The DOI® Handbook

Edition 4.4.1 5 October 2006

Please send any comments or questions to

Dr Norman Paskin
International DOI Foundation, Inc.,
5 Linkside Avenue
Oxford
OX2 8HY
United Kingdom

Email: n.paskin@doi.org
Tel: +44 (0) 1865 559070
Fax: +44 (0) 1865 853120

www.doi.org

Contents

DOI® Glossary 6
1 Introduction 12
1.1 The DOI® Handbook 12
1.2 Identification and the Internet 13
1.3 What is an Identifier? 13
1.4 What is a DOI® Name? 15
1.4.1 DOI names are persistent identifiers 15
1.4.2 DOI names are actionable identifiers 15
1.4.3 DOI names are interoperable identifiers 16
1.4.4 Identifying at the appropriate level 16
1.4.5 Identifying copies and versions 17
1.5 Components of the DOI System 17
1.6 What can be identified by a DOI name? 18
1.6.1 Intellectual property 18
1.6.2 Identification of abstractions 19
1.6.3 Formal definition 19
1.6.4 Granularity 19
1.7 Benefits of the DOI System to Publishers, Intermediaries, and users 20
1.7.1 Summary of benefits 20
1.7.2 Benefits in internal content management 21
1.7.3 Benefits in the distribution and sales life-cycle 21
1.7.4 Benefits in the production life-cycle 22
1.7.5 Quantified benefits: case studies 23
1.8 The DOI System as social infrastructure 23
1.9 The DOI System as a managed system 25
2 Numbering 27
2.1 Assigning numbers 27
2.2 The structure of a DOI name 27
2.2.1 The DOI name prefix 27
2.2.2 The DOI name suffix 28
2.3 Uniqueness 29
2.4 Case sensitivity 29
2.5 Character sets 30
2.6 Publishing DOI names in print 30
2.7 DOI System and Legacy Identifiers 31
2.7.1 Using existing identifiers as a DOI Suffix 31
2.7.2 Using DOI names to relate existing legacy identifiers 32
2.7.3 Benefits of using legacy identifiers with DOI names 33
2.8 DOI System and check digits 33
2.9 DOI System and Internet identifier specifications 34
2.9.1 Generic identifier standards 34
2.9.2 URI implementation 35
2.9.3 URN implementation 36
3 Resolution 37
3.1 What is resolution? 37
3.2 Simple resolution 37
3.3 Multiple resolution 38
3.4 Handle System technology 39
3.4.1 Overview 39
3.4.2 Technical support of DOI name Resolution 39
3.4.3 Software support for use of DOI names 40
3.4.4 Handle System RFCs 41
3.5 The DOI System and other actionable systems 41
3.5.1 Relationship between the DOI System and the Handle System 41
3.5.2 Relationship between the DOI System and other “actionable identifier” systems 41
3.6 The resolution interface with Handle System technology 43
3.6.1 Native resolver 43
3.6.2 Proxy server 43
3.6.3 Other mechanisms 44
3.7 The maintenance of DOI name “state data” 44
3.8 The development of services 45
3.9 DOI System and OpenURL 45
3.10 DOI names and Persistent URLs (PURLs) 47
3.11 DOI System and the Domain Name System (DNS) 48
4 DOI® Data Model 49
4.1 Overview of the DOI Data Model 49
4.2 Aims of DOI Data Model policy 50
4.2.1 Interoperability 50
4.2.2 Administrative capability 51

4.3 DOI Data Model tools 51
4.3.1 DOI Kernel Declaration 51
4.3.2 indecs Data Dictionary (iDD) 53
4.3.3 Resource Metadata Declaration (RMD) for metadata interchange 53
5 Applications 56
5.1 Introduction 56
5.2 Basic Approach: DOI names, DOI APs, and DOI System® Services 56
5.3 Managing APs and Services 57
5.4 DOI Data Model and APIs 58
5.5 Management and administration of APs and Services 60
5.5.1 Registration of APs and Services 60
5.5.2 AP and Service Management 61
5.5.3 Ownership of Intellectual Property in DOI APs 62
5.6 Benefits 62
6 Policy 63
6.1 Policy formulation 63
6.2 Current areas of policy development 63
6.3 Current policies of the IDF 63
6.4 Data associated with a DOI name 65
6.5 Restricted use of the DOI System 65
6.6 Exclusivity and restrictions 66
6.7 Ensuring persistence 67
6.7.1 Persistence of the DOI System 67
6.7.2 Persistence of the IDF 68
6.7.3 Persistence of the resolution technology (Handle System®) 68
6.7.4 Persistence of the metadata technology 68
6.7.5 Acceptance by users 69
6.7.6 Persistence of the identified object 69
6.8 Ensuring uniqueness 70
6.9 Community autonomy 70
6.10 Separation of metadata, identifier, and entity 71
6.10.1 Separation of metadata and identifier 71
6.10.2 Separation of identifier from object 71
6.11 Use of DOI names by services other than the original publisher 72
7 The International DOI Foundation 74
7.1 Origins and status 74
7.2 Business and organizational principles 74
7.2.1 Cost recovery and its protection 74
7.2.2 The creation of Registration Agencies 75
7.3 DOI System business model 75
7.4 DOI System Oorganisation model 76
7.4.1 Initial model 76
7.4.2 Future model 76
7.5 Relationships between Registration Agencies 77
7.6 The IDF as the DOI System Maintenance Agency 78
7.7 The IDF and DOI System development 78
7.7.1 IDF Development process 79
7.7.2 TWG (Technical Working Group) 79
7.7.3 RAWG (Registration Agencies Working Group). 80
7.7.4 Application Working Groups 80
7.7.5 Governance of Working Groups 81
7.7.6 Intellectual Property Rights and Working Groups 81
7.7.7 Disclosure of information in Working Groups 82
7.7.8 Working Group Agreements 82
7.7.9 Third party development 83
7.8 IDF Email discussion lists 83
7.9 DOI System and Handle System® development 83
7.9.1 The Handle System Advisory Committee 83
7.9.2 Further development of related technology 83
7.10 DOI System and indecs development 84
7.10.1 The initial indecs project (Dec 1998-June 2000) 84
7.10.2 DOI System-specific development (2001) 85
7.10.3 indecs as input to MPEG (2001-2003) 85
7.10.4 The indecs Data Dictionary (2003 onwards) 85
7.10.5 InterParty (2002-2003) 85
7.11 Alliances and liaisons with other organizations 86
7.11.1 Alliances and liaisons 86
7.11.2 Formal relationships 86
7.12 Governance of the DOI System 88
7.13 Membership of the International DOI Foundation 88
7.13.1 Benefits of membership 89

7.13.2 Classes of membership 89
7.13.3 Membership fees 90
7.14 Working with user communities 91
8 Registration Agencies 92
8.1 How to obtain a DOI Prefix 92
8.1.1 Selecting a Registration Agency 92
8.1.2 Registering a DOI® name with the IDF 92
8.2 The role of Registration Agencies 92
8.2.1 Overview 92
8.2.2 Development of Registration Agency role 93
8.2.3 RA technical requirements 93
8.2.4 RA information management requirements 93
8.2.5 RA general requirements 94
8.2.6 Summary of functions 94
8.3 Business models for Registration Agencies 95
8.4 Criteria for becoming a Registration Agency 95
8.5 Formalizing the relationship between an RA and the IDF 96
8.5.1 Current procedures 96
8.5.2 Future developments 97
8.6 Registration Agencies and IDF governance and policy 97
8.7 Use of DOI® trademark in domain names 98
8.8 Fee structure for Registration Agencies 98
8.9 Operational and technical requirements for Registration Agencies 99
8.10 Registration Agency Local Handle Service (LHS) Operation 101
8.10.1 Advantages of using a Local Handle Service at an RA 101
8.10.2 DOI® Handle Service configuration 101
Figure 1 102
8.10.3 RA responsibilities 102
Figure 2 103
8.10.4 Prefix allocation and service configuration 103
Appendix 1 Excerpt from ANSI/NISO Z39.84-2005 Syntax for the Digital Object Identifier 105
ANSI/NISO Z39.84-2005 Syntax for the Digital Object Identifier 105
1. Introduction 105
2. Standards and References 105
3. Definitions 105
4. Format and Characteristics of the DOI 106
5. Maintenance Agency 107
APPENDIX A DOI Specifications 107
APPENDIX B Designation of Maintenance Agency 107
APPENDIX E Application Issues 109
Appendix 2 The Handle System® 111
A2.1 Handle System overview 111
A2.1.1 Introduction 111
A2.1.2 History and Applications 111
A2.1.3 Need for a General Purpose Naming System 111
A2.2 Handle Syntax 113
A2.3 Handle System Architecture 114
A2.4 Handle System scalability 114
A2.4.1 Storage 115
A2.4.2 Performance 116
A2.5 Building Handle System applications - tools 116
A2.6 Conclusion 117
A 2.7 References 117
Appendix 3 DOI® Data Model and sector applications 119
A3.1 DOI Data Model and text applications 119
A3.2 DOI Data Model and learning object applications 120
A3.3 DOI Data Model and MPEG applications 120
Appendix 4 indecs Data Dictionary 122
A4.1 Introduction 122
A4.2 iDD Structure 122
A4.3 iDD Terms: Attributes 123
A4.4 iDD Terms: Scope 124
A4.5 iDD Maintenance Agency 124
A4.6 Semantic Analysis and Mapping of Terms 124
A4.7 Benefits of mapping through iDD 126
Appendix 5 DOI® Resource Metadata Declaration 129
A5.1 Basic Attributes of the RMD 129
A5.2 Semantic engineering of RMDs 130
A5.2.1 Subtyping Attributes 131
A5.2.2 Substituting Subtypes 132

A5.2.3 Chaining Attributes 132
Resource 133
A5.2.4 Transferring Attributes 134
A5.2.5 Reciprocating Attributes 135
A5.2.6 Compositing Attributes 135
A5.2.7 Allowed Values and Codes 136
A5.3 Contexts as Attributes 137
A5.4 Journal-RMD XML schema 139
Appendix 6 DOI® Kernel Metadata Declaration: XML Schema 140
A6.1 The DOI® Resource Metadata Kernel Declaration 140
A6.2 Kernel Metadata Declaration: specification of data elements 141
A6.3 Allowed iDD values for Kernel elements 143
A6.4 Namespace 146
A6.5 DOI Resource Metadata Kernel XML schema – version 0.9 146
A6.6 DOI Resource Metadata Kernel XML schema enumerations – version 0.9 150
A6.7 Example DOI Resource Metadata Kernel Declaration 154
Selected Bibliography 156

DOI® Glossary

This glossary defines selected terms unique to the DOI® System, and other terms with meanings used in a specific way within the DOI System. Refer to relevant sections of the Handbook for more information. Italicised terms are cross-references to other entries.

Actionable:
Of *identifiers (1,2,3):* implying that in the context of a particular piece of digital network infrastructure, the identifier can be readily used to perform some action: e.g. in an Internet Web browser, it can be "clicked on" and some action takes place.

Application Profile:
A grouping mechanism for DOI® names; a set of DOI names which share some common characteristics. 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. Includes a set of metadata, comprising the kernel metadata (compulsory for all DOI names) and additional information applicable only to that particular sort of object. Each DOI name is associated with one or more Application Profiles. Special cases of Application Profiles are the *Zero-AP; Base-AP*; and *Restricted-AP* (q.v).

Application Programming Interface (API):
A technical tool used in managing DOI names in relation to underlying technologies. A description of functionality in a modular conceptual layer above the technology that provides the functionality; in the case of the *DOI System* the API provides specifications for using the *Handle System®* but avoids the need for users to address the Handle system directly and in depth. The API ensures the portability of any code written to address DOI System services and applications.

Base Application Profile (Base-AP):
A special case of an *Application Profile* in which the metadata set is identical to the *kernel*.

Creations:
Entities which are products of human imagination and/or endeavour in which intellectual property rights exist; resources made by human beings, rather than other types of resource (natural objects, people, places, events, etc). These may be manifested as Digital Objects, Physical Packages, Spatio-temporal Performances, or Abstract Works. They correspond to intellectual property as defined by the World Intellectual Property Organisation: "creations of the mind: inventions, literary and artistic works, and symbols, names, images, and designs used in commerce". The initial focus of DOI System applications is Creations; other types of resource are identified by DOI names only if involved in intellectual property transactions (or in internal DOI System administration), and may be identified by DOI names where appropriate.

Context:
The circumstances in which an act occurs, encompassing the entities of time, place, agent and resource. The context model is the basis of the *indecs* data model and ontology.

Digital Object:
A data structure whose principal components are digital material, or data, plus a unique identifier for this material. Note that a DOI name may be an identifier of such objects; but a DOI name is not solely an identifier of such objects as it may also identify other *entities* and *creations.*

DOI® name:
An *identifier (1)* which conforms to the *DOI System*. Since a DOI name may be used to identify any entity, it is a digital identifier of any object (in the sense of *entity*), not solely an identifier of *Digital Objects*.

DOI® Application Profile Framework:
An approach to efficiently grouping DOI names and connecting those groups to available services, giving application developers a way to go from a given DOI name to a set of services specific to that DOI name. This approach has been codified in the form of an API and answers the fundamental question of 'given a DOI name, what can I do with it?' In the framework, DOI names are linked into *Application Profiles*. Any single DOI name can be a member of one or Application Profiles; each Application Profile can be linked into one or more Services; and each Service can be made available through one or more interfaces. This layered approach is required both for usability and data integrity. As we add procedures and structures for layering DOI System information over the Handle System® it becomes increasingly important to define those procedures and structures in a separate layer. The Handle System is still there and its functionality is what enables the DOI System layer, but understanding and using the system will require building and using abstractions such as APs and Services.

DOI® Data Model:
The DOI® Data Model defines the connections between all of the DOI System operational components, enabling DOI names to be created and used in a consistent fashion. It combines the *DOI Application Profile Framework* and the *DOI® Metadata System* and the interaction between the two. Both of these major components have their own models and procedures, including the *Handle System* and the *indecs Data Dictionary*, which have been brought together and harmonized to effect true interoperability across DOI names from different domains and used within different applications.

DOI® Metadata System:
The metadata component of the *DOI System*. A set of three components for semantic definition that provide well-formed and interoperable metadata to support the use of DOI names: the *Kernel Metadata Declaration; indecs Data Dictionary*; and *Resource Metadata Declaration*. The specification is designed to maximise semantic interoperability with existing metadata element sets, and to work with the *DOI Application Profile Framework*.

DOI® name record:
The set of data (resolution values and administrative data) held in the *DOI System* as current *state data* associated with a specific DOI name and returned on *resolution.*

DOI® System:
The integrated system – comprising enumeration, description, resolution and policymaking – managed by the International DOI Foundation (IDF), providing an *identifier (4)*, i.e. an implemented system of labels through a numbering scheme in an infrastructure using a specification.

