kernel must be declared upon registration of each DOI.
2. Kernel metadata is both stable and persistent. "Stable" implies the true nature of the kernel metadata and signifies the need for accuracy when registering metadata.
3. Kernel metadata must be available to any enquirer through the resolution system
4. <indecs> principles must be embodied in the kernel metadata and core elements must be used as defined in the DOI Handbook.
5. The metadata should be recognizable as having come from the authority of the IDF.
6. The source authority of the metadata (the party asserting its veracity) should be identified. Unless otherwise specified this is assumed to be the registrant.

The XML instance should be encoded in UTF-8. This is the recommended character encoding.

The schema defines kernel metadata for a DOI that is provided by an assertor. The DOI itself is an opaque string that must obey the syntax specified in the ANSI/NISO Z39.84 specification. The entity represents information about that which is identified by the **DOI**, which is a mandatory element of the schema.

There may be one or more other **Identifiers**. Each **Identifier** should be equivalent to the DOI (can also be used to refer to the entity that the DOI is associated with) and must be defined though the use of an **IdentifierType** which names the identifier realm, and the **IdentifierValue**.

There may be one or more **Titles**. Each **Title** contains a **Language** (the ISO693 code for the language in which the title is described) and a **TitleValue** which contains the actual title.

The **StructuralType** is mandatory and describes the nature of the entity being identified (a single value must be selected from a list). The **Mode** describes the means by which the

identified entity or its intended manifestation or expression is perceived (multiple values are allowed from a defined set).

There must also be one or more **PrimaryAgents**. PrimaryAgents consist of at least one **Agent** which may be a person, group or organization. Each agent has a **Name** and a **Roles** element. The Name is comprised of a **NameType** (which may be a person, group or organization) and a **NameValue**. The Roles element contains one or more **Role** elements permissible values for whichcan be defined within a particular profile and are outside the scope of this specification (designers may wish to consult the DOI-namespace database for guidance). As identifier is a type of name, then this element can be used to express agent identifiers if required.

The **Assertor** contains the **Registrant** that provides the metadata information and the registration **Authority** that validated that information – as well the date on which the metadata was asserted within the DOI System (Date). Both these elements can contain a single **Name** and **Identifier** element.

3. Namespace

All elements within the **KernelMetadata** tags belong from to the http://www.doi.org/ namespace, which is declared as an IDF namespace with permissible names as defined within this document. Kernel metadata could be derived from other namespaces: however by establishing a DOI namespace we provide a stable meaning with a DOI definition of terms under this namespace. The doi namespace must always be declared in the XML instance using the URI http://www.doi.org/.

4. DMS kernel metadata XML schema draft 0.1

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:doi="http://www.doi.org/">

<xs:annotation>
	<xs:documentation>DOI Metadata Kernel Version 0.1</xs:documentation>
</xs:annotation>

<xs:complexType name="doi:KernelMetadata">
	<xs:sequence>
		<xs:element name="DOI" type="DOI"/>
		<xs:element name="Identifiers" type="Identifiers" minOccurs="0"/>
		<xs:element name="Titles" type="Titles" minOccurs="0"/>
		<xs:element name="StructuralType" type="StructuralType"/>
		<xs:element name="Modes" type="Modes"/>
		<xs:element name="PrimaryAgents" type="PrimaryAgents"/>
		<xs:element name="Assertor" type="Assertor" minOccurs="0"/>
	</xs:sequence>
</xs:complexType>

<xs:simpleType name="DOI">
	<xs:restriction base="xs:string">
		<xs:pattern value="[^\./]+(\.[^\./]+)*/.+"/>
	</xs:restriction>
</xs:simpleType>

<xs:complexType name="Identifiers">
	<xs:sequence maxOccurs="unbounded">
		<xs:element name="Identifier" type="Identifier"/>
	</xs:sequence>
</xs:complexType>
```

```
<xs:complexType name="Identifier">
	<xs:sequence>
		<xs:element name="IdentifierType" type="xs:string"/>
		<xs:element name="IdentifierValue" type="xs:string"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Assertor">
	<xs:sequence>
		<xs:element name="Registrant" type="Party"/>
		<xs:element name="Authority" type="Party"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Party">
	<xs:sequence>
		<xs:element name="Name" type="Name"/>
		<xs:element name="Identifier" type="Identifier" minOccurs="0"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Titles">
	<xs:sequence maxOccurs="unbounded">
		<xs:element name="Title" type="Title"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Title">
	<xs:sequence maxOccurs="1">
		<xs:element name="TitleValue" type="xs:string"/>
		<xs:element name="Language" type="xs:language"/>
	</xs:sequence>
</xs:complexType>

<xs:simpleType name="StructuralType">
	<xs:restriction base="xs:string">
		<xs:enumeration value="PhysicalFixation"/>
		<xs:enumeration value="DigitalFixation"/>
		<xs:enumeration value="Performance"/>
		<xs:enumeration value="Abstraction"/>
	</xs:restriction>
</xs:simpleType>

<xs:complexType name="Modes">
	<xs:sequence maxOccurs="4">
		<xs:element name="Mode">
			<xs:simpleType>
				<xs:restriction base="xs:string">
				<xs:enumeration value="Abstract"/>
				<xs:enumeration value="Audio"/>
				<xs:enumeration value="AudioVisual"/>
				<xs:enumeration value="Visual"/>
				</xs:restriction>
			</xs:simpleType>
		</xs:element>
	</xs:sequence>
</xs:complexType>
```

```
<xs:complexType name="PrimaryAgents">
	<xs:sequence maxOccurs="unbounded">
		<xs:element name="Agent" type="Agent"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Agent">
	<xs:sequence>
		<xs:element name="Name" type="Name"/>
		<xs:element name="Roles" type="Roles"/>
	</xs:sequence>
	<xs:attribute name="sequence" type="xs:positiveInteger" use="required"/>
</xs:complexType>

<xs:complexType name="Name">
	<xs:sequence>
		<xs:element name="NameType" type="xs:string"/>
		<xs:element name="NameValue" type="xs:string"/>
	</xs:sequence>
</xs:complexType>

<xs:complexType name="Roles">
	<xs:sequence>
		<xs:element name="Role" type="xs:string"/>
	</xs:sequence>
</xs:complexType>

</xs:schema>
```

5. Example metadata

```
<?xml version=”1.0” encoding="UTF-8"?>
<doi:KernelMetadata xmlns=”http://www.doi.org/”>
<DOI>10.1025/abio.4372.9898</DOI>
<Identifiers><Identifier>
	<IdentifierType>PII</IdentifierType>
	<IdentifierValue>S1031-5806(95)00403-9</IdentifierValue>
<Identifier></Identifiers>
<Titles><Title>
	<TitleValue>DRM in Streaming Media</TitleValue>
	<Language>en</Language>
</Title></Titles>
<StructuralType>visual</StructuralType>
<Modes>
	<Mode>visual</Mode>
</Modes>
<PrimaryAgents>
	<Agent sequence=”1”><Name>
		<NameType>Person</NameType>
		<NameValue>Lincoln Smith</NameValue>
	</Name><Roles>
			<Role>author</Role>
	</Roles></Agent>
	<Agent sequence=”2”><Name>
		<NameType>Person</NameType>
		<NameValue>Marsha Grossman</NameValue>
	</Name><Roles>
		<Role>illustrator</Role>
	</Roles></Agent>
```

```
</PrimaryAgents>
<Assertor>
        <Registrant><Name>
                <NameType>Organization</NameType>
                <NameValue>Elsevier Science</NameValue>
        </Name><Identifier>
                <IdentifierType>PublisherCode</IdentifierType>
                <IdentifierValue>123321</IdentifierValue>
        </Identifier></Registrant>
        <Authority><Name>
                <NameType>Organization</NameType>
                <NameValue>PILA</NameValue>
        </Name></Authority>
</Assertor>
</doi:KernelMetadata>
```

Appendix 3: 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)

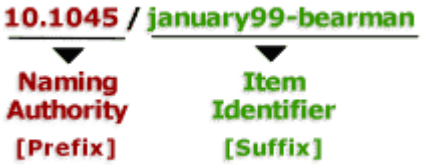


Figure 1

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)

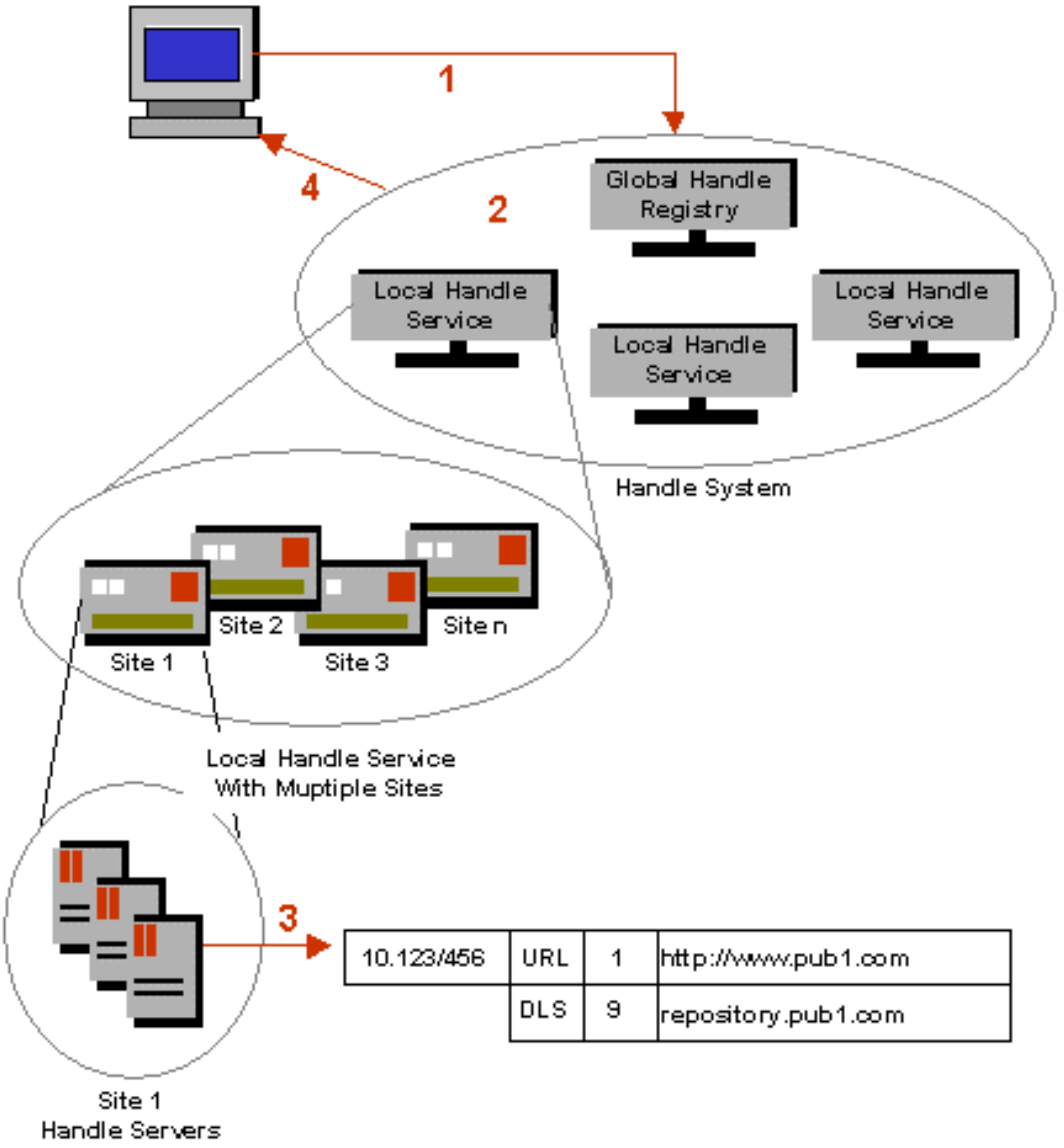


Figure 2

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.

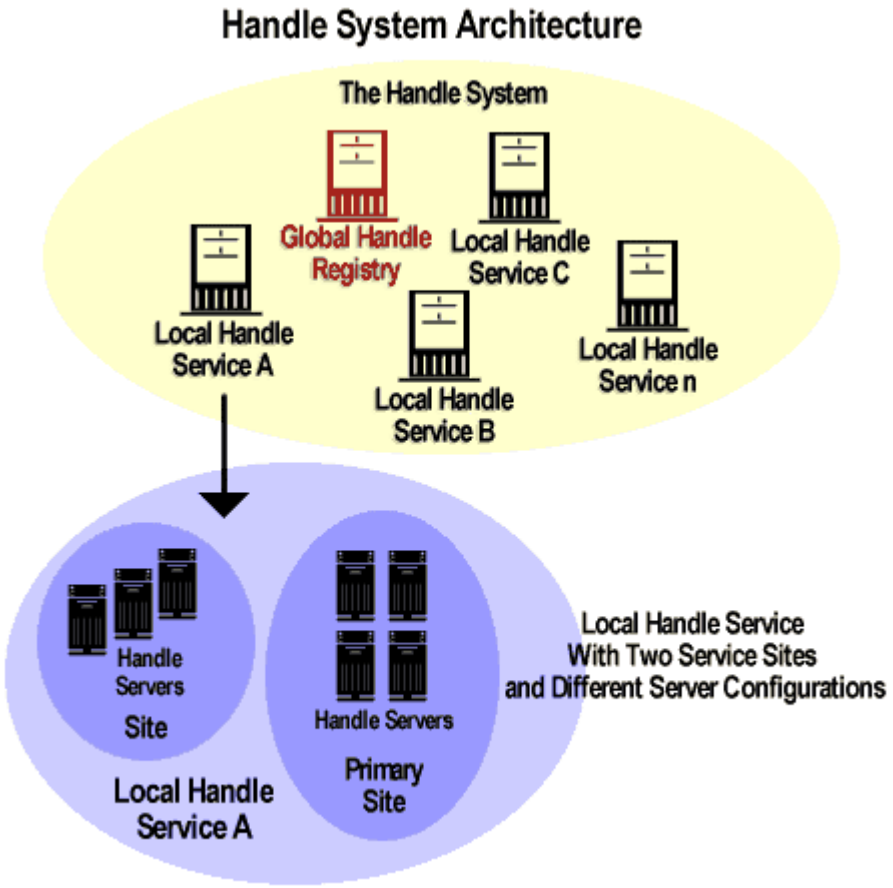


Figure 3

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.

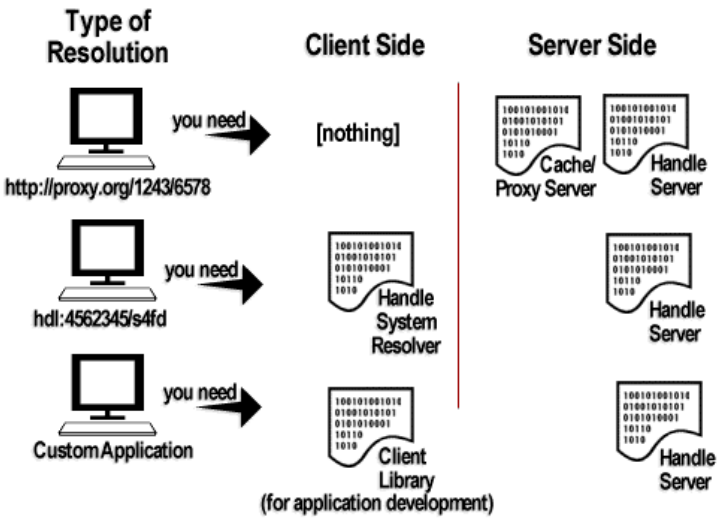


Figure 4: Client Side and Server Side Software

Currently available client side software includes:

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 JAVA Version), including 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.

Currently available client side software:

- the **CNRI Handle System Resolver** which enables web browsers to recognize the handle protocol.
- the Handle System **Client Library** (JAVA™ Version), 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), 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.

Appendix 4: An introduction to <indecs>®

Section 1 of this introduction to <indecs> draws heavily on the introduction to the <indecs> Metadata Framework, a final project document from the <indecs> project published in July 2000. This document summarizes the technical work of the original <indecs> project (Dec 1998-June 2000) which produced ***the <indecs> metadata framework****, a reference model.*

Section 2 describes the evolution of this work in the context of the Contecs:DD work, which produced the indecs ***Rights Data Dictionary (RDD)****, a common dictionary or vocabulary for intellectual property rights to be based on the <indecs> framework. The dictionary resulting from this activity was adopted (in Dec 2001) as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard, and is adopted by IDF as the basis for the DOI namespace.*

See chapter 8 for a summary of the evolution of the <indecs> work.

1. The <indecs> metadata framework

The <indecs> project (and its successors the not-for-profit **Indecs Framework Ltd** and **Contecs:DD consortium)** 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.

The initial <indecs> (interoperability of data in e-commerce systems project) project was established at the end of 1998, with support from the European Commission's DGXIII Info 2000 Programme and a wide range of partners and affiliates, representing a very broad cross-section of international bodies representing all aspects of the content industries' value chain from creators to users. The project documentation and Summary Final Report may be found at www.indecs.org.

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?

1.1 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

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.

1.2 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.

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	(conceptual framework model)
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.

1.3 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.

1.4 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.

1.5 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.

2. Contecs:DD - Rights Data Dictionary

2.1 Introduction

Contecs:DD is a consortium-based initiative to develop a rights data dictionary, originally known as the <indecs>rdd consortium and renamed in July 2002 (the renaming was merely for technical administrative purposes). Its purpose is to support the implementation of a rights language for secure exchange of intellectual property on networks. The initiative, based on the original <indecs> analysis (1998-2000: www.indecs.org), has been underway since mid-2001. At that time a consortium of rights holder representatives and providers of services funded the work and engaged the services of Rightscom to lead the technical work of the project.

The <indecs> analysis was a reference model. In order to implement the reference model, a practical reference implementation was required. In April 2001 the International DOI Foundation funded a feasibility study for the development of a Rights Data Dictionary (RDD), a common dictionary or vocabulary for intellectual property rights to be based on the <indecs> framework. The aim was to propose a consortium which would fund the development of standard rights terms to enable the exchange of key information between content industries for eCommerce trading of intellectual property rights. <indecs>rdd was the next phase in <indecs> development. Because rights metadata is inseparable from other metadata, and because the <indecs> framework specifies a general metadata framework, the work done in developing <indecs>rdd also deepened and expanded the original <indecs> framework, building on it.