Entity:
Something that is identified. In principle a DOI name can be used (like any other URI implementation) to identify any entity. In practice 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 System applications is *Creations*. However other types of resource are also necessarily involved in intellectual property transactions, and so may be identified by DOI names where appropriate.

First-class:
Of an *entity*: an object in itself, not some attribute of an object; an address is an attribute of something, whereas the something itself is a first class object. A DOI name references an entity as a first-class entity, not simply the place where the object is located (it may then resolve to a location).

Global Handle Registry®*:*
A component of the *Handle System.* A unique *local handle service* which stores naming authority handles. A query to the Global Handle Registry® reveals which local handle services store which handles.

Handle:
An *identifier (1)* within the *Handle System.* DOI names are a subset of handles, but not the only one. The relationship Handles:Handle System is analogous to *DOI names:DOI System*. DOI names are distinguished from other handles by additional features and functionality, specifically *metadata* and *policy* forming the totality of the *DOI System* described in this Handbook.

Handle System®:
The *resolution* component of the *DOI System*. 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 DOI System is one implementation of the Handle System; hence a DOI name is a *Handle.* The Handle System is made up of *local handle services*.

Identifier:
(1) An unambiguous string or "label" that references an *entity* (e.g. ISBN 0-19-853737-9)

(2) A numbering scheme: a formal standard, an industry convention, or an arbitrary internal system providing a consistent syntax for generating individual labels or *identifiers (1)* denoting and distinguishing separate members of a class of entities (e.g. ISBN, or DOI Syntax NISO Z39.84).

(3) An infrastructure specification: a syntax by which any *identifier (1)* can be expressed in a form suitable for use with a specific infrastructure, without necessarily specifying a working mechanism (e.g. URI).

(4) A system for implementing labels (*identifiers (1)*) through a numbering scheme (*identifiers (2)*) in an infrastructure using a specification (*identifiers (3)*) and management policies (e.g. *DOI System*)

indecs™:
Acronym of "interoperability of data in e-commerce systems". The set of principles and tools forming a Framework deriving from the indecs project and subsequently developed into the *indecs Data Dictionary*; the basis for the DOI approach to metadata.

indecs™ Data Dictionary (iDD):
Part of the *DOI Data Model.* A structured ontology, developed from the indecs Framework, containing all Terms used in DOI® AP Metadata Declarations, ONIX messages and other schemes, and formal mappings of the relationships between them, based on *indecs.*

indecs™ principles:
Principles specified by the *indecs* project:

- 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.
- Metadata is a relationship that someone claims to exist between two entities

International DOI Foundation (IDF):
The organization established to develop and manage the *DOI System*.

Interoperability:
Enabling information that originates in one context to be used in another in ways that are as highly automated as possible.

Kernel metadata:
A compulsory small standardised set of *metadata* elements for every DOI name; the kernel is common to every associated *Application Profile*. Each Resource identified with a DOI name must have a supporting *Kernel Metadata Declaration.*

As the initial focus of the *DOI System* is on Creations, the initial kernel has been devised for Creations Other types of resource are identified by DOI names only if involved in intellectual property transactions, or in internal DOI System administration for the identification and management of Application Profiles and Services. When DOI names are applied to such other entities, the appropriate kernel will be defined.

Kernel Metadata Declaration (KMD):
The declaration of the *Kernel metadata.* For every resource identified with a DOI name, the metadata declaration mandatory for all DOI names with associated metadata (i.e. all but those in the *Zero-AP*), with standardised semantics and a supporting XML schema.

Local handle service:
A component of the *Handle System.* A local handle service is made up of one or more sites, and a site is made up of one or more handle servers. Handle servers store *handles.* One local handle service is unique, the *Global Handle Registry®.* "Local" implies simply that the particular service may be physically or logically confined to storing some administratively convenient subset of handles, such as all DOI names. Local handle services may also be implemented by a Registration Agency for their DOI names, for convenience.

Metadata:
(1) a single item of metadata is a relationship that someone claims to exist between two entities, each of which may have an identifier (an unambiguous string denoting an entity) and must, in an automated environment; a piece of data comprising the two entities and the relationship between them.

(2) (plural) a set of these relationships (data) having one entity in common, and thus forming data about that entity.

Multiple resolution:
Resolution returning a specific output of several pieces of current information related to the DOI name-identified entity, specifically at least one URL plus defined data structures representing *Application Profile* and *Service* information. Multiple resolution goes hand-in-hand with managing the object and its available services through structured metadata.

Native resolver:
A *resolver* plug-in that extends a web browser's functionality so that it understands the Handle protocol. It will recognize a DOI name 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 name (or types the DOI name into the address line in their browser) and the DOI name is resolved directly. A plug-in has significant advantages (performance, functionality, non-reliance on other capabilities) when compared with the use of the *proxy server* and an unextended browser, the more common user interface to the *DOI System* today. The development of additional services which depend on utilizing the full *multiple resolution* potential of the *DOI System* will necessitate the user being able to manage DOI name resolution directly.

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 *indecs Data Dictionary.*

Persistence:
The consistent availability over time of useful information about a specified entity: ultimately guaranteed by social infrastructure (through *policy*) and assisted by technology such as managed *metadata* and indirection through *resolution* which allows reference to a *first class entity* to be maintained in the face of legitimate, desirable, and unavoidable changes in associated data such as organisation names, domain names, URLs, etc. DOI

names resolve to information (metadata) about an identified object in a manner intended to persist over changes in location, ownership, description methods, and other changeable attributes. If the object ceases to be available, the DOI name at minimum indicates a valid but now defunct identifier.

Policy:
Rules determined by the International DOI Foundation for the operation of the *DOI System* in a consistent, predictable, and controlled manner to ensure longevity of the system as infrastructure.

Primary URL:
The default URL value which is provided from a DOI name *single resolution*; and also provided as one value in the case of DOI name *multiple resolution* and specifically denoted as such in the *DOI name record.* The default URL may represent an instance of the identified entity, or some relevant metadata.

Proxy server:
A gateway between the Handle System and HTTP which enables resolution of a DOI name in the URL http:// syntax (e.g. doi:10.123/456 would be resolved from the address: http://dx.doi.org/10.123/456). Any standard browser encountering a DOI name in this form will be able to resolve it without the need to extend the web browsers capability, unlike the use of a *native resolver.* The use of the proxy server and an unextended browser provides the more common user interface to the *DOI System* today. The core DOI name resolution service is used by the proxy but is not constrained by the proxy. DOI names used through a HTTP proxy server (in the "http://dx.doi.org" formulation as a URL) will continue to be persistent.

Registrant:
An organization that registers DOI names through a DOI® *Registration Agency.*

Registration Agency (RA):
An organization which participates in the *DOI System* by offering services to *Registrant*s by agreement with the *International DOI Foundation*, including prefix allocation, registration of DOI names, and maintenance of sufficient infrastructure to allow Registrants to declare and maintain DOI name *metadata* and *state data*, and optionally may offer additional services based on any business model.

Resolution:
The process in which an identifier is the input (a request) to a network service to receive in return a specific output of one or more pieces of current information (*state data*) related to the identified entity: e.g. a location (such as URL) where the object can be found. Resolution provides a level of managed indirection between an identifier and the output. In the case of the *DOI System*, the Handle System® technology provides the capability of DOI name resolution. In the *DOI System* resolution is from a DOI name to one or more pieces of typed data: e.g. URLs representing instances of the object, or services, or one or more items of metadata.

Conceptually two cases are distinguished which are of significance in the *DOI System*: *single resolution* and *multiple resolution*.

Resource Metadata Declaration (RMD):
Part of the *DOI Data Model.* A non-mandatory model with comprehensive semantics and an XML schema for declaring resource *metadata* of any complexity, based on i*ndecs.* A tool to extend DOI System metadata declarations to any desired level of specification whilst ensuring these are interoperable with kernel metadata and other Application profile metadata declarations.

Restricted Application Profiles (Restricted-AP):
Special cases of an *Application Profile* in which the metadata set is not available to all users.

Service:
A defined result from a defined action i.e., do X and the result will be Y. DOI System Services perform specific functions when presented with data from DOI Application Profiles. This may involve specific servers on the network or abstract notions such as a defined method for comparing dates in documents with dates in DOI name records. "DOI System Service" is used in the same generic sense as Web Services and the Grid Service architecture, but is not restricted to either of those models. Like Application Profiles, Services are managed separately, as classes; DOI names, APs and Services have many-to-many relationships.

Single resolution:
Resolution returning a specific output of one piece of current information related to the DOI name-identified entity, specifically a URL. The basis of early applications of the *DOI System*, and of persistent identification through naming.

State data:
The current values of the data held in the DOI name resolution service in association with a particular DOI name, returned in response to a *resolution* request.

Uniform Resource Identifier (URI):
An *identifier (3)* specification for "any resource" on the Internet. A DOI name is a URI, and hence the DOI System forms an *identifier (4)* system which implements the URI specification.

Well-formed:
Metadata elements and schemes which adhere to the *indecs principles* and so can be understood outside a particular metadata scheme.

Z39.84:
The formal specification of the syntax for creating the DOI name identifier string (i*dentifier (2)*), "ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier"

Zero Application Profile (Zero-AP):
A special case of an *Application Profile* in which the metadata set is empty. Only limited functionality is possible with DOI names registered in the Zero-AP.

1 Introduction

This Chapter describes the DOI® Handbook and its updating process; explains the environment which leads to the need for the DOI® System, and outlines the components and use of the DOI System.

1.1 The DOI® Handbook

This Handbook is intended as:

- a definitive reference to the DOI System for the non-technical reader;
- a central point of reference for more complex technical content, through the appendices;
- a means of providing updated information to be disseminated; through the release of new versions.

The Glossary of Terms defines selected terms unique to the DOI System, and other terms with meanings used in a specific way within the DOI System which are discussed in the Handbook.

Other introductory material may be found in:

- The Frequently Asked Questions section of the DOI.ORG® website.
- Slide presentations, overview and demonstrations on the DOI.ORG website.
- Documents listed in the Selected bibliography.

The Handbook is regularly updated to reflect progress in DOI System development. The primary publication medium of the Handbook is the DOI.ORG website; users working with print versions are advised to ensure that they are using the most up-to-date version by checking the version on the DOI.ORG website (http://dx.doi.org/10.1000/182). Earlier versions of development documentation are superseded by any later edition.

The numbering system of versions 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.

Edition 1 of this Handbook was issued in February 2001. Edition 2 was issued in February 2002 and followed by several updated releases throughout 2002. Edition 3 was issued in May 2003 followed by several updated releases throughout 2003. This Edition 4 (first release April 2004) incorporates substantial additional material especially on the DOI® Data Model and DOI Systems Applications, related appendices and revision of all other chapters.

If you have any questions or suggestions relating to this Handbook, please contact us on contact@doi.org; your input will help to improve future versions of the Handbook.

1.2 Identification and the Internet

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 and manage pieces of digital content. A widely implemented and well understood approach to naming digital objects is essential if we are to see the development of services that will enable content providers to grow and prosper in an era of increasingly sophisticated computer networking. The 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.

> *"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, 1999).

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. The IDF has developed, and continues to evolve, a fully implemented solution to this challenge: the DOI System, using the DOI name, an "actionable identifier" for intellectual property on the Internet. The DOI System is now widely implemented by hundreds of organisations through millions of identified objects.

1.3 What is an Identifier?

For detailed information on concepts of identification and metadata, see the documents referenced in the bibliography. As the use of numbering in digital networks has developed, the use of the word "identifier" 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 distinguished. It is not possible to compare two "identifiers" unless it is clear which of the following is implied by each:

(1) ***A single unambiguous string*** or "label" that references an entity (e.g. ISBN 0-19-853737-9)

(2) ***A numbering scheme***: a formal standard, an industry convention, or an arbitrary internal system providing a consistent syntax for generating a series of labels *(identifiers (1))* denoting and distinguishing separate members of a class of entities (e.g. ISBN, or DOI® Syntax NISO Z39.84). 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 intention is establishing a one-to-one correspondence between the members of a set of labels (numbers), and the members of the set counted and labelled. The product of the process is enumeration, a cardinality judgement, and assigned numbers for each cardinal member.

The numbering scheme may or may not be accompanied by some policy apparatus – for example, a registration agency and maintenance agency. An important point is that the resulting number is simply a label string. It does not of itself create a string that is actionable in a digital or physical environment 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.

Common standard numbering schemes of interest in digital content management 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

(3) ***An infrastructure specification***: a syntax by which any *identifier (1)* can be expressed in a form suitable for use with a specific infrastructure, without necessarily specifying a working mechanism (e.g. URI). 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. The set of Internet specifications known as Uniform Resource Identifiers (embracing URLs and URNs) provides mechanisms for taking labels and specifying them as actionable within the Internet. The same principles apply in the physical 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. 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?

(4) ***A system for implementing labels*** (*identifiers (1)*) through a numbering scheme (*identifiers (2)*) in an infrastructure using a specification (*identifiers (3)*) and management policies (e.g. DOI System). The DOI System is an "identifier system" in the digital supply chain, just as the UPC/EAN is an "identifier system" in the physical 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 System is one of the

better developed, with several million DOI names currently in use by several hundred organisations.

1.4 What is a DOI® Name?

A DOI® (Digital Object Identifier) name is an identifier (not a location) for an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks. Unique identifiers are essential for the management of information in any digital environment.

It is an identifier in sense (4) above. One of the components is a syntax specification (*identifier (2)*). The DOI name conforms to a URI (*identifier (3)*) specification. It provides an extensible framework for managing intellectual content based on proven standards of digital object architecture and intellectual property management. It is an open system based on non-proprietary standards. It has the following notable features:

1.4.1 DOI names are persistent identifiers

A DOI name differs 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. The DOI System is not solely designed for use on the World Wide Web; the same functionality can be made available through any digital network and protocol, but the Web demonstrates its advantages well.

1.4.2 DOI names are actionable identifiers

The purpose of the DOI System is to make the DOI name an actionable identifier: a user can use a DOI name to do something. The simplest action that a user can perform using a DOI name is to locate the entity that it identifies. In this respect, a DOI name may look superficially like a URL. However, the technology which underlies the DOI System facilitates much more complex applications than simple location; and the DOI name identifies the intellectual property entity itself rather than its location. 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.

This does not imply that the DOI name will necessarily resolve to the entity that it identifies – although that will sometimes be the case. The DOI name, 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 name does identify a digital file, this will not always be the most appropriate or useful data for the DOI name 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 name identifies from what the DOI name resolves to. They may be the same thing, but they will often be very different.

The technology used to manage the resolution of the DOI name is the Handle System®; a description can be found in Appendix 2. The Handle System is unlike most other resolution technologies in supporting multiple resolution. A DOI name 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 name can resolve to different data, depending on the way in which the Handle System is queried. This enables the DOI name, 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 name to be resolved, the Registrant (or the Registration Agency he uses) needs to maintain the data associated with that DOI name 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 name can resolve to many other forms of data.