The first stage of the initiative ended with the submission of the <indecs>rdd fundamental design to MPEG (ISO/IEC JTC1/SC29/ WG11) for the purposes of standardisation: in December 2001 the initial <indecs>rdd design specification was selected as the baseline for the MPEG-21Part 6 Standard for a Rights Data Dictionary.

The second stage of the initiative followed this acceptance, and the <indecs>rdd consortium, now renamed Contecs:DD, continued, with a revised set of partners, with two objectives:

The primary objective is to complete the MPEG-21000 Part 6 standard for a rights data dictionary. To achieve this, the consortium has retained Rightscom to work in collaboration with MPEG experts to create and edit the standard. It is expected that a Committee Draft will be produced in October 2002, at the 62nd meeting to be held in Shanghai, China. The Committee Draft will be balloted by MPEG national bodies and comments returned in time for discussion at the 63rd meeting, in March 2003. The International Standard could then be declared in the autumn of 2003.

The second objective, which both supports and is dependent on the first objective, is the development of a version of the rights data dictionary, complete with an extensible database design and interface, as the basis of an operational system. This objective is dependent upon the MPEG process, in that the database and dictionary must be conformant with the MPEG standard. Furthermore it also supports the first objective by providing proof that the standard can be implemented fully and that there is nothing in the standard that cannot be completely realised in operational terms. Some consortium members will use the developing dictionary in early implementations as a practical demonstration of their confidence in the work.

2.2 The need for a standard rights data dictionary

“Rights management” has to work in a computer environment. The huge amount of digital content now being traded – legally and illegally – requires an infrastructure for rights. What does “rights management” mean in this context?

The *terms* used in various “rights expressions” which mediate the use of digital items – ownership statements, licenses, permissions, offers, requests and agreements – need to be

unambiguously understood by computers. Together these terms are often called *rights metadata*.

For instance, if a license agreement states that a commercial *consumer* must *pay* a particular fee to *copy, play* and *keep* a particular format of a digital file in a particular time and place, and that a *student* may do the same for a reduced price, all those *terms* must be interpreted by a computer or user to mean what is intended by the licensor.

To achieve such a level of unambiguous interpretation, there must be a common *data dictionary* of terms involved in rights. This is a common requirement in computing, but in the area of rights management there are three problems which make it especially challenging.

2.3 Three Problems

First, rights are complex. Rights metadata can quickly become much more complicated than the simple license example given above: all kinds of media, content and usage might be involved, including rights in underlying abstract works; ownership of rights often changes over time. A rights data dictionary must be capable of supporting the simplest through to most complex of rights expressions.

Second, rights expressions will be mixed with other types of information. Agreements, offers and licenses may include any terminology taken from descriptive, legal or financial systems. A rights data dictionary must be broad enough to embrace terms from any other kind of metadata which might occur in a rights expression.

Third, many dictionaries are already in use. Different market sectors, individual companies and organizations may have their own working dictionaries and schemes (often called different namespaces). Some deal with rights, some don't. Many groups will not want, or be able, to change to a new dictionary, or use a new one alongside the terms from their own namespace. Yet because these groups are now all co-operating in common multimedia areas, some way of connecting them is essential. A rights data dictionary must be allow the use of terms from existing and future namespaces.

2.4 The solution: Contecs:DD

The architecture for Contecs:DD has been developed over a number of years to cope with just these problems of complexity and interoperability. A call by MPEG early in 2001 for proposals for an RDD standard provided the impetus for turning this architecture into a practical tool in the form of a distributable data dictionary with a range of powerful features. It combines the main elements of a data dictionary, a multi-lingual dictionary, an ontology and a thesaurus.

Contecs:DD is well suited to this task because:

It has a powerful conceptual base. Contecs:DD is based on a strong and mature underlying data model (the "Context Model") in which verbs are the starting point for all the most important definitions. This model provides a core of several hundred terms to which any number of others may be added in a systematic way.

It is highly structured. Every Contecs:DD term has a unique identifier and a "Genealogy" which defines precisely and logically how it relates to others. Because the underlying model is very rich, it can accurately describe very complex relationships between terms.

It is inclusive. Any terms from other dictionaries can be added to Contecs:DD (by assigning a unique identifier and a genealogy). Other terms are not just "extensions" or "mapped" words, they become an integral part of <indecs>rdd itself.

It is highly granular. Contecs:DD can support terms at any level of detail, fragmentation or versioning required by users.

Users can "mix and match" terms. Because any "mapped" scheme is part of Contecs:DD, terms from different namespaces can be combined to form rights expressions without loss of meaning. Each namespace may have and use its own set of Contecs:DD terms to create its own rights expressions.

It supports "transformations". Contecs:DD provides the semantic tools needed to translate terms from one scheme to another in a highly automated way. This is critical to allow different metadata schemes to co-exist in the multimedia environment. Software applications are still required as well, but Contecs:DD provides all the underlying semantic relationships.

It is legally neutral. Contecs:DD does not define legal terms. It can be used to make rights expressions which draw on any existing legal definition, or none.

It is business-model neutral. Contecs:DD terms can be used to describe any situation in which any kind of rights are owned, managed, protected or used, at any point in the life of content from origination to "end use" or archiving.

It is not a Rights Expression Language (REL). A data dictionary is not an expression language (such as XrML, now adopted as baseline technology for the MPEG-21 REL standard). An REL deals with the way in which terms are expressed in computer language; the dictionary defines the terms. An REL will use terms defined in an RDD.

It has uses beyond rights. Because of its generalized model, Contecs:DD can provide a comprehensive basis for metadata expressions and schemes for purposes other than rights – such as resource description, workflow management and event reporting. It could be used as a tool for the deployment of semantic based web services.

2.5 How Contecs:DD will be used

Contecs:DD is a tool which will be used in an automated way (often invisibly) to help to create, transform and interpret rights expressions.

It will:

Provide a ready-made standard terminology for rights management. Organizations needing to create rights expressions, or to enhance their existing metadata schemes, will be able to use Contecs:DD as a source for terminology. Apart from providing a structured basis for metadata selection, it ensures interoperability with other compliant schemes.

Be available in a variety of forms. The dictionary will grow constantly as other schemes are mapped, and so (as with "virus checking" software) regular updates will be an essential component.

Support application software at all points in the "content chain". Contecs:DD will be available to support the making, transforming and interpreting of rights expressions from origination to end use.

2.6 The Contecs:DD consortium

The Contecs:DD consortium represents major groups of rights owners and ancillary service providers. Currently its members are:

Enpia Systems Ltd.
International DOI Foundation
International Federation of the Phonographic Industry (IFPI)
MMG Ltd.
Motion Picture Association (MPA)
Recording Industry Association of America (RIAA)

2.7 The adoption of <indecs> principles as the basis of MPEG-21's Rights Data Dictionary

In November 2001, we announced that IDF was a founding sponsor of the <indecs> consortium to develop a Rights Data Dictionary - a common dictionary or vocabulary for intellectual property rights (Contecs:DD). The dictionary resulting from this activity was adopted as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard. The MPEG-21 Rights Data Dictionary, based on Contecs:DD principles, will provide a key part of the architecture required to deliver interoperability between develop a digital rights management (DRM) standard systems.

In July 2002 the MPEG-21 Multimedia Description Schemes Subgroup completed the Committee Draft for MPEG-21 Part 6 'Rights Data Dictionary (RDD). (See Bibliography – The Internet and e-commerce in Intellectual Property) RDD forms the basis of all expressions of rights and permissions as defined by the MPEG-21 Rights Expression Language. The MPEG-21 REL and RDD work together to allow the machine-readable expression of rights associated with the use of multimedia. These parts will be finalised by MPEG over the next year.

The goal of MPEG-21 is to "define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. Its scope is the integration of the critical technologies enabling transparent and augmented use of multimedia resources across a wide range of networks and devices to support functions such as: content creation, content production, content distribution, content consumption and usage, content packaging, intellectual property management and protection, content identification and description, financial management, user privacy, terminals and network resource abstraction, content representation and event reporting".

From the Scope statement of Part 6, Rights Data Dictionary: "The Rights Data Dictionary (RDD) comprises a set of clear, consistent, structured, integrated and uniquely identified Terms (as defined in Clause 3) to support the MPEG-21 Rights Expression Language (ISO/IEC 21000-5). Clause 5 of the MPEG-21 Standard specifies the structure and core of this Dictionary, and specifies how further Terms may be defined under the governance of a Registration Authority, requirements for which are described in Annex C.