1.4.3 DOI names are interoperable identifiers

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

- "Legacy" identifiers can form an integral part of DOI names. Businesses can continue to use familiar – and proven – naming or numbering systems in this new environment.
- The Handle System is an efficient, extensible system designed to operate on any existing or future Internet service.
- The metadata (DOI® Data Model) component of the DOI System is designed to offer maximum interoperability of data through a structured, extensible, design resulting from significant work over the past few years in the indecs project and its successors.

1.4.4 Identifying at the appropriate level

An achievement of DOI® work has been a practical implementation of the idea of rethinking the Net as management of information, not movement of data packets. Managing information on the Internet at the appropriate level is a recurring theme in the vision of the future of the Internet.

As will be seen in what follows, a DOI name is not (only) an identifier of digital objects but (more widely) a digital identifier of objects – that is, it facilitates digital management of any entities (focussing on those involved in intellectual property transactions). Identification of non-digital entities, such as underlying abstractions (the "work") and physical manifestations are also needed in expressing real world transactions, and any technology which considers only "digital representations" is inadequate for digital rights management. 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 and persistently identified, and associated with data.

The indecs framework recognises the concept of functional granularity ("it should be possible to identify an entity when there is a reason to distinguish it"); this is echoed in the DOI System treatment of an identified entity as a first class object (an object in itself, not some attribute of an object). Whereas URLs are grouped by domain name and then by some hierarchical structure (originally based on file trees), DOI names offer a more finely grained approach to naming, where each name stands on its own, unconnected to any Domain Name System (DNS) or other hierarchy. The most common mechanism for resolution on the Internet is DNS (http as used in URL is a use of DNS). The Handle system used by the DOI System uses TCP/IP but avoids the need to use the DNS, and this has significant advantages. One advantage is that names are not implicated in trademark disputes. Another is flexibility over time as the document origins reflected in a hierarchy lose meaning, such as a change in ownership (if acmeco.com sells some assets to newco.com, all URL filenames beginning acme.com/ which pertain to the sale need to be changed. This benefit has already been seen in the case of Crossref, where millions of DOI names identified through the Academic Press IDEAL system were merged into Elsevier's Science Direct system when the companies merged). In order to manage DOI names we have created tools

that allow more flexible management of sets of DOI names, in a more useful way than as a fixed sub-domain: a DOI name, DOI® Application Profile and DOI System services can all be thought of as layers of abstraction which allow this. Functionality such as URL partial redirection and relative URLs (which assume as "known" or inherited a part of a URL / domain name address) make a lot of sense in the context of URLs. However since DOI names deliberately have a more finely grained approach to naming things, functionality such as partial redirection is dealt with through tools that capitalise on that finer granularity: precise definition of components and their associated services.

Identifying at the appropriate level is key to managing information. Too low a level of granularity makes it impossible to pick out important differences: too high a level of granularity makes it too complex to group similarities. Here is a good analogy from Jorge Luis Borges: *"Locke, in the seventeenth century, postulated (and rejected) an impossible idiom in which each individual object, each stone, each bird and branch had an individual name; Funes had once projected an analogous idiom, but he had renounced it as being too general, too ambiguous. In effect, Funes not only remembered every leaf on every tree of every wood, but even every one of the times he had perceived or imagined it. He determined to reduce all his past experience to some seventy thousand recollections, which he would later define numerically. Two considerations dissuaded him: he thought the task was interminable and that it was useless. He knew that at the hour of his death he would scarcely have finished classifying even all the memories of his childhood." ("Funes the Memorious").*

1.4.5 Identifying copies and versions

A common question is: if I identify entity A with a DOI name, and then I adapt it in some way to create entity B, should I assign a new DOI name to entity B?

The answer is: there can be no general rule which applies to all cases and each must be treated in context. If a registrant finds it useful to do so, they may. The rules of Application profiles, and business rules of Registration Agencies, will help in deciding for DOI names registered in Application Profiles. The key point is that one should precisely specify what A is and what B is; two digital entities are never the same in any absolute sense and can be considered copies of each other only in the context of some defined purpose.

For a more detailed explanation of this fundamental topic, see the article "On Making and Identifying a Copy" http://dx.doi.org/10.1045/january2003-paskin

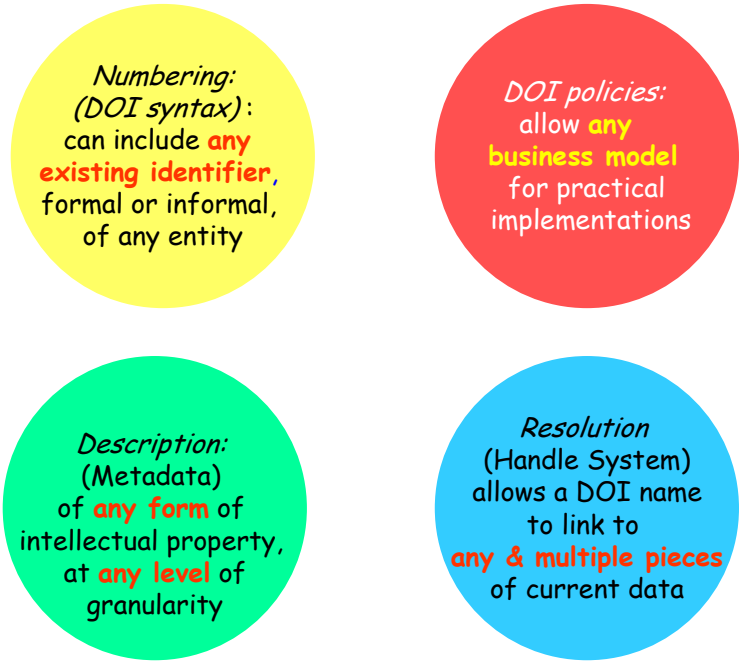
1.5 Components of the DOI System

The DOI System has four components:

- Numbering: assigning a number (or name) to the intellectual property entity that the DOI name identifies (It is more correct to talk about the DOI name as an alphanumeric string, since a DOI name may contain characters as well as numbers but we use the term "name" to apply to this string, to avoid unnecessary complexity). This is the NISO syntax, standardised as ANSI/NISO Z39.84-2000.
-
- Description: creating a description of the entity that has been identified with a DOI name, through the DOI Data Model, which is based on the indecs Data Dictionary.

- Resolution: making the identifier "actionable" by providing information about what the DOI name should resolve to, and the technology to deliver the services that this can provide to users; this uses the Handle System. The DOI System is an implementation of URI (Uniform Resource Identifier, sometimes-called Universal Resource Identifier, IETF RFC2396).

- 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", the DOI System 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 DOI names in a consistent quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

1.6 What can be identified by a DOI name?

1.6.1 Intellectual property

A DOI name can be used to identify any resource involved in an intellectual property transaction including, for example, text, audio, images, software, etc., and the agreements and parties involved. While the scope of intellectual property transactions is quite broad, it is unlikely that DOI names would be appropriate for identifying entities such as people or natural objects unless they are involved in such a transaction, or entities such as trucks. Intellectual property transactions don't necessarily involve money: DOI names can be used to identify free materials and transactions as well as entities of commercial value.

While a DOI name can be used like any other URI to identify "anything that has identity", the DOI System is a combination of components (identification, resolution, data model and policies) devised with the specific primary aim of identifying any "intellectual property entity". The initial focus of DOI System applications was "Creations" – that is, resources made by human beings, rather than other types of resource (natural objects, people, places, events, etc.). Other types of resource are also necessarily involved in intellectual property transactions, and so may be identified by DOI names where appropriate. As an example, the initial aim of the DOI System 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 the DOI System to museum artefacts or pharmaceutical components

which would be appropriate. Similarly, a DOI name 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 DOI names alongside the intellectual property that they govern.

Critically, the DOI name 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 name changes, but not the DOI name itself.

1.6.2 Identification of abstractions

Creations may be in both tangible and intangible forms. DOI names 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. This may be necessary for applications such as rights management or citation. 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. DOI names 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 name 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 name 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.

1.6.3 Formal definition

Formally, DOI System scope is defined in terms of a data model, the model underlying the indecs work: a DOI name can be assigned to any entity which is a Resource within the indecs context model. 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 used by the DOI System. The practical outcome of this is important and provides a pragmatic functional specification: a DOI name 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 Data Model.

1.6.4 Granularity

The DOI name 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 name. Rather the decision as to what a DOI name identifies is taken by the Registrant on a purely functional basis – what is it that I need to be able to identify? This is an application of what the indecs analysis calls Functional Granularity. The principle of functional granularity proposes that "it should be possible to identify an entity whenever it needs to be distinguished".

A DOI name 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 DOI names 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 con-

sider 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 name. Or a publisher could consider them two separate underlying works, hence two DOI names. 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 name 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 name for the work could be resolved to multiple additional DOI names for versions of the work, e.g., language, and each of those DOI names 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 DOI names), retaining the flexibility to aggregate them in one or more ways at a later date.

1.7 Benefits of the DOI System to Publishers, Intermediaries, and users

1.7.1 Summary of benefits

The DOI System offers a unique set of functionality:

- *Persistence, if material is moved, rearranged, or bookmarked;*
- *Interoperability with other data from other sources;*
- *Extensibility by adding new features and services through management of groups of DOI names;*
- *Single management of data for multiple output formats (platform independence);*
- *Class management of applications and services;*
- *Dynamic updating of metadata, applications and services.*

For users, these features provide the ability to:

- *Know what you have*
- *Find what you want*
- *Know where it exists*
- *Be able to get it*
- *Be able to use it in a transaction*

Some benefits which the DOI System enables:

- Links appear as standard hyperlinks, but (unlike URLs) are persistent: changes to URLs or other defined data as resolution points may be made without affecting the DOI name and any links bookmarked using it.

- Once DOI names are registered, they are available to anyone who wants to use them.

- Multiple-resolution options are registered along with the DOI name, and may be added or modified at any time. Using DOI System-aware software, links can pop up on a menu dynamically, in real-time, out of the global DOI System Directory – if the registrant adds an option to the DOI name record, it will show up the next time a user clicks on a link based on that DOI name.

- DOI names may be cut and pasted: users might encounter the DOI name as text in e-mail rather than on a web site, etc;

- A DOI name is a universal, machine-readable numbers allowing cross-system communication – like the ISBN for physical books, or the CUSIP # for Securities, yet without the need for a separate standard – and so usable in all back office functions, which are not visible on the surface – e.g., allowing the retailer to track sales according to a specific number, then report sales back to the publisher, have the publisher's internal systems tracking sales, performing accounting, calculating royalties, etc.

- DOI names can incorporate existing identifiers such as ISBN, ISTC, proprietary identifiers, etc and add value to them and enforce interoperability; DOI names can utilise existing metadata schemes.

- DOI System metadata mapping is based on the indecs (interoperability of data in e commerce) framework. Conformance with this framework facilitates the use of DOI names with MPEG-and ONIX compliant tools for multimedia content management and digital rights management, and with other schemes following the same principles.

Some specific benefits of DOI names in various aspects of the supply chain are described below in more detail.

1.7.2 Benefits in internal content management

DOI names and associated metadata ensure accurate, interoperable and efficient product information is available both externally but also internally, reducing costs in many places:

- System Management cost overhead for existing systems: the costs of managing multiple systems holding the same data, system and interface maintenance, ongoing data cleansing, etc;

- Data handling cost associated with manual entry of data, data cleaning;

- Development efforts for new systems – costs associated with having to re-identify and validate data sources, complexity of data mapping and increased risk of errors, cleansing of data for every new project;

- Reconciliation costs associated with having to reconcile different data sources;

- Customer Query/Complaint cost associated with handling customer queries/complaints and loss of revenue and customers due to poor quality data;

- Lost sales opportunities: revenue loss associated with not selling existing stock due to missing or incorrect product information;

- Misinformation: costs associated with making decisions based on incorrect management information;

- Dissatisfaction with the company: inability to provide accurate product information results in dissatisfaction amongst staff and customers;

- Lack of knowledge of product suit: difficult to identify the entire relevant product suite, meaning that many customer-facing staff are unable to offer the complete range to customers;

- New system development is delayed, as each project needs to address issues such as identifying and cleansing data. Increasingly this may impede product development particularly for electronic offerings. Standardisation of product type and data definitions would increase speed of new product development.

1.7.3 Benefits in the distribution and sales life-cycle

- More sales. Any hyperlink anywhere on the Internet which refers to a product's DOI name now becomes an active, dynamically-controlled, updateable sales & marketing tool – clicking on it can pop up a menu of all the actions or services which the Seller wants to provide to the consumer, a single click away for the consumer.

- Managing new options. Changing these choices dynamically is easy and cheap for the Seller: 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 Seller simply updates one central DOI name record – at which point, thousands of links all over the Internet which refer to that DOI name will now reflect the new or changed menu options, even if already stored, bookmarked, or printed.

- Less labour, errors, and costs: Sales can be tracked more easily and more precisely (at the object level, or aggregated into new collections). Sales are tracked more cheaply and accurately 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 and licensing revenue can be tracked more easily, cheaply and accurately, not only reducing costs but also capturing revenue more fully.

- Cross-linking on a wide scale is feasible and inexpensive, both among a 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.

- Wider dissemination of content, through enhanced "discoverability" for all third parties and 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 System).

- Additional, incremental product revenue enabled by offering highly-targeted, customized information products from the same asset base to specific audiences who might not otherwise purchase that information. The DOI System enables object-level control over digital assets – finding them, recombining them, distributing them, tracking their usage, etc.

1.7.4 Benefits in the production life-cycle

- Unique, unambiguous, universal content identifier ID throughout the entire content lifecycle:

  - 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 analysis, etc.

- Greater corporate-wide leverage over assets which are otherwise invisible behind separate divisional content management systems keyed with separate content IDs.

- Cheap to implement: existing content management systems internal content ID field can be used to create a DOI name; so the internal ID also becomes usable by distribution/sales partners and all others.

- Reduces costs, streamlines efficiency, and increases the functional capabilities of all internal systems which manage digital assets; 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.

1.7.5 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 DOI names 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:

- Know what you have (users able to look at catalogues of content available throughout the enterprise);
- Find what you want (users able to search and browse for content to be used or repurposed);
- Know where it exists (able to see where the item exists within the organisation);
- 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.

1.8 The DOI System as social infrastructure

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 name is free at the point of use, there is a small fee to an assigner for creating a DOI name (a few cents). This is because we have deliberately chosen to make the DOI System a self-funding (though not for profit) system. Our task now is to show that the DOI System 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 DOI names 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 System: 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 name registration and charging fees for participation.

At the outset of the DOI System development, a very simple model was introduced whereby a prefix assignment was purchased for a one-off fee from the IDF. It was recognized at the outset that this fee structure was a starting point but would be insufficiently flexible for the long term. DOI names allocated using these prefixes purchased directly from IDF are registered without structured metadata: they are now defined as being in the zero Application Profile.

We are now in a process of migration to the long term aim of a wide variety of potential business models, using third party Registration Agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. The disadvantage of using direct prefix purchase is that IDF cannot offer the level of metadata support and social infrastructure support of the type which can be given by a Registration Agency. DOI name prefixes obtained directly from IDF may however be useful if you wish to experiment or consider developing your own applications. DOI name prefixes will now only be issued through this direct route at the discretion of the Managing Agent.

Our intention is that eventually all DOI names 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 DOI names allocated, numbers of DOI names 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.

DOI names 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 name prefix free of charge to organizations for limited experimental non-commercial uses at the discretion of the Managing Agent. 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 name 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 DOI names without charge. Hence DOI names 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 DOI names 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.

1.9 The DOI System as a managed system

Like Domain Name registration, DOI name assignment requires a fee and agreement to follow the defined standard and rules. This does not make the system closed, or commercial, but it does make it managed. The International DOI Foundation is a not-for-profit organization, not a commercial operation; however, the system has costs that need to be met. Persistence is a function of organizations, not technology: to support a persistent identifier system, a persistent organization needs to exist. The principle concern of a persistent organization is of continuing funding; hence the model selected for a long-term position for a DOI System organization was a body that is not reliant on external sources, such as grants or membership, but is a self-funding system that can be supported in perpetuity from its own resources. The IDF is currently undergoing controlled migration from its initial member-funded organization (like W3C) to an organization that is operationally funded.

The implementation of the DOI System adds value, but the implementation necessarily incurs some resource costs in data management, infrastructure provision and governance, all of which contribute to persistence. The mechanism chosen to recoup those costs incurred by the organization is a self-funding “franchise” business model, as used by the physical bar code UCC/EAN system, and other proven systems. This is funded by a fee for participation (which may optionally be passed on to registrants, waived, or subsidised by the operating entity), but not for use of a DOI name once issued.

To make such a system work effectively requires protection of the assets within the system (1) from illicit exploitation, and (2) for assured quality control. Illicit exploitation would include someone calling something a DOI name when it is not part of the system; this could be damaging to one or both of the financial health (avoiding payment of an issuing fee) or the quality (poor data) of the system. To prevent this exploitation requires the availability of legal remedies: specifically, the DOI System relies on copyright and trademark law to protect the “DOI” brand and reputation. The DOI System is not a patented system; the IDF has not developed any patent claims on the DOI System and does not rely on patent law for remedy.

The underlying technologies used by the DOI System also have similar considerations. The Handle System is used by IDF under licence from Corporation for National Research Initiatives, who have certain intellectual property claims to protect the misuse of the Handle System; <indecs> intellectual property (IP) is assigned to, jointly and solely, IDF and EDItEUR and made available freely but under stated terms to others (an example being the <indecs>RDD work contributed to MPEG 21).

There is a widespread recognition of the advantages of assigning identifiers as well as a widespread misconception that an abstract-free specification (like a URN or URI) actually delivers a working system rather than a namespace that still needs to be populated and managed. URLs, for example, have a clear technical infrastructure (standards for how they are made) but a very loose social infrastructure (anyone can create them once a domain name has been obtained, with the result that they are unreliable: they have no guarantee of stability, let alone associated structured metadata). Product bar codes, Visa numbers, and DOI names have tighter social (business) infrastructures, with rules and regulations, costs of maintaining and policing data, and corresponding benefits of quality and reliability. From this need for management stems some misconceptions about the DOI System funding and business model. The most common myths are:

- Myth: the DOI System is for, run by, or only to the benefit of, commercial publishers. The publishing community was the first to see the benefits of persistent identification and to attempt to build an open system (rather than a system for, e.g., a library or a campus); several publishers have not only joined the IDF but provided initial loan funding, and the initial Crossref application is in the publishing

sector. However, there is nothing to prevent any other application, or any non-publisher involvement.

- Myth: the DOI System is “a commercial packaging of something that is available for free elsewhere”. The practical implementation offered by the DOI System is more than a collection of the underlying technical specifications.
- Myth: the DOI System is “only for rights management”. Whilst that was the initial impetus, since rights management requires an extensible system, it is in fact applicable for any use.
- Myth: the DOI System is “untested” or unrelated to other activities. All of the components are proven in other contexts, and there are millions of working DOI names. The DOI System builds on the Handle System and <indecs>, and so it inherits the strengths and real-world testing of these: for example, the <indecs> approach has been validated by rigorous analysis in the MPEG 21 framework development. These underlying technologies (rather than the DOI System per se) are often appropriate to answer the question of “how the DOI System relates to X”.
- Myth: the DOI System “allows only one business model” (seeing a swan and claiming that all birds are white and swim). As more applications are developed, the flexibility of a system that deliberately allows *any* business model will be appreciated.

2 Numbering

This chapter explains how a DOI® name is constructed and assigned. It discusses the use of the DOI name prefix as a naming authority, and the DOI name suffix as a mechanism for assigning individual numbers within that naming authority, incorporating (if required) existing identifiers. The ability of the DOI name to incorporate existing identifiers and the benefits of that approach are discussed in detail. Character sets, case sensitivity, uniqueness, and internet identifier specifications are also discussed.

2.1 Assigning numbers

Each DOI name is a unique "number", assigned to identify only one entity. Although the DOI® System will assure that the same DOI name is not issued twice, it is a primary responsibility of the Registrant (the company or individual assigning the DOI name) to name each object within a DOI name 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 DOI names 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 name).

The DOI name 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 name, which should make it much easier for registrants to issue DOI names to all their existing "content assets".

However, the DOI name 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.

2.2 The structure of a DOI name

The numbering system of the DOI name follows a syntax standardised as ANSI/NISO Z39.84-2000 (see Appendix 1). The DOI name is also an implementation of a URI (Universal Resource Identifier) and is defined as such in an IETF RFC document.

In use, the DOI name is an "opaque string" or "dumb number" – nothing at all can or should be inferred from the number in respect of its use in the DOI System. The only secure way of knowing anything about the entity that a particular DOI name identifies is by looking at the metadata that the Registrant of the DOI name 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 name is called a "persistent identifier".

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

10.1000/123456

In this example, the prefix is "10.1000" and the suffix is "123456"

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 DOI names available.

2.2.1 The DOI name prefix

The prefix itself has two components.

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

There is no limitation placed on the number of DOI name prefixes that any organization may choose to apply for. For example, a publishing company might have a single 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 prefix, it can make transferring responsibility for the relevant DOI names rather more straightforward

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

The prefix may be further divided into sub-prefixes, for example:

10.1000.10/123456

Remember, though, that the DOI name 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 name has a prefix issued by a particular organization should not be used to identify the owner of any given intellectual property – the DOI name remains persistent through ownership changes, and the prefix is unaltered.

2.2.2 The DOI name 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 names.

The 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 (see more on this topic below). There are two possibilities on assigning a suffix: either (1) the entities are already numbered in some way, or (2) they are not yet numbered.

- DOI names may be used to identify entities which already have standard numbering schemes – e.g. books, which have ISBNs – and/or internal production numbers unique within a registrant's system. These numbering schemes can be used as a suffix, if that is convenient. The metadata component of the DOI name also deals with declaration of existing (or "legacy") standard identification schemes. Using a "legacy identifier" in this way also enables any "intelligence" in the legacy identifier to be retained for use in systems which recognize it.

  e.g. each of the following would be valid as DOI names:
  10.1000/123456
  10.1000/ISBN1-900512-44-0

  When a legacy identifier is incorporated into the DOI name in this way it is not intended to be interpretable as such within the DOI System (it may be useful as such outside the system in other applications). The check digit in such a number is not used by the DOI System, but may be retained without any problems arising; see more on this below.

- If there is no current numbering scheme used to identify the entity at present, the DOI name may become the numbering scheme *de novo*. In that case, the registrant will need to adopt or devise a suffix scheme: this could simply be a sequential production number.

It is not essential that all the registrants in a DOI System sector use the same mechanism for generating the suffix. In fact, use of a DOI name obviates the need for such standardization. A good example is the use of DOI names in identifying articles in CrossRef. Publishers use many different schemes which all form DOI names that can then be used together: e.g.

Publisher A uses PII**: S1384107697000225**
Publisher B uses SICI: **0361-9230(1997)42:<OaEoSR>2.0.TX;2-B**
Publisher C uses "C-numbers": **JoesPaper56**

These three schemes are not at all interoperable, but become so in the DOI System as:

doi:10.2345/S1384107697000225
doi:10.4567/0361-9230(1997)42:<OaEoSR>2.0.TX;2-B
doi:10.6789/JoesPaper56

Each publisher can retain his own scheme and does not need to switch to a new one, though all publishers need to agree on a common metadata set for their DOI names.

2.3 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 DOI names 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.

2.4 Case sensitivity

DOI names are case insensitive. 10.123/ABC is identical to 10.123/AbC. All DOI names 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 name 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 name already existed.

The DOI name syntax standard allows suffixes to be case sensitive, i.e. 10.123/ABC as completely different to 10.123/AbC and the two are different identifiers. The Handle System is configurable by service so as to be either case sensitive or case insensitive and therefore allows this. The IDF imposed the further parameter on the specification, to remove case sensitivity, after a detailed review of the consequences. As this restriction has been implemented from an early stage, we have not introduced any cases of two DOI names 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, since DOI names can only be assigned via the IDF.

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 DOI names 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").

This argued in favour of case insensitivity being the safer, and more robust, option for future evolution and development of the DOI System, even though the DOI name syntax had

already been set. It is expected that this additional restriction parameter will be reflected in a revised version of the formal syntax standard in due course.

2.5 Character sets

DOI names 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).

When thinking about prefixes, suffixes and character sets, it is important to distinguish the DOI System from the underlying technology, the Handle System. The DOI System 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 System 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 DOI names.

Encoding. The Handle System at its core uses UTF-8, which is a Unicode implementation 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 name 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 name syntax is considered within the NISO standard. 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.

2.6 Publishing DOI names in print

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

Whereas in a digital context a DOI name 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 DOI names in print for e.g. journal articles tells people what an article's DOI name is, but it doesn't tell people how to access it on the Web; readers will not necessarily know that the DOI name is actionable. To do that, one may print the DOI name in a readily recognised form such as the http proxy server 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 name 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 DOI names in some other formulation. 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 name and a way to resolve it online (a shorthand way of saying "the DOI name 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

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

DOI names 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 name 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.

2.7 DOI System and Legacy Identifiers

An aim of the DOI System is to allow existing numbering systems to be retained, and the functionality of DOI names added to them easily.

2.7.1 Using existing identifiers as a DOI Suffix

An existing standard identification system number may be incorporated into a DOI name, if the registrant finds it convenient to do so (it is course recommended that precisely the same entity be identified by the two systems). The DOI System 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 name might consist of the ISTC (International Standard Textual Abstraction Code number): the DOI name 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 DOI names identifying related entities, such as manifestations of the textual work, or related textual works, even if these are not identified by ISTCs. The DOI name/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). PII is used by several scientific publishers as an internal numbering system. (PII and

DOI System 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 name, a specific DOI System implementation can create interoperability where none existed before. For example, in the CrossRef implementation of the DOI System, some publishers create their DOI names 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 DOI names, 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 name 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 System service, using the kernel declaration of this element. Since, as we have stated earlier, DOI names are inherently opaque non-parsable strings, the legacy identifier will not be securely recovered from the DOI name 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 name more easily human readable, and be administratively desirable, even though it is not a requirement of DOI name creation.

2.7.2 Using DOI names to relate existing 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. 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").

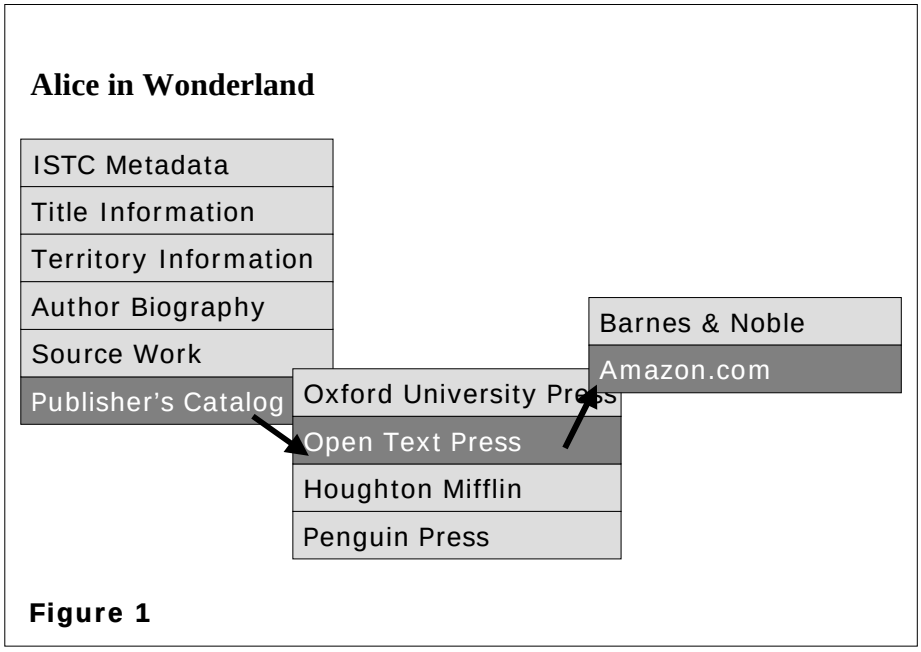


Figure 1

Figure 1 shows as an example a possible DOI System implementation of textual work identification, where the DOI names correspond to ISTCs (i.e. the ISTC may be incorporated as the DOI name suffix), in this case implemented as pop-up cascaded windows in a web browser (note that DOI name 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 name resolution (from the DOI name to the current value of the associated datum). The specific choices selected for implementation is not specified by ISTC or DOI System 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 name (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 name.

- 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 name 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 name (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 DOI names 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. The DOI System detailed technical architecture of Application Profiles and DOI System Services would be used to instantiate this example application.

2.7.3 Benefits of using legacy identifiers with DOI names

In addition to the benefits common to any DOI name there are some benefits specific to the incorporation of an existing standard numbering scheme into a DOI name:

- DOI names have a common, controlled prefix structure but allow complete domain-specific structures at the suffix. 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. Use of the DOI System 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.

- Users building applications of these DOI names are free to invoke all or only some of available multiple-resolution options – e.g., a DOI name 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 name for its own purposes – the Registration Agency offers registration of the full superset of options.

2.8 DOI System and check digits

A check digit is not compulsory or necessary in a DOI name, but if you wish to include one you may. Identifiers such as URL and URI specifications, deriving from an Internet envi-

ronment, do not have check digits: the underlying TCP/IP protocol they use has an error-correction component. Identifiers such as ISBN and similar bibliographic or documentation identifiers do have check digits: these act as aids to readability or keyboard data entry in the absence of any automated protocol correction.

The DOI name is deliberately designed as an opaque string, so that it is suitable for any use. The DOI System does not itself make use of check digits. However, other applications may make use of them, or may require them: so if you wish to incorporate a checksum digit into a DOI name you may. This could be useful for some other application. You may use as the suffix an existing string with a checksum (e.g. ISBN). You can also calculate the checksum across the whole DOI name if you wish (that would be akin to what the EAN/UPC does when it encapsulates an ISBN). Such a use of checksums in a particular DOI System application could be a rule of the DOI® Application Profile concerned: "your DOI names must include a checksum".