Taken together, these specifications and the Dictionary and Database together make up the RDD System. Use of the RDD System will facilitate the accurate exchange and processing of information between interested parties involved in the administration of rights in, and use of, Digital Items, and in particular it is intended to support the MPEG-21 REL. Clause 6 describes how the RDD relates to the MPEG-21 REL.

As well as providing definitions of Terms for use in the REL, the RDD System is designed to support the mapping and transformation of metadata from the terminology of one namespace (or Authority) into that of another namespace (or Authority) in an automated or partially-automated way, with the minimum ambiguity or loss of semantic integrity.

The Dictionary is a prescriptive Dictionary, in the sense that it defines a single meaning for a Term represented by a particular RDD name (or Headword, see Clause 3), but it is also inclusive in that it recognizes the prescription of other Headwords and definitions by other Authorities and incorporates them through mappings. The RDD also supports the circumstance that the same name may have different meanings under different Authorities. The RDD has audit provisions so that additions, amendments and deletions to Terms and their attributes can be tracked.

RDD recognises legal definitions as and only as Terms from other Authorities that can be mapped into the RDD. Therefore Terms that are directly authorized by RDD neither define nor prescribe intellectual property rights or other legal entities."

IDF is now investigating effective ways for DOI Application Profiles to be built using the MPEG-21 RDD.

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.

DOI specific articles, papers and reports

- ANSI/NISO Z39.84 -2000 Syntax for the Digital Object Identifier http://www.niso.org/standards/standard_detail.cfm?std_id=480

- 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.

- Bide, Mark; "In Search of the Unicorn: The Digital Object Identifier From a User Perspective"; *BNBRF Report 89,* Book Industry Communication; February 1998.

- Paskin, Norman; "International DOI Foundation Annual Review 2000"; International DOI Foundation; August 2000; [doi:10.1000/193]

- Paskin, Norman; "Digital Object Identifiers"; Information Services and Use, Vol 22, 97-112; IOS Press, 2002

- Paskin, Norman; "International DOI Foundation Annual Review 1999"; International DOI Foundation; 1999; available from the IDF

- Paskin, Norman; "From one to many"; International DOI Foundation, Version 1.0, August 2000. [This paper explores the mechanisms and consequences of moving from "one DOI resolves to one URL" to "one DOI resolves to multiple data types". Now superseded by DOI Handbook]

- Paskin, Norman; "Digital Object Identifier: implementing a standard digital identifier as key to effective digital rights management"; International DOI Foundation; April 2000.

- Paskin, Norman; "DOI: Current Status and Outlook", *D-Lib Magazine*, Vol. 5, No. 5, May 1999. [a summary of DOI progress as of mid 1999]

- Paskin, Norman; Neylon, Eamonn; Hammond, Tony; Sun, Sam; The "doi" URI scheme for Digital Object Identifier (DOI); http://www.ietf.org/internet-drafts/draft-paskin-doi-uri-02.txt (October 2002)

- 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. ¾; 1998; JAI Press Inc., pp. 12-20.

- Rosenblatt, Bill; "Enterprise Content Integration with the Digital Object Identifier: A Business Case for Information Publishers", June 2002, http://dx.doi.org/10.1220/whitepaper5

- Software & Information Industry Association, Content Division Digital Rights Management Working Group: "The Digital Object Identifier: The Keystone for Digital Rights Management" May 2001. http://www.siia.net/divisions/content/doi.pdf

- Sidman, David; "Digital Object Identifiers: Not Just for Publishers"; *CMS Watch*; 31 March 2002;
  http://www.cmswatch.com/Features/TopicWatch/FeaturedTopic/?feature_id=66

- Sidman, David & Davidson, Tom; "A Practical Guide to Automating the Digital Supply Chain with the Digital Object Identifier (DOI)"; Content Directions, Inc. (First published in *Publishing Research Quarterly,* Vol. 17, No. 2, Summer 2001.)

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

- Wildstrom, Stephen H.: "A Library to End All Libraries". Business Week, July 23 2001
  http://www.businessweek.com/magazine/content/01_30/b3742031.htm

---

Identifiers

- Brown, Stephen A; Dunlop, John T (Introduction); Rivkin, Jan (Introduction). "Revolution at the Checkout Counter: The Explosion of the Bar Code". Harvard University Press. Oct 1997. ISBN: 0674767209

- Causton, Laurie; "Identifying and describing Web resources"; *Luxembourg: Office for the Official Publications of the European Communities;*1999 55pp; ISBN 92-824-1516-3

- Connolly, Dan. "Naming and Addressing: URLs, URIs...". *W3C Architecture Domain,* Updated 2001.

- Dack, Diana; National Library of Australia; Persistent Identification Systems - Report on a consultancy; May 2001
  http://www.nla.gov.au/initiatives/persistence.html

- Ehlers, Hans-Jurgen. "Identification Numbering in the Book, Library and Information World". ISBN Review 15 (1994) pp. 89-214

- Green, Brian and Bide, Mark. "Unique Identifiers: a brief introduction". *Book Industry Communication/EDItEUR.*

- Hill, Keith. "CIS - A Collective Solution for Copyright Management in the Digital Age". Copyright World, issue 76, 1997, pp18-25

- Hock, Dee. Birth of the Chaordic Age. Berrett-Koehler Publishers Inc. 1999 [development and social infrastructure of the VISA credit card identifier system]