A check digit is usually the last in the sequence within an identifier string, algorithmically derived from the preceding digits, rather than being part of the identifier itself. The aim is to ensure that if one digit is incorrectly transcribed, the check digit will change as an alerting mechanism, and that if two digits are incorrectly transcribed, the chance of their combined effect on the check digit cancelling each other out is minimised. Recalculation of the check digit from the body of the number, followed by comparison with the stated check digit, can be performed automatically at key points in processing. Note that this provides error detection, but not error correction. In a typical check digit algorithm, each digit is assigned a different weighting factor (ideally a prime number). Digits and their corresponding factors are individually multiplied and summed, the resulting sum divided by a prime modulus number, leaving a remainder being the check digit; using prime numbers minimises the chances of internal cancellation. Check digits occur in for example ISBN and ISSN numbers and in other contexts, e.g. bank account numbers; ISO has published a standard ISO 7064 for check digits. Check digits are typically of importance in an entry step (where identifiers have to be manually transcribed as input) and less important in a transmission step where error correction protocols are already in place, although their original introduction was to ensure consistency in both types of activity. This has led to the assumption that check digits are of less importance, in an Internet-enabled world, than had been assumed in earlier automation phases. Whether or not this is true depends to some extent on the consequences of an error slipping through: whether inputting an incorrect identifier generates an error message, or simply locates the wrong object. A message may be transmitted correctly, but contain incorrect initial input. Omitting check digits in bank account numbers would not provide adequate error protection for most users. So the choice of whether to include a check digit will depend on the nature of the application. DOI names can accommodate them if required.

2.9 DOI System and Internet identifier specifications

2.9.1 Generic identifier standards

Persistent and actionable object names are required for coherence in the digital realm. "Persistent and actionable object names" thus necessarily require mechanisms for *persistence* (provided by social infrastructure); *actionability* (resolution from a name to some service); *specification of an object* (either through simple referencing or more formal description); and *naming syntax* (prescriptive rules for assigning identifiers in a standard format and ensuring uniqueness). The DOI System uses as its naming syntax the NISO standard *DOI syntax Z39.84.* The DOI System uses for its name resolution the Handle System (*IETF RFCs 3650, 3651, 3652*). The DOI System uses for its optional object specification a DOI Data Model and the *indecs Data Dictionary* and its subset the *ISO MPEG 21 Rights Data Dictionary, ISO/IEC 21000-6*. (The data dictionary component is designed to maximise semantic interoperability with existing metadata element sets; the data model allows descriptions to be grouped in meaningful ways so that certain types of DOI names all behave the same way in an application). DOI name persistence is guaranteed through the IDF social infrastructure which provides rules for registration, formal resilience procedures in the event of any single agency failing, etc.

A standard represents an agreement by a community to do things in a specified way to address a common problem. Whilst the DOI System community has developed the DOI

System, it has also ensured conformance with relevant generic external formal standards. This note discusses those relevant in the Internet communities IETF and W3C. There is currently considerable debate here on the issue of generic standards for naming objects. The DOI System is capable of being used in any specification which may finally be endorsed. Until a clear consensus is reached in the Internet communities on which approach is to be preferred the DOI System remains agnostic as to formal registration as a generic scheme, but useable and widely implemented for millions of objects.

The DOI System conforms to the functional requirements of the two generic approaches for naming first-class objects on the Internet: the Uniform Resource Name (URN) and the Uniform Resource Identifier (URI). URI and URN specifications deal only with syntax and (in part) associated implementation through resolution, not with description or persistence policy. Broadly, the URN approach is favoured by IETF and the URI approach by W3C, though there is considerable ongoing debate about each; some documentation on these specifications is incomplete. Crucially, widespread practical implementations of these specifications as object naming do not exist: both URI and URN are specifications, not in themselves working implementations. The DOI System is de facto a practical implementation of URI and URN.

The DOI System can also be implemented using current URL (http) specifications. The DOI System is also a defined Digital Item Identifier within the ISO MPEG 21 multimedia framework specification.

2.9.2 URI implementation

The Uniform Resource Identifier (URI) specification is IETF RFC 2396, URI Generic Syntax, currently under revision as RFC 2396 bis.

URIs formally encompass URNs as a sub set. In practice, the URI specification defines (1) an implementation more often called the Uniform Resource Locator, a location on a file server, commonly accessed using the http protocol though other protocols are allowed; (2) a syntax for referencing in XML, through which e.g. ISBNs can be specified as URIs. This provides a single framework which can accommodate any other identifier for referencing, but it is not as such persistent (since persistence is not determined by the specification but by the practical implementation). Conflating these two causes confusion. URLs as currently understood are demonstrably not persistent; redefining them as URIs doesn't fix that.

URL implementation. Users may resolve DOI names using the URL syntax through the DOI System proxy server (http://dx.doi.org). A DOI name of the form doi:10.123/456 would be resolved from the address: "http://dx.doi.org/10.123/456". Any standard browser encountering a DOI name in this form will be able to resolve it. The use of the proxy server does not interfere with any http requirements, so DOI names may be used with other http-based mechanisms such as OpenURL, PURL, parameter passing, etc. The proxy server is maintained by the IDF and the DOI System community for use by all.

URI syntax implementation. In the URI specification, the network path of the URI is implicitly DNS based; there are no real provisions to include systems that are not DNS based. Original URI specifications, and good design practice, assume the URI to be opaque (that is, it is not assumed that software can parse the body of the URI but that it would simply recognize the name of the scheme and hand it off to some other software that understood the scheme). The current URI specification, however, assumes that the initial URI parser will look into every URI, no matter what the scheme, looking for certain meaningful characters such as dot and slash. This version of the URI proposed in RFC 2396 bis is so restrictive that it is difficult to see what system could make use of it.

A specification for DOI name as a URI exists as an Internet Draft: this document defines the 'doi' Uniform Resource Identifier (URI) scheme for DOI names, which allows a DOI name to be referenced by a URI for Internet applications. The current revision of the URI specification, plus ongoing debate within the IETF and W3C communities on several proposed URI specifications, have delayed the processing of this Draft. DOI System implementation does not depend on implementation of this specification.

2.9.3 URN implementation

The URN (Uniform Resource Name) specification is RFC 2141 URN Syntax.

In practice, the URN specification defines (1) a formal registration process as a urn name-space, e.g., urn:doi:10.1000/1 and (2) accompanying specifications to implement a series of functional requirements for such namespaces.

Namespace referencing. One may specify any existing identifier as a URN: e.g urn:isbn:123456789, but this has no advantage over the simpler isbn:12345678. Such identifiers may be implemented using a specially written URN plug-in and resolved to URLs: functionally this gives nothing beyond the functionality achieved by coherent management of the corresponding URLs.

URN implementation. In order to implement the functional requirements, the URN architecture assumes an additional network service: a DNS-based Resolution Discovery Service (RDS) to allow a client to deal with a previously unknown URN type by finding the specific service appropriate to the given URN scheme. URN resolutions are then delegated to that scheme-specific resolution service. However no such deployed RDS schemes currently exist: browsers cannot action URN strings without some additional programming in the form of a "plug-in". The lack of any wide-spread infrastructural support will require any URN implementation to develop its own resolution mechanisms, such as plug-ins or proxy servers. Resolution mechanisms which require functionality beyond 1 URN to 1 URL also require the creation of data models. Several such implementations have been developed for specific uses where deployment to a closed group of users may be achieved; these carry no guarantee of ready interoperability with other deployments, which may require a different plug in for each implementation and may use conflicting data approaches.
The DOI System is not registered as a formal URN, despite fulfilling all the functional requirements, since URN registration appears to offer no advantage to the DOI System. It requires an additional layer of administration for defining the DOI System as a URN namespace (the string urn:doi:10.1000/1 rather than the simpler doi:10.1000/1) and an additional step of unnecessary redirection to access the resolution service, already achieved through either http proxy or native resolution. If RDS mechanisms supporting URN specifications become widely available, the DOI System will be registered as a URN.

2.9.4 DOI System functional requirements

The DOI System is designed to fulfil several functional requirements which we believe offer significant advantages in generic naming, notably:

Neutral as to implementation. The DOI System allows but does not require http or other protocols. The design principle is that DOI names are not specific to the web or any other implementation (e.g. information may be delivered in non-web platforms such as PDAs). The DOI System is designed to be applicable in any environment on the Internet (the global information system linked by a globally unique address space based on the Internet Protocol (IP) using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite).

Granularity of naming and administration at the object level. Allows but does not mandate coarser level granularity tools such as domain names. Specifically, DOI name resolution in native resolver form does not require the use of the DNS (Domain Name System): the DNS administrative model argues against using it as a general-purpose name system and has well-recognised problems of security and updating.

Neutral as to language/character set. Compatible with, but not restricted to, the ASCII character set. DOI names can use the Unicode capability of the Handle System to develop DOI names in Japanese, Chinese, etc characters. The current DOI System syntax restricts initial implementations to ASCII simply for ease of adoption, but is intended to be widened (backward compatibly) to Unicode at the next revision.

3 Resolution

This chapter explains the first main technical component of the DOI® System, resolution, and its ability to provide persistent association of the identifier and related data. Readers are advised to consult the Glossary of Terms at the start of the Handbook in conjunction with this chapter. The chapter describes in outline the Handle System used for DOI name resolution, and also discusses related technologies.

3.1 What is resolution?

A DOI® name is a name (identifier) for an entity in a network environment. Entities identified by a DOI name 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 System, using the Handle System® as a reference implementation, the resolution is from a *DOI name*, 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.

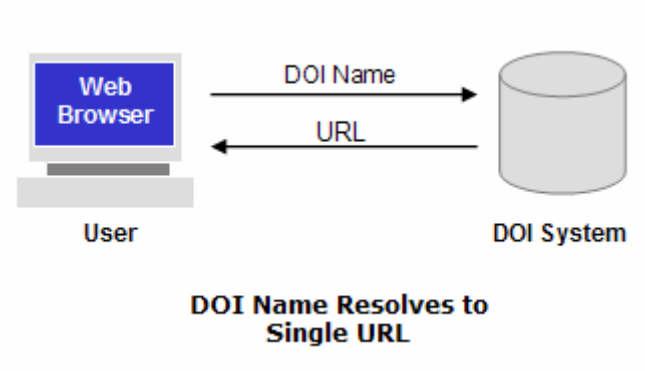
Using multiple resolution, a DOI name can be resolved to an arbitrary number of different associated values: multiple URLs, other DOI names, or other data types representing items of metadata. Resolution requests may return all associated values of current information, or all values of one data type; 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.

3.2 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 name 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 System was for simple, single point resolution. Each DOI name 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 System 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.

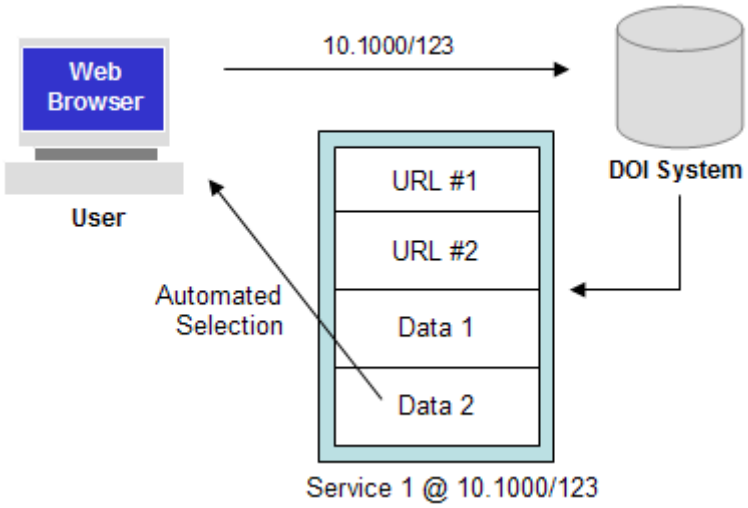
However, this lack of persistence of the URL is only the first – and the simplest – challenge that the DOI System was designed to manage.

3.3 Multiple resolution

Multiple resolution, allows one entity to be resolved to multiple other entities; it can be used to embody e.g a parent-children relationship, or any other relationship. This is a feature of the Handle technology, but the Handle System per se (deliberately) has no pre-existing constraints to make a useful framework (think of it as like spreadsheet software): the DOI System is an application of the Handle System which adds this constraint (think of it as like a spreadsheet application already written for you to add data to). In the DOI System the constraints come from the metadata which defines the entities, which is the data dictionary approach: hence our role in MPEG-21 RDD and the indecs Data Dictionary). That enables one to express relationships.

The solution to these challenges lies in automated “multiple resolution”. A DOI name can be resolved to an arbitrary number of different points on the Internet: multiple URLs, other DOI names, and other data types. If the DOI name 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 name will increasingly depend on automation of “service requests”, through which users (and, more importantly, users’ application software) can be passed seamlessly from a DOI name to the specific service that they require.

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



3.4 Handle System technology

3.4.1 Overview

Handle System technology, developed by Corporation for National Research Initiatives, 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

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

The Handle System is made up of local handle services (LHS). A local handle service is made up of one or more sites, and a site is made up of one or more handle servers. Handle servers store handles. One local handle service is unique, the Global Handle Registry®: the handles it stores, which are the naming authority handles, makes it the LHS to query to find out which services store all the other handles.

For further information on the Handle System, see the appendices and the General and Technical FAQs about Handle at http://www.handle.net/faq.html.

3.4.2 Technical support of DOI name 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 System technical infrastructure configuration and performance.

- Monitor research and commercial computing and networking developments that could enhance or limit the use of DOI System.
- Recommend, help initiate, and guide new prototypes to evolve the utility and expansion of DOI System usage, excluding production-level support when prototypes transition into operational mode.
- Represent the DOI System 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 System 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 System Directory Infrastructure

- Provision of 24x7 commercial hosting and secure facilities, back-up and disaster recovery, etc.
- Operational Management of the DOI System Service (DOI System Primary server (DOI System Local Handle Service); Mirror (secondary) of the DOI System 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 name 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 DOI names 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 name deposits.
- Administration tools to manage Prefix inventory
- Statistical analysis reporting
- Technical Consultation

4. Analysis and Communications Services

- DOI.ORG website and mailing list infrastructures

3.4.3 Software support for use of DOI names

There are a few servlets and tools that some users and programmers may find useful. (Contact the Handle System Administrator at hdladmin@cnri.reston.va.us for information.) These include:

net.handle.batch.DOIBatch

A batch loader for DOI names.

net.handle.apps.admin_servlets

The servlets used for administering handles via the web, useful if you'd like to allow DOI name 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 client 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 System specific libraries will soon be available with the Acrobat/DOI System services prototype.

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

3.4.4 Handle System RFCs

Here is a link to the current Handle System RFCs:

http://www.handle.net/rfcs.html

3.5 The DOI System and other actionable systems

3.5.1 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 a data model, 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 the Handle System, 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 name 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 name registrant; in return the IDF provides tools (e.g., The DOI® Handbook) to support DOI name 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 the DOI System is that our structured data model ensures ready interoperability between applications, which is of increasing importance. For this reason we are already in early discussions with some other Handle System applications about how we can encourage convergence of our approaches using the Handle System.

3.5.2 Relationship between the DOI System and other “actionable identifier” systems

The relationship between the DOI System and non-actionable identifier systems such as standard numbering systems, and the relationship with standard protocols such as URI and URN are discussed in Chapter 2. This section is concerned with other actionable systems,

i.e. systems which set out to deliver some or all of the functionality intended by the DOI System. 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 the Handle System)
- data model 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 the DOI System 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 data model, 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***
  DOI names, 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.

- ***Prefixes***
  Each registered handle type in the Handle System is itself represented by a handle of the form 0.TYPE/XYZ, where 0 is a prefix (also known as a naming authority) reserved for identifying system components, 0.TYPE is the prefox for handle types, and XYZ is the character string of the type itself. The type "NA" is reserved for prefixes, hence the Global Handle Registry® contains handles of the form 0.NA/10.1080 which resolve to "where do I find 10.1080"? Local handle services then consist of handles in the form 10.1080/456 which resolve to information about that specific handle.

  It is logically possible to envisage that e.g. "XIDs are DOI names in the form of a separate prefix". Note that there is no difference between X being allocated prefixes 10.XXXX, 10.100.XXXX or 20.XXXX; all handle prefixes are "peers". 10.100 is not a hierarchy below 10, all NAs are opaque. A handle 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 prefix issue is largely a business policy decision.

- ***Local Handle Service (LHS) usage***
  The question of where the NA information lives. The mapping of prefiox to an LHS is most efficient if done one-to-one (1 LHS to many prefixes 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 name users). Whereas DOI-AP and NA issues are a matter for IDF and X, the LHS issue should involve the Handle System Advisory Committee (HSAC) too (as it involves other non-DOI System applications issues).

3.6 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 names 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.

3.6.1 Native resolver

A "resolver plug-in" was developed by CNRI to enable a browser to resolve a DOI name in the form "doi:L10.123/456" without using a proxy server. The user simply downloaded and installed the plug-in and then "clicked" on the DOI name (or typed the DOI name into the address line in their browser) and the DOI was resolved directly. The resolver plug-in is currently being re-engineered and will be made freely available once again when the new version is complete. Software developers are encouraged to use the Handle System client library to develop their own custom client resolution software. (For information, see http://www.handle.net/client_download.html.)

3.6.2 Proxy server

Alternatively, without the need to extend their web browsers' capability, users may resolve DOI names that are structured to use the DOI System proxy server (http://dx.doi.org). The resolution of the DOI name in this case depends on the use of URL syntax: the example DOI name 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 name 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 System today. However, it has significant disadvantages when compared with native handle resolution. 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 System (and the Handle System technology) will necessitate the user being able to manage DOI name resolution directly.

The use of the DOI System 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.

DOI names 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 name (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 name (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 name 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 name 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 name-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

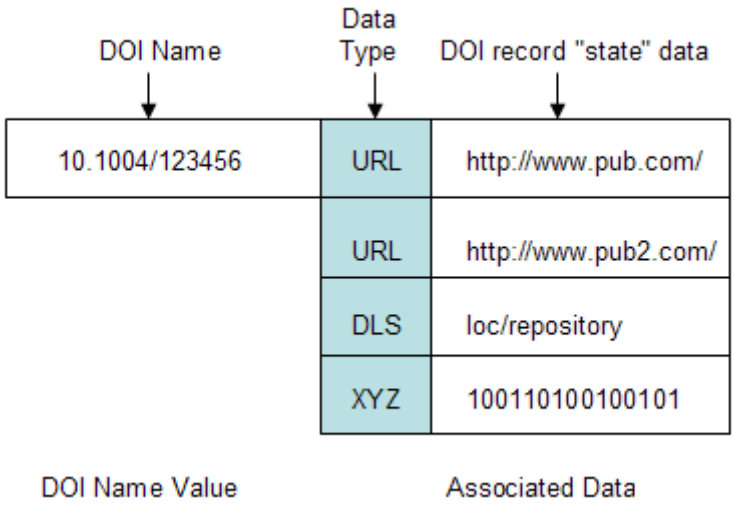
System. This, of course, is not at all unique and is just another variation on the Internet theme of layering services on top of one another. dx.doi.org is itself dependent on the Domain Name System (DNS), which is itself dependent on IP addressing and routing, etc. This picture will probably grow more complex as time goes on (we hope it does), with the core DOI name resolution facilities used in multiple ways and by multiple services. OpenURL resolvers, for example, will find DOI names 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 name proxy server(s), or using the handle protocol to query the DOI System directly.

3.6.3 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 System/Handle System 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.

3.7 The maintenance of DOI name "state data"

The effective operation of the DOI System depends on accurate resolution of a DOI name 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 name. 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 name record are conceivable and the requirements for these are currently being investigated by the IDF.

The data types to which a DOI name can resolve are fully extensible within the Handle System, to permit the DOI name 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 name data be entered as a full path, for example:

http://www.somepublisher.com/photo/photo#1.gif

rather than a relative reference. Whilst a relative link could be used as the DOI name data, we cannot predict the context in which the DOI name will be resolved, i.e. what the current base html reference will be.

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

3.8 The development of services

The development of services that make use of the potential of the DOI System 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 8) 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 System 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.

3.9 DOI System and OpenURL

> *"When you don't have decent metadata, it's hard to provide decent services. That's why I am an enormous fan of unique identifiers for objects, and systems that allow you to obtain well-structured metadata by using those identifiers. For me the big deal of the DOI/CrossRef framework is not necessarily the links they provide, because that might be done in other ways. The crucial importance of that work is in the mere fact that objects are being identified, and that identifiers can lead to metadata about objects. That changes the whole game."*
>
> Herbert Van de Sompel, Creator of OpenURL/SFX, in an interview with Dennis Brunning, The Charleston Advisor, Volume 4, Number 4, April 2003.

OpenURL is a NISO standard syntax for transporting information (metadata and identifiers) about one or multiple resources within URLs. OpenURL provides a syntax for encoding metadata (but not a source of it), restricted to the world of URLs (unlike the DOI System's wider application). This interface can be used to tie together otherwise disparate services such as centralized resolution systems and local knowledge of available resources.

The DOI System 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 DOI names to provide a richer user experience that incorporates both the global and the local requirements of the user. 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.

The logically centralized resolution service maintained by the content producers for DOI names has no way to resolve a DOI name to a locally held copy of the identified entity. So the synergy between the DOI System and OpenURL is clear: OpenURL needs a source of identifiers and authoritative metadata; the DOI System provides a 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 is a significant accomplishment in and of itself, but there are many opportunities for productive collaboration beyond that.

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).

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 with 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.

DOI names can be used within the OpenURL syntax to query local services about availability of resources at a local level, e.g. 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 DOI names that it has a local service for and offer that alongside the global information services obtained by resolving the DOI name through the Global Handle REgistry. OpenURL also allows more complex constructs that those illustrated above.

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. Several library service vendors provide such functionality. If the user is a bona fide member of a group, the local resolution service will be available to that user.

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 DOI names (provided by OpenURL-compliant service components) that can operate in a context-sensitive manner. The persistence of OpenURL is dependent 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 name 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: the DOI System is a global system; all information about a resource is the same in the global system wherever the DOI name is resolved from; the data that is associated with a DOI name 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 name 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.

> *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 has now been developed further into an active production level service used by CrossRef and a number of library service vendors to deploy localisation of DOI names referencing articles.*
>
> *For further information, see "OpenURL and CrossRef" at:*
> http://www.crossref.org/02publishers/16openurl.html

3.10 DOI names 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 as 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. PURLs are an approach to fixing the problem of unstable URLs. The 2002 OCLC Web Survey includes measurement of percent of IP addresses identifying a Web site in Year A also identifying a Web site in Year B: almost half of web addresses registered in one year are no longer reachable after one year. As time goes on this compounds: only 13% of the web addresses registered in 1998 were still around in 2002 (19% of the sites created in 1999 survived to 2002, as did 33% of the 2000 ones and 51% of those from 2001). The folly of relying on URLs alone for persistence is dramatically brought home by this statistic.

PURLs are all http and inherit both the strength and weakness of that approach. PURLs provide one level of indirection, just like a single value DOI name, but 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. The redirection is functionally equivalent to the way the DOI System uses a proxy server, dx.doi.org, which re-interprets DOI name queries into http. PURL is equivalent to a local DOI name which never goes beyond the proxy server approach and never makes use of the multiple resolutions and data types, metadata approach, and enforced common policy. The DOI System also provides a centrally managed redirection service rather than local purl server management.

We recommend that interested parties refer to independent comparisons e.g.: "To attach truly archival, long-lived names to network-accessible resources, I think PURLs should not be considered. My primary objection is that PURLs rely on DNS for labelling namespaces, which has at least two problems in the long run: DNS names are controlled by outside agencies at many levels (i.e. not just local administrators, but our ".EDU" parent domain is subject to the Internet governing bodies). Also, I believe the entire DNS naming system will be revised within the next 100 years, which is probably shorter than the range MIT Archives routinely anticipates. Although the Handle System currently needs the crutch of HTTP proxies which have the same DNS naming problem, it is inherently free of the domain name system and even the current Internet implementation. The handle namespace is not connected to any other protocol or standard, because it was properly designed to persist as a meaningful, resolvable naming system well into the foreseeable future." (http://web.mit.edu/handle/www/purl-eval.html)

The DOI System sits on top of a system explicitly designed to name digital objects on networks. This system, the Handle System, can provide the web-centric functions of a PURL through the use of a proxy server that returns a PURL-like single redirection. But underneath that is a much more extensive set of functionality that can be used as needed now or in the future.

A PURL can be resolved only by its designated PURL server. A DOI name can be separated into resolver and identifier. A given DOI name can be resolved by any handle resolver. A counter-argument involves the use of PURL partial redirects which can allow, for example, the single server at purl.org to route to what might be considered PURL subdomains to other PURL servers and to change this routing over time. But this puts the purl.org server in the middle of PURL resolutions forever. The dx.doi.org server could serve that same function and also be subject to the same problems, but it is only one way into the DOI System. dx.doi.org is just as deployed as purl.org and adds a more robust underlying system.

A PURL-based Object Identifier (POI) is a simple specification for resource identifiers based on the PURL system. The use of the POI is closely related to the use of the Open Archives Initiative Protocol for Metadata Harvesting and with the OAI identifier format used within that protocol. The main argument for POIs seems to be that they fit with OAI-compliant repositories. As described in the POI Resolver Guidelines, POIs are not explicitly assigned to resources – they are implied by the existence of an OAI metadata 'item' with an identifier that can be mapped to a POI. There are a number of assumptions: that an oai

metadata item will correspond to an available Resource; that resource will be available through a URL that can be derived from the oai-identifier; that identifier will work because the right kind of PURL partial redirect has been made. This can be contrasted with DOI names, which are not considered to exist before being explicitly registered and once registered are by definition part of the resolution system. Any implementation of persistent identifiers using existing material must accommodate DOI names, unless it plans to ignore the great bulk of scholarly journal literature.

The partial redirection rules that PURL uses to map OAI identifiers to POIs are simple and result in web redirections with a large degree of granularity. Currently POIs can only partially redirect OAI identifiers based on its namespace-identifier to a specific web server. Although this is not necessarily an issue, it does pose long term-collection management issues. This high level granularity is however not an intrinsic limitation of PURLs, indeed, PURL servers could use more sophisticated partial redirection algorithms and obtain a much finer level of redirection. This would however require PURLs servers to have a mechanism for expressing complex redirection mechanisms and the ability to promptly and accurately distributed them across all PURL servers. The partial redirection lets you move entire sub-trees from one location to another but doesn't let you rearrange the trees. This requires, as noted in the POI Resolver Guidelines, that the URLs are used in a consistent manner or at least that the base URL apply to all POIs in a given namespace. The registration of a separate PURL for each POI, which is the only way to begin to introduce the same level of flexibility given by the simplest use of DOI names, is given as a last resort case and recommended only for small numbers of POIs. One of the advertised strengths of POIs, that you don't have to register each one, also has the usual weakness of deriving identifiers from some aspect of the resource: defining who owns it and where it fits in their organization of resources. The DOI System policies and social engineering aspect we have referred to elsewhere (DOI names are backed by an organization dedicated to their growth and survival) add value as well as functionality to the DOI System.

3.11 DOI System and the Domain Name System (DNS)

DOI name resolution in native resolver form does not require the use of the DNS (Domain Name System), though does of course when used with the proxy resolver. The Handle System is more appropriate for large numbers of digital objects than DNS, and the DNS administrative model argues against using it as a general-purpose name system (DNS administration typically requires a network administrator, and has no provision for administration per name by anyone other than a network administrator). DNS also has well-recognised problems of security and updating which suggest that it will not be sufficient to assume that existing DNS technology should be adapted to deal with new requirements, rather than inventing something new: peer-to-peer networks already presage this.

URLs are grouped by domain name and then by some sort of hierarchical structure, originally based on file trees, now possibly unconnected from that but still a hierarchy. DOI names offer a more finely grained approach to naming where each name stands on its own, unconnected to any DNS or other hierarchy. This offers beneficial flexibility, especially over time, as the document origins reflected in that hierarchy lose meaning, such as a change in ownership which is reflected in DNS. A DOI name, DOI Application Profile and DOI System services are layers of abstraction which allow more flexible management of sets of DOI names, in a more useful way than as a fixed sub-domain.

Functionality such as URL partial redirection and relative URLs (which assume as "known" or inherited a part of a URL/domain name address) make a lot of sense in the context of URLs. However since DOI names/handles deliberately have a more finely grained approach to naming things, functionality such as partial redirection is dealt with through tools which capitalise on the finer granularity made available through DOI names.

For a detailed analysis, see the article cited in the bibliography: "Online Registries: The DNS and Beyond...", Release 1.0, September 2003. [doi:10.1340/309registries].

4 DOI® Data Model

This chapter explains the basis for the second main technical component of the DOI® System, the DOI Data Model, and its ability to ensure interoperability of DOI name metadata assigned through existing metadata schemes. The chapter gives an overview of the system, and then separate sections discuss the aims of the DOI Data Model policy – interoperability and good administration – and the three tools of the Metadata System – kernel metadata, the data dictionary and schemas for metadata interchange. Readers are advised to consult the Glossary of Terms at the start of the Handbook in conjunction with this chapter. For RAs and those wishing to explore this further, more extensive discussion of the issues and detailed specification of the components of the DOI Data Model are found in appendices 4-6.