- IETF Network Working Group. "Uniform Resource Identifiers (URI): Generic Syntax". August 1998.

- IETF URN Working Group. "Uniform Resource Names (urn)". Modified 31 July 2001.

- Library of Congress: Bibliographic Control of Web Resources: A Library of Congress Action Plan
  http://lcweb.loc.gov/catdir/bibcontrol/actionplan.html

- Liebowitz, Ed. "Bar Codes: reading between the lines". Smithsonian Magazine February 1999 (volume 29 number 11) pp130-146 [development and social infrastructure of the physical bar code identifier system]

- 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.

- Konicki, Steve. "Shopping for Savings". Information Week, July 1, 2002, pp37-48 http://www.informationweek.com/story/IWK20020628S0003

- Paskin, Norman; Toward Unique Identifiers; *Proceedings of the IEEE;* 87 (No.7) July 1999; pp. 1208-1227 http://www.ieee.org/organizations/pubs/pub_preview/PROC/87proc07_toc.html

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

- Paskin, Norman. "Identification and Metadata". To be published in Digital Rights Management: Technical, Economical, Juridical, and Political Aspects in the European Union), in the series "Lecture Notes in Computer Science" (Springer-Verlag, 2003)

- Payette, Sandra. "Persistent Identifiers on the Digital Terrain", *RLG DigiNews*, Vol. 2, No. 2, April 15, 1998.

- Powell, Andy. "Unique Identifiers in a Digital World". *Ariadne*, April 1997.

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

- Reynolds, Regina R. "ISSN and Seriality". *Library of Congress CONSER,* May 1998.

- Sun, Sam X; "Establishing Persistent Identity Using the Handle System"; *Tenth International World Wide Web Conference;* Hong Kong, May 2001; http://www.handle.net/papers.html

- Svenonius, Elaine; The Intellectual Foundation of Information Organization; MIT Press; 2000

- Uniform Code Council Website: http://www.uc-council.org [Physical bar code identifiers current information]

- URN implementers "Uniform Resource Names: a progress report" *D-Lib Magazine,* February 1996.

Metadata

- 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]

- Blanchi, Christophe; Petrone, Jason; "Distributed Interoperable Metadata Registry"; D-Lib Magazine; December, Vol 7, No.12; December 2001 http://www.dlib.org.ar/dlib/december01/blanchi/12blanchi.html

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

- Caplan, Priscilla; "International Metadata Initiatives: Lessons in Bibliographic Control"; *Library of Congress;* 6 November 2000.

- Denny, Michael. “Ontology Building: A Survey of Editing Tools”. O’Reilly Xml.com, Nov 2002
  http://www.xml.com/pub/a/2002/11/06/ontologies.html

- Dublin Core Metadata Initiative
  http://dublincore.org/

- Erickson, John. "The role of metadata supply chains in DOI-based, value-added services", *ICSTI Forum,* No.30, April 1999.

- Lagoze, Carl; "Keeping Dublin Core Simple: Cross-Domain Discovery or Resource Description?"; *D-Lib Magazine*; January 2001. [doi:10.1045/january2001-lagoze]

- Lagoze, Carl; Hunter, Jane; “The ABC Ontology & Model”; *Journal of Digital Information*; Vol2, Issue 2; Nov 2001
  http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/

- Munchen, KG ; “IFLA Study Group on the Functional Requirements for Bibliographic Records - Functional Requirements for Bibliographic Records”, 1998, http://www.ifla.org/VII/s13/frbr/frbr.pdf

- ONIX Product Information Standards
  http://www.editeur.org/onix.html

- Paskin, Norman. "Towards a Rights Data Dictionary -- Identifiers and Semantics at work on the net". *imi insights*, EPS Ltd, June 2002. EPS is a strategic consultancy and research company specialising in the impact of digital media on content businesses of all kinds - http://www.epsltd.com/

- Paskin, Norman. “Identification and Metadata”. To be published in Digital Rights Management: Technical, Economical, Juridical, and Political Aspects in the European Union), in the series “Lecture Notes in Computer Science” (Springer-Verlag, 2003)
  http://www.doi.org/topics/drm_paskin_20030113_b1.pdf

- Paskin, Norman. “On Making and “Identifying” a Copy”. *D-Lib Magazine*, Jan 2003
  http://www.dlib.org/dlib/january03/paskin/01paskin.html

- PRISM: Publishing Requirements for Industry Standard Metadata.
  http://www.prismstandard.org

- Rust, G; Bide, M;” The <indecs> Metadata Framework: Principles, model and data

  http://www.indecs.org/pdf/framework.pdf

- Rust, G.; Bide M; “indecs Summary Final Report”; 2000
  http://www.indecs.org/pdf/SummaryReport.pdf

- Rust, G; “Metadata: The Right Approach, An Integrated Model for Descriptive and Rights Metadata in E-commerce”; D-Lib Magazine
  http://www.dlib.org/dlib/july98/rust/07rust.html

- SCORM: Sharable Content Object Reference Model
  http://www.adlnet.org/

- Sowa, John F; “Knowledge Representation: Logical, Philosophical and Computational

Technology

- Active Digital Object Repository Architecture (ADORA)
  http://www.handle.net/wkshp_000920/ADORA-DOI.ppt

- Cross-Industry Working Team; "Managing Access to Digital Information"
  http://www.xiwt.org/documents/ManagAccess.html

- Kahn, Robert E. and Cerf, Vinton G. "What is the Internet (And What makes it Work)?", a copy of a paper prepared by the authors at the request of the Internet Policy Institute, December 1999.
  http://www.cnri.reston.va.us/what_is_internet.html

- Kahn, R.E and Wilensky, R; "A Framework for Distributed Digital Object Services"; 1995; Reston, VA: Corporation for National Research Initiatives (CNRI).
  http://www.cnri.reston.va.us/home/cstr/arch/k-w.html.

- Rey, Catherine A; "A System for Persistent Names"; *Web Techniques;* June 2001.