4.1 Overview of the DOI Data Model

Without metadata, an identifier is of very little value. Metadata, which may be defined in this context as *information about an identified Resource*, provides human beings or machines with the data they need to enable them to make use of that identified Resource. Metadata may include names, identifiers, descriptions, types, classifications, locations, times, measurements, relationships and any other kind of information related to a Resource. *For a fuller review of the relation of metadata to the DOI System in general see the factsheet "DOI® System and Data Dictionaries" (www.doi.org/factsheets.html)*

There are two ways in which every IDF Registration Agency is bound to deal with metadata. An RA will gather *input metadata* from Resource providers (typically, descriptions of the Resources and associated rights and policies); and an RA will need to provide some level of output or *service metadata* to support DOI System services. Input metadata will provide some, but not necessarily all, of the service metadata. In some cases, a *metadata declaration* will itself be a complete DOI System service (for example, "provide an ONIX Product message for this Resource"). These two flows of metadata declarations are illustrated in figure 1.

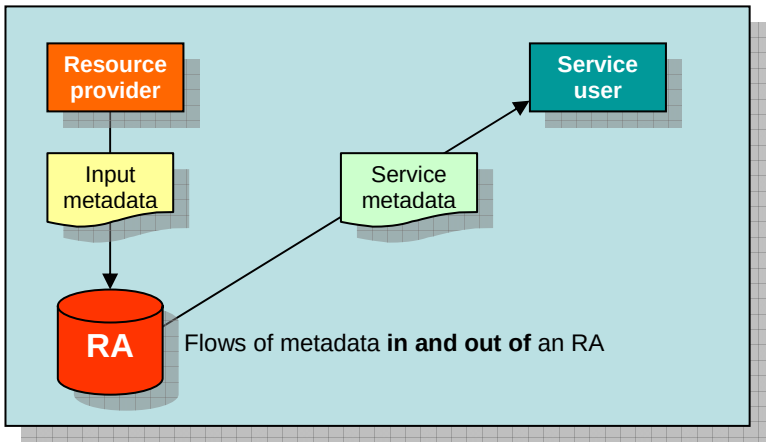


Figure 1: Flows of metadata in and out of an RA

DOI System policy places no restrictions on the form and content of an RA's input and service metadata declarations, except insofar as input metadata must support the minimum requirements implicit in the *DOI® Kernel* (see below). RAs may specify their own metadata schemes and messages, or use any existing schemes in whole or part for their input and service metadata declarations.

DOI Data Model policy is concerned with the internal management and exchange of metadata between RAs within the "RA network", and is designed to achieve two aims:

1. To promote *interoperability* within the network of DOI System users (see 4.2.1), and

2. To ensure minimum standards of quality of *administration* of DOI® names by Registration Agencies, and facilitate the administration of the DOI System as a whole (see 4.2.2).

The DOI Data Model has three tools to support its metadata policy:

- The DOI® Kernel Metadata Declaration (see 4.3.1)
- DOI® Resource Metadata Declaration schemas for data interchange between RAs (see 4.3.3)
- The indecs Data Dictionary ("iDD") (see 4.3.2)

The **responsibilities** of RAs can be summarized in these three statements:

1. An RA must be capable of producing a Kernel Metadata Declaration for each DOI name issued.

2. Metadata exchanged between RAs supporting DOI System services should be exchanged using an agreed DOI System Resource Metadata Declaration ("RMD") for the Resource or Service type.

3. Proprietary terms (data elements and values) used by RAs in Kernel and Resource Metadata Declarations should be registered in the IDF's data dictionary ("iDD").

These responsibilities are not mandatory for all DOI names: exceptions are discussed in terms of the requirement for interoperability described in the next section.

4.2 Aims of DOI Data Model policy

4.2.1 Interoperability

The first aim of DOI Data Model policy is to promote interoperability within the network of DOI System users. It does this by providing ways of achieving "semantic compatibility" between different RAs described in this chapter.

Standardization of any kind is driven by a need for interoperability. If an RA is issuing DOI names for Resources for use within a private domain where that RA is able to command all aspects of metadata gathering and output, then it has no need for standardization or conformance with DOI Data Model obligations. The RA will lay out its schemas and declarations, and its providers and users will, hopefully, conform to them. Such a situation is described as *restricted use of the DOI System*, and applies typically where an organization becomes an RA for the specific purpose of issuing DOI names for use only within its own private organization. *Restricted use is discussed more fully in section 6.5 of the Handbook.*

However, such isolation is unusual. Normally, when a DOI name is issued to a Resource, one fundamental assumption may be made about interoperability: *the RA or the Resource provider may wish (now or in the future) that the DOI name should be available for use in services provided by other RAs.* For example, where several RAs are issuing DOI names to journal articles from different publishers, it is likely that some RAs and publishers will want their DOI names to be included in journal-related services supported by other RAs.

In a similar way, many RAs will want DOI names issued by other RAs to be available for inclusion in services they themselves are providing. Such interoperability is one of the principal benefits of the DOI System.

As the RA network grows, such requirements are emerging, and where specific opportunities do not yet exist they are anticipated. In such circumstances neither the RA nor the

Resource provider wishes to issue a second DOI name for the Resource, nor to provide and capture the input metadata all over again from its source.

In addition, some DOI System services may not, in future, be the direct responsibility of RAs. Any service provider making use of DOI names issued by different RAs under different Application Profiles will be faced with the question of metadata interoperability.

Any DOI name which is *intended for interoperability* – that is, which has the possibility of use in services outside of the direct control of the issuing RA – is subject to DOI Data Model policy. The aim of metadata interoperability can therefore be expressed in these two objectives:

1. To ensure that metadata held by different RAs is *not fundamentally inconsistent*, and
2. To ensure that an *efficient and extensible means of interchange* exists for transporting metadata between RAs (and in future other service providers).

The first objective is dealt with by the DOI Kernel, and the second by the interchange provisions of the RMD and iDD.

The above provisions do not apply to DOI names registered under the legacy "Zero AP" described in Chapter 5.

4.2.2 Administrative capability

The second aim of DOI Data Model policy is "To ensure minimum standards of quality of *administration* of DOI names by Registration Agencies, and facilitate the administration of the DOI System as a whole". This aim may also be seen as supporting the first aim of interoperability, but it specifically addresses the need to ensure that a prospective RA is competent to issue DOI names responsibly and that ambiguous DOI names do not enter the network.

The Data Model policy provides a simple test for an RA's competence: the ability to make a DOI® Kernel Declaration, which ensures that the RA has an internal system which can support the unambiguous allocation of a DOI name and is fundamentally sound enough to support interoperability within the network. In addition, Data Model policy also requires that an RA maintains records of the date of allocation of a DOI name and the identity of the registrant on whose behalf the DOI name was allocated.

The metadata policy also exists to support the future development of mechanisms for facilitating the administration of the DOI System as a whole. This might be done, for example, through the use of iDD-registered terms as types to classify DOI names, Services or Application Profiles.

4.3 DOI Data Model tools

4.3.1 DOI Kernel Declaration

The Kernel Declaration, which is formally specified in an XML schema, answers a number of basic questions about the identified Resource (see Table 1). The answers to these questions should all be known by the RA at the time the DOI name is issued: if they are not, it will be questionable that the DOI name has been allocated unambiguously.

Questions about the Resource	***Kernel element(s)***
What is the DOI name being allocated?	DOI
Is it commonly referenced with another identifier (e.g. an ISBN)?	resourceIdentifier(s)
What is it usually called?	resourceName(s)

Who is principally responsible for its creation or publication? What role did they play?	principalAgent(s), agentRole(s)
Is it a *physical fixation*, a *digital fixation*, a *performance* or an *abstract work*?	StructuralType
How is it perceived – is it *audio, visual, audiovisual* or *abstract*?	mode(s)
What particular kind of Resource is it? (e.g. an *audio file, scientific journal*, *musical composition, dataset, serial article, eBook, pdf etc*)	ResourceType

Table 1: Kernel elements

There may also be a few questions about the issuing of the DOI name and Kernel itself (Table 2):

Which RA issued this DOI name?	RegistrationAgency
When was this Kernel issued?	IssueDate
Which version is it?	IssueNumber

Table 2: Administrative Kernel elements

The Kernel has one major function: it ensures that a basic set of interoperable, descriptive metadata exists so that DOI names can be discovered and disambiguated across multiple services and Application Profiles in a coherent way. The “AP1” Application Profile for Kernel metadata is under development to enable access to Kernel metadata for any DOI name. It is not mandatory that all DOI names should be accessible through such a service: but no such cross-network tool, however limited, would be feasible without a standard such as the Kernel.

Values of some Kernel elements (names and identifiers) are simply data strings. The other elements are drawn from sets of *allowed values*: for example, an agentRole might be “Publisher”, “Composer” or “Distributor”. These values may be expressed in different ways in code lists or “pick-lists”, and they may be more or less well defined, but what they share is that, for interoperability to succeed, the values used by different providers or RAs must be reconciled at some point through mapping.

Two Kernel elements (structural Type and mode) have a small, prescribed set of allowed values which all RAs must recognize. For the other elements and sub-elements, RAs may use their own choice of values, and add to them as and when required. These value sets must be registered in the data dictionary (iDD) for mapping purposes, so that any application using Kernel metadata from more than one source may be capable of presenting an integrated set of values to its users.

The use of certain standard values and the registration and mapping of other key values has another essential purpose: to ensure that metadata from RAs is *not fundamentally inconsistent*. For example, if one RA is issuing DOI names for digital fixations of journal articles, and another is issuing DOI names for abstractions of the same articles, the two cannot be used in the same way in the same service. Such distinctions are by no means self-evident, and unless they are made explicitly, using a common or mapped vocabulary, confusion is inevitable. As the RA network grows such confusion would result in costly problems and constraints on commerce.

The Kernel Declaration described here applies to resources in the form of *Creations* (items of intellectual property which represent the scope of early DOI System implementation).

However, other types of resource (such as Parties and Places) are also necessarily involved in intellectual property transactions and may in principle be identified by DOI names. As DOI names are applied to entities other than Creations, an appropriate Kernel will be defined.

Kernel metadata for all DOI names may be published under Application Profile AP1. Technical arrangements for the provision of Kernel metadata records through a generalized Kernel Metadata service are under development.

The detailed specification and XML schema for the Kernel Declaration is given in Appendix 6.

4.3.2 indecs Data Dictionary (iDD)

The indecs Data Dictionary (iDD) is under development as the repository for all data elements and allowed values used in Kernel Metadata declarations and Resource Metadata Declarations (RMDs).

The iDD enables the definition and ontology of all metadata elements to be available to all RAs, and provides the necessary mappings to support metadata integration and transformations required for data interchange between RAs. For example, if an RA wishes to consolidate metadata provided by several other RAs for a specific service, the iDD will provide the data mappings required to enable the RA to present the consolidated metadata as if from a single set.

iDD also contains mappings of "third party" schemes such as ONIX, the MPEG-21 Rights Data Dictionary and ISO Territory, Currency and Language codes.

The iDD is based on a contextual metadata framework developed under the <indecs> project to support interoperability of multiple metadata schemes (the IDF was a partner in the original indecs activity). The contextual structure of iDD supports mapping and transformation in a richer and more comprehensive way than conventional one-to-one "crosswalks". It is explicitly designed to enable metadata to be expressed in the simplest or most complex ways and transformed from one to the other.

iDD is a structured ontology compliant with logical axioms and constructors common to ontology languages such as W3C's OWL (Web Ontology Language). It can, for example, support the production of legal OWL ontologies.

All allowed values used by an RA in its Kernel Metadata, and all data elements used by an RA when mapping to an RMD, must be registered in the iDD. The iDD is administered on behalf of the IDF by an agency appointed for the purpose.

Each iDD-registered Term will have its own DOI name to support DOI System services accessing the dictionary.

A more detailed description of the iDD is given in Appendix 4. See also see the factsheet "DOI® System and Data Dictionaries" (www.doi.org/factsheets.html)

4.3.3 Resource Metadata Declaration (RMD) for metadata interchange

A DOI Resource Metadata Declaration (RMD) is a message designed specifically for metadata exchange between RAs. The format may also be used for input or service metadata, but it is not intended as a replacement for other domain or service specific schemes. An RMD is in the form of an XML document which conforms to an XML Schema (xsd). All its elements and allowed values are mapped into the iDD.

The first RMD ("Journal-RMD") was designed in the spring of 2004 for exchange of journal metadata used by several RAs to support different services. An RMD may be developed for any "domain", which may be defined in any way that a group of RAs requires. Typically these are expected to be for domains such as "eBooks", "sound recordings", "multimedia rights" or "educational coursepacks", which may be centred on a type of resource, or sector or function, supporting any group of Application Profiles or DOI System services.

Interoperability of RMDs will be ensured by a common structure and the underlying dictionary. The RMD uses a generic metadata structure of ten basic data element classes, developed from the <indecs> framework model and designed to incorporate all types of Resource metadata in a structured and flexible way. Table 3 shows the ten RMD basic elements, and to which class each of the more specialized Kernel elements belong:

Questions about the Resource	***RMD element class***	***Includes Kernel elements***
By what unique names is it known?	identifier	DOI, resourceIdentifier
By what non-unique names is it known?	name	resourceName
How is it described?	annotation	
What are its measurements?	quantity	
What kind of Resource is it?	category	structuralType, resourceType, mode
What has happened to it?	context	
Who has done something to or with it?	agent	primaryAgent, agentRole
When has something happened to it?	time	
Where has something happened to it?	place	
What other Resources are related to it?	resource	

Table 3: RMD basic element classes

Subtypes can be added to the ten RMD elements to any level of granularity: for example, an *identifier* might have a subtype of *ISBN*, or *vehicle registration number*, and a *relative* might be a *page* or an *edition*. The elements can be nested in any way required: for example, a *place* may have a *name* which has an *annotation*, or an *agent* may have a *category* which has an *identifier.* Elements can be grouped together in any combination in *composite* elements.

RMDs may incorporate data elements, allowed values, codes and composites from any other standard or proprietary message or metadata schemes (for example ONIX, SCORM or MARC) and draws on standard ISO codes and formats for Languages, Territories, Currencies, Measures and Dates and Times.

All element types and allowed values for an RMD are registered in the iDD. Every RA wishing to make use of an RMD must register the corresponding data elements and values in its own database to ensure reliable mapping by other RAs. A set of standard element groups or “composites” will be developed to form a core XML schema so that these composites can be re-used in different RMDs. The generic RMD structure and early iDD vocabulary is particularly appropriate for multimedia resource and rights metadata, but is in extensible for any Resource and domain.

RAs are free, of course, to use existing standards to communicate metadata between them where they are suitable. If, for example, two RAs are providing services requiring ONIX metadata, then it would be expected for one to provide ONIX message to the other. Likewise, one RA may wish to make different metadata records available to its users: a MARC-based RA may provide users with ONIX metadata records supplied by another RA for the same DOI name. The RMD is not a replacement for these, but to deal with a different issue: the integration of metadata from RAs and other sources using different standards (or none) where it is required.

The RMD and iDD combine to provide a generic solution for this problem, ensuring that all such interchange schemas within the RA network are themselves compatible and maximise the opportunities to re-use data and formats.

An RMD is developed with contributions from two or more RAs. RMDs are available for use by any RA. Any RA making use of a specific RMD may contribute to the editorial development of the RMD. An RMD will include the metadata elements required for all nominated services by any participating RA. Specific data elements within an RMD may be required only for specific RAs or Application Profiles, enabling the same RMD to be used flexibly within a community. The metadata flows between RAs, using RMDs and the iDD, are illustrated in figure 2 below.

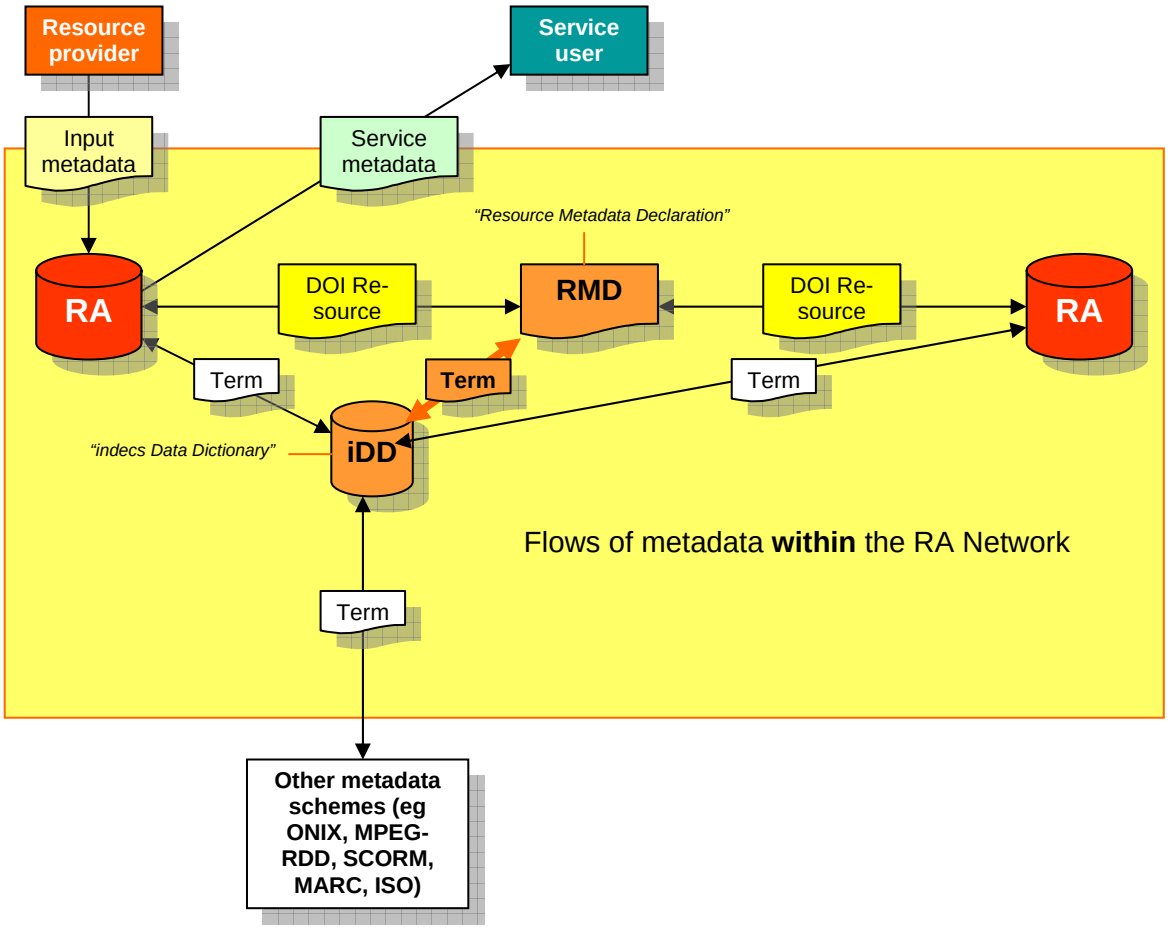


Figure 2: Flows of metadata within the RA network

More detailed discussion of the RMD data elements and structure can be found in Appendix 5.

5 Applications

This chapter introduces the concept of DOI® Application Profiles (APs) and associated Services. A DOI® Data Model and Application Programming interface allow application developers to work independently of the underlying technologies. Management of APs and Services is discussed.

5.1 Introduction

DOI® names can do more than provide the current location of an object; they can also provide or point to any useful information about the object. This information can include any type of descriptive metadata and any type of service related to the object, e.g., rights clearance or alerting services. Building applications to take advantage of this extended information requires that it be standardized in some way and that the DOI name owners have some way to manage all of this data.

This chapter introduces a few specific mechanisms as well as a data model and some policies, all aimed at making this extended DOI System data both understandable and manageable.

- ***Application Profile*** (AP) – a group of DOI names with some shared characteristic(s). This grouping will be used to connect DOI names to Services which are common to the group.
- ***Service*** – a specific transaction that can be performed on or with a DOI name.
- The ***DOI Data Model*** defines the way in which DOI names, Application Profiles, and Services relate.
- Lastly, multiple ***API***s have been built which allow application programmers to tap into the data model, using DOI APs and Services within their applications, without understanding the details of how they are implemented.

This approach has been proven in a concept trial used to build the DOI name/handle plug-in for Adobe Acrobat Reader (see the DOI.ORG web site).

5.2 Basic Approach: DOI names, DOI APs, and DOI System® Services

Each DOI name is associated with one or more Application Profiles, and each AP is associated with one or more defined *Services*.

DOI names are grouped into Application Profiles in order to manage their association with services. Any DOI name may, potentially, be associated with any number of services, e.g., extended metadata may be obtained by sending the DOI name in a certain manner to a certain address, a price quote for a translation service can be obtained here, related documents can be found there in a certain way, and so on. And the specifics of those services and the connections between a given DOI name and a set of services will likely change over time. Managing that complexity at the level of each DOI name, in a world of many millions of DOI names, is not tractable. Grouping DOI names into named Application Profiles and further associating APs with named DOI System Services allows simple changes, e.g., the address of a service location, to be done once and to reflect back on many DOI names.

A *service* is simply a defined result from a defined action, i.e., do X and the result will be Y. This will frequently involve specific servers in specific locations on the network but we also include more abstract notions such as a defined method for comparing dates in documents with dates in DOI name records. This involves a server but the client is then critical for finishing off the 'service.' We are using it in more or less the same generic sense as Web

Services and the Grid Service architecture, but what we mean is specific to DOI names and is not restricted to either of those models.

One of the services, possibly the only one for some DOI names, is the provision of kernel metadata (see Chapter 4) for each DOI name. Other sets of metadata may also be available for some DOI names and this, as with other services, would be known through the inclusion of a given DOI name in an AP and the association of that AP with the given service. Each AP is itself identified by a DOI name, e.g. doi:10.profile/7, and it is that DOI name, instead of each of the content level DOI names, that will be directly associated with the set of services relevant to each member of its set. Each of the registered services will also be identified by a DOI name, e.g. 10.Service/3. Associated with each service will be a natural language description of the service plus one or more interface descriptions (e.g., Interface Definition Language, Web Services Definition Language, or bindings and other information needed for using the service). The same service could be used by multiple APs through the simple mechanism of registering the service under multiple APs, just as a single AP will be 'used' by multiple DOI names.

APs may themselves be grouped into APs, i.e., they can be nested. This makes it possible to group them together for greater levels of indirection yet still maintain the flexibility to create new groupings and associations as needed. One of the challenges of applying the DOI AP and Service model will be finding the right level of granularity for the AP groupings. In some cases it may be that all DOI names from a certain producer may be grouped together in reasonable anticipation that any service that would apply to any of them would apply to all. In other cases the right level might be one particular product, e.g., a specific journal, again in anticipation that this will represent the group of DOI names for which deals will be made and services designed.

Granularity is also an issue in the creation and identification of services. If the same service, where 'same' refers to the function as opposed to the structural specifics, is implemented by multiple service providers and used by multiple APs, it should nonetheless be identified as a single service whenever possible. The details of the profile/service association can carry the information needed to go from the DOI name to the appropriate version of the service. Consider, for example, a rights clearance service in which the same input, a DOI name, yields the same output, e.g., current conditions attached to re-use. If multiple providers make that service available it would be advantageous to DOI System users if third party applications were able to offer all variations of the service in a transparent manner, going from one vendor to another as needed without any action on the user's part.

5.3 Managing APs and Services

The IDF will maintain the registry of APs and Services, i.e., it will serve as the registrar for the 10.AP and 10.Service prefixes just as the current RAs maintain the registry for the 10.1001, 10.1002, etc. prefixes. That relationship is summarized in the following table:

Type of DOI name	Create and Maintain	Registrar
Creation (Content) DOI names	Content Providers	RAs
APs	RAs	IDF
Services	RAs*	IDF

(Although all these entities have DOI names, APs and Services are assigned DOI names simply as a management mechanism, i.e., attributes of the Creation DOI name rather than as meaningful "containers" in their own right; we do not think it useful to treat APs as self-standing entities other than for internal management purposes – so we do not consider APs for DOI APs, etc.)

Further, the key connections between DOI names and APs and between APs and services are regulated by the RAs through the process of registering and maintaining the content DOI names and the AP DOI names.

*Note that creation of services by organizations other than RAs would require additional mechanisms and rules for registering services; initially services will be created only by RAs or in collaboration with RAs or IDF.

In creating a new AP an RA could choose to associate it with existing registered services or could create and register a new service or services with which to associate it. Existing APs can be used by new RAs, but the AP will continue to be maintained and governed by the RA which registered it in the first place. Contractual agreements could, of course, be put in place to manage the evolution of a shared AP, but only the original RA would be the manager of that piece of the infrastructure.

As illustrated in the following section, both APs and Services carry their own descriptions within the system. These descriptions are currently open-ended natural language explanations of the intent of their developers, i.e., what sort of DOI names are in this AP and why and, for Services, what will happen if a particular binding is used to instantiate this particular service. In order to both better manage and understand APs and Services it will be necessary to describe them with much greater precision. The proposed approach to that precision is to describe both APs and Services using DOI names for terms registered in the iDD (see Chapter 4). Thus, future versions of the Data Model will allow application builders to more precisely understand the intended relations between DOI names, APs, and Services and potentially link them together directly through the iDD taxonomy.

None of this constrains the creation of Services in any way. One could imagine these mechanisms being used to point at specific servers on the Net that would provide a defined service, e.g., rights and permissions information for the identified entity, and one could imagine a service that provides software for a client to use at its discretion, e.g., apply this script to validate the signature attached to any entity of this AP.

These mechanisms provide a great deal of flexibility but it should be noted that there is little way for any client to bootstrap its way into understanding what to do with completely new services: either a software client has been programmed ahead of time to know what to do with a Service or it hasn't and won't know what to do with it now. The bindings and descriptions that are associated with Service DOI names are intended as reference material to be consulted by application developers in designing and building applications. The variety of Services will be broad and implemented in a great number of clients; it may be difficult to build an application capable of bootstrapping its way into a Service that was not explicitly known to the application's developer.

5.4 DOI Data Model and APIs

Up to this point in DOI System evolution, most technical plans and descriptions, with a few notable exceptions, have been made in terms of the underlying Handle System®. The current level of complexity, where DOI names mostly resolve to single URLs, has made this possible. The added complexity of APs and Services, however, requires us to formally define a separate DOI System layer. This is required both for usability and data integrity. As we add procedures and structures for layering DOI System information over the Handle System it becomes increasingly important to define those procedures and structures in a separate layer. The Handle System is still there and its functionality is what enables the DOI System layer, but understanding and using the system will require building and using abstractions such as APs and Services.

Figure 1 shows an abstract view of this model, which has also been referred to as the Application Profile Framework. DOI names are linked into Application Profiles. Any single DOI name can be a member of multiple APs. Each AP can be linked into multiple Services. That linkage is to one or more specific instances of a Service. Each defined Service can be made available in multiple ways, referred to as instances. Using the rights clearance example above, and working from right to left in Figure 1, a rights clearance Service would be precisely defined and would be made available from multiple vendors using a variety of specific approaches. Each of those approaches would be an instance of the rights clearance service. Each of the instances would be usable by all of the members of one or more APs. This makes it possible to add a Service to many DOI names by adding that Service to relatively few APs.

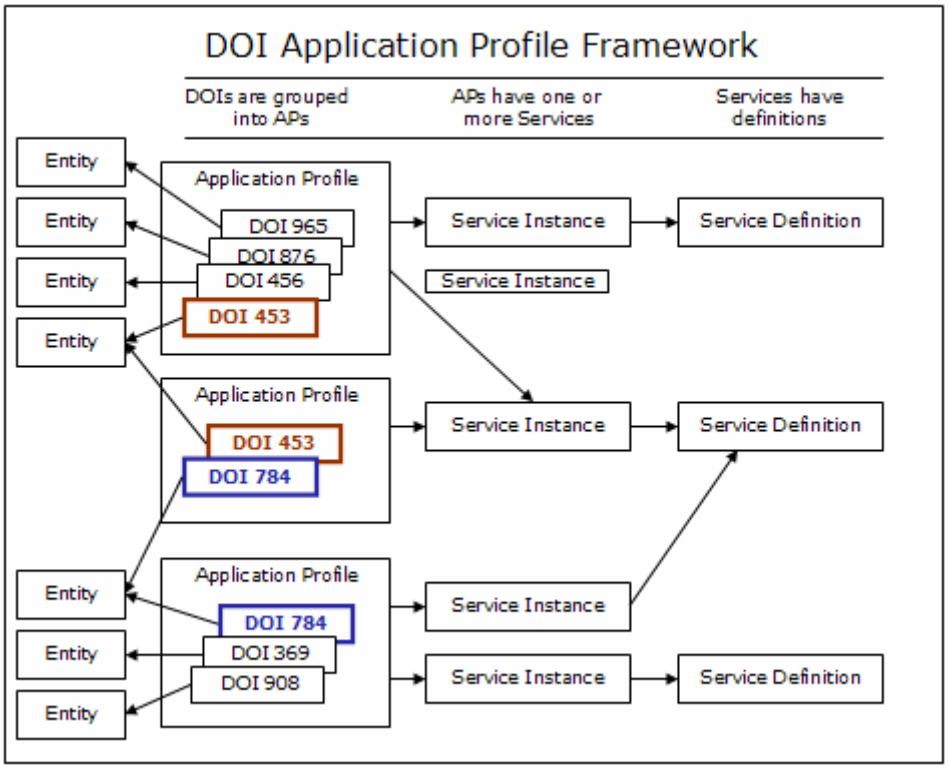


Figure 1: DOI Application Profile Framework

This framework is implemented in the handle system, using DOI names for both APs and Services and linking them together through typed handle values.

APIs (application programming interfaces) have been created that abstract out the details of the handle