  http://www.webtechniques.com/archives/2001/06/rey/

- Sun, Sam; Lannom, Larry. "Handle System Overview". CNRI, July 2002.
  http://hdl.handle.net/4263537/4069

- Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle System Namespace and Service Definition". CNRI, July 2002.
  http://hdl.handle.net/4263537/4068

- Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". CNRI, July 2002.
  http://hdl.handle.net/4263537/4086

Infrastructure

- Borgman, Christine; "From Gutenberg to the Global Information Infrastructure"; MIT Press; 2000

- Biddle, P; England, P; Peinado, M; Willman, B. "The Darknet and the Future of Content
  http://crypto.stanford.edu/DRM2002/darknet5.doc

- De Soto, Hernando; "The Mystery of Capital"; Basic Books; 2000

- Gupta, A; Stahl, D; Whinston, A; "The Economics of Network Management *Communications of the ACM*; Sept 1999; Vol 42, No.9
  http://cism.bus.utexas.edu/works/articles/cacm99.pdf

- Liebowitz, Stan; Margolis, Steve; "Network Externalities (Effects)"; *The New Palgraves Dictionary of Economics and the Law*; MacMillan; 1998
  http://www.utdallas.edu/~liebowit/palgrave/network.html

- Norman, Donald; "The Invisible Computer"; MIT Press;1999 [see Chapter 6 – The Power of Infrastructure]

- Whitehead, Jr, EJ; "Control Choices and Network Effects in Hypertext Systems"; *Information & Computer Science*;1999
  http://www1.ics.uci.edu/~ejw/papers/whitehead_ht99.html

Applications: Digital journals and reference linking

- 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]
- Arms, William; Hillman, Diane; Lagoze, Carl; Krafft, Dean; Marisa, Richard; Saylor, John; Terrizzi, Carol; Van de Sompel, Herbert; A Spectrum of Interoperability; D-Lib Magazine; January 2002
  http://www.dlib.org/dlib/january 02/arms/01arms.html
- Atkins, H; Lyons, C; Ratner, H; Risher, C; Shillum, C; Sidman, D and Stevens, A; "Reference Linking with DOIs: A Case Study", *D-Lib Magazine,* February 2000. [doi:10.1045/february2000-risher]
- Brand, Amy; “CrossRef turns One”; May 2001.
  http://www.dlib.org/dlib/may01/brand/05brand.html
- Caplan, Priscilla and Arms, William Y. 'Reference Linking For Journal Articles", *D-Lib Magazine,* July 1999. [doi:10.1045/july99-caplan]
- Caplan, Priscilla; Beit-Arie, Oren; Blake, Miriam; Flecker, Dale; Ingoldsby, Tim; Lannom, Laurence; Mischo, William; Pentz, Edward; Rogers, Sally; Van de Sompel, Herbert; “Linking to the Appropriate Copy: Report of a DOI-Based Prototype”.
  http://www.dlib.org/dlib/september01/caplan/09caplan.html
- Davidson, Lloyd and Douglas, Kimberly. "Promise and Problems for Scholarly Publishing". The Journal of Electronic Publishing, Volume 4, Issue 2, December 1998.
- Hitchcock, Steve et al. "Linking Electronic Journals". *D-Lib Magazine,* December 1998. [doi:10.1045/december98-hitchcock]
- Paskin, Norman. doi:10.1000/170 "E-citations: actionable identifiers and scholarly referencing. [Also available from "Learned Publishing", Vol 13, No 3, July 2000, [doi:10.1087/09531510050145308]
- Spedding, Vanessa; “JOURNAL CROSS-LINKING - The Web's potential untangled”; Research Information; July 2002
  http://www.researchinformation.info/feature1a.html
- Van de Sompel, Herbert & Hochstenbach, Patrick. "Reference Linking in a Hybrid Library Environment", *D-Lib Magazine,* October 1999 - Part 3. [doi:10.1045/october99-van_de_sompel]
- Van der Werf, Titia. "Identification, version control and future availability of electronic publications". Koninklijke Bibliotheek.

Applications: Learning Objects

- Barefoot, T; Flynn, T; Secure Internet Delivery of High-Value Content; Learning Objects Network; 2001
  http://www.learningobjectsnetwork.com/WhitePaper_SecureInternetDeliveryOfHighValueContent.pdf

- Cohen, Edward J; "The Emerging Standards Effort in E-Learning: Will SCORM lead the way? "; *e-learning Magazine*, January 1, 2002.
  http://www.elearningmag.com/elearning/article/articleDetail.jsp?id=6787

- Degen, B; Strategic Benefits of Standard Learning Objects; Learning Objects Network; 2001
  http://www.learningobjectsnetwork.com/WhitePaper_StrategicBenefitsOfStandardLearningObjects.pdf

---

Applications: eBooks

- Walter, Mark. "E-book Project Highlights Role of DOI in Selling Digital Content", *The Seybold Report,* Analyzing Publishing Technologies, June 18, 2001.

Applications: Data

- CODATA Task Group Web Site
  http://www.codata.org/2000tg.html

- E-BioSci - a European platform for access and retrieval of full text and factual information in the Life Sciences
  http://www.e-biosci.org/E-BioSci_overview.html

- IUPAC Task Group on Standardization of Physico-Chemical Property Electronic Data Files (IUPAC-CODATA Project)
  http://www.iupac.org/projects/1999/024_1_99.html

- Paskin, Norman; "Digital Object Identifiers and Digital Preservation of the Record of Science"; Proceedings of ICSTI Seminar:Symposium on Digital Preservation of the Record of Science - 14/15 February 2002; IOS Press; 2002
  http://www.doi.org/topics/020210_CSTI.pdf

Applications: Digital Preservation

- Bide, M; Potter, E J; Watkinson, A; Digital Preservation: an introduction to the standards issues surrounding the deposit of non-print publications; September 1999
  http://www.bic.org.uk/digpres.doc

- The Internet Archive Wayback Machine
  http://www.archive.org/

- Jones, Maggie; Beagrie, Neil; Preservation Management of Digital Materials: A Handbook; The British Library; October 2001

- Paskin, Norman; "Digital Object Identifiers and Digital Preservation of the Record of Science"; Proceedings of ICSTI Seminar:Symposium on Digital Preservation of the Record of Science - 14/15 February 2002; IOS Press; 2002
  http://www.doi.org/topics/020210_CSTI.pdf

- Zorich, Diane; " Introduction to Managing Digital Access: Options for Cultural and Educational Organizations"; Getty Information Institute;1998; ISBN 0-89236-546-3

Applications: Graphics

- Paskin, Norman; "Digital Object Identifier"; Proceedings from Technical Association for the Graphic Arts 54th Annual Technical Conference (April 14-17 2002); Feb 2002

The Internet and e-commerce in Intellectual Property

- Barlas, Chris; Rust, Godfrey; “Information Technology -- Multimedia Framework -- Part 6: Rights Data Dictionary. Committee Draft”; International Organization For Standardization. ISO/IEC JTC 1/SC 29/WG 11: Coding of Moving Pictures and Audio. Document Reference: ISO/IEC JTC 1/SC 29/WG 11/N4943. Text of ISO/IEC CD 21000-Part 6; Date: 2002-07-26.
  http://mpeg.telecomitalialab.com/working_documents.htm

- Brin, David; The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?; Addison-Wesley; (1999) 384 pages

- Cox, Brad; “Superdistribution: Objects as Property on the Electronic Frontier”; Addison-Wesley; 1996.

- Erickson, John S. "A Digital Object Approach to Interoperable Rights Management*". D-Lib Magazine*, June 2001. [http://dx.doi.org/10.1045/june2001-erickson]
  http://www.dlib.org/dlib/june01/erickson/06erickson.html

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

- Gladney, Henry M. "Are Intellectual Property Rights a Digital Dilemma? Controversial Topics and International Aspects". *Information Impacts,* February 2000.

- Innella, Renato "Digital Rights Management (DRM) Architectures". *D-Lib Magazine*, June 2001. [doi:10.1045/june2001-iannella]

- Kahle, Brewster; Prelinger, Rick; Jackson, Mary E; “Public Access to Digital Material”, *D-Lib Magazine*, Vol 7 No.10, Oct 2001 [doi: 10.1045/october2001-kahle]
  http://www.dlib.org/dlib/october01/kahle/10kahle.html

- Kaser, Richard.T: “If Information Wants to Be Free . . . Then Who's Going to Pay for It?”. *D-Lib Magazine*, Vol 6 No.5, May 2000 [doi: 10.1045/may2000-kaser]
  http://www.dlib.org/dlib/may00/kaser/05kaser.html

- Lessig, Lawrence; “Code and Other Laws of Cyberspace”; Basic Books; 1999.

- Lessig, Lawrence; “The Future of Ideas: The Fate of the Commons in a Connected World”; Random House; ISBN: 0375505784; (2001)

- Levinson, Paul; “The Soft Edge: A Natural History and Future of the Information

- Levy, Steven; “Crypto”; Penguin Books; 2002; ISBN 0-140-244-328

- Paskin, Norman. "Position Paper for W3C Workshop on Digital Rights Management for the Web", January 2001

- Paskin, Norman: "Rights Management Through Digital Objects", presentation to the *Second International Conference on Electronic Commerce and Intellectual Property,* Geneva, September 19-21, 2001. (HTML version or PPT version)
  http://ecommerce.wipo.int/meetings/2001/conference/presentations/pdf/paskin.pdf

- Rosenblatt, Bill. "Solving the Dilemma of Copyright Protection Online". *The Journal of Electronic Publishing,* Volume 3, Issue 2, December 1997.

- Rosenblatt, Bill; Trippe, Bill; Mooney, Steve: “Digital Rights Management: Business and Technology”; Hungry Minds; ISBN 0764548891, 2002

- Seely-Brown, John; Duguid, Paul; The Social Life of Information; Harvard Business School Press; 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.

Briefings & Presentations from IDF

Please refer to the HTML version of the DOI handbook for links to the material cited.

- Introductory DOI System Slide Presentation - A high level summary of DOI System policies and operations, including a graphical illustration of DOI administration and resolution.
  http://www.doi.org/overview/Overview_DOI_System.ppt

- Detailed DOI System Slide Presentation - A detailed look at the DOI System policies and operations.
  http://www.doi.org/doi_presentations/021008shortDOI.ppt

- Presentation given by Norman Paskin at the BIC/NISO Meeting: Digital Preservation on 5 Dec 2002
  http://www.doi.org/doi_presentations/021205DigPres.ppt

- Workshops on DOI. The following presentations (HTML versions, including original PowerPoint slides) come from a series of DOI workshops.
  - Metadata
  - Resolution
  - Registration Agencies

- Description of a prototype of advanced DOI functionality using application profiles to inform client software of available services for a given DOI. (PPT version or PDF version)

- "DOI: The developing Standards Infrastructure" presentation by Norman Paskin, December 2001. (PPT version or PDF version)

- "DOI: The Case for Supporting the International DOI Foundation", presentation by the IDF, November 2001. (PPT version or PDF version)

- Frankfurt Book Fair 2001***
  - "DOI Update 2001: DOI Commercial Implementations," (HTML Version or PPT Version) presented by Norman Paskin, IDF.
  - "DOI Update 2001: Summary," (HTML Version or PPT Version) presented by Norman Paskin, IDF.
  - "DOI Update 2001: Rights Management through Digital Objects," (HTML Version or PPT Version) presented by Norman Paskin, IDF.
  - "The DOI-EB Project: Bringing the DOI to Market for the Book Industry," (HTML Version or PPT Version) presented by David Sidman, Content

Directions, Inc.
(See also the DOI-EB Project Demo presented at Frankfurt)

- "CrossRef Update," (HTML Version or PPT Version) presented by Ed Pentz, PILA, Inc.
- "Learning Objects and Learning Objects Network," (HTML Version or PPT Version) presented by John Purcell, Learning Objects Network, Inc. Inc.
-

Due to the variety of fonts and font sizes, HTML versions of these presentations are best viewed with Internet Explorer 5

- Additional presentations from previous DOI Workshops.

- Handle System Developers' Workshop, agenda and links to presentations, CNRI, Reston, Virginia, held 20 Sep 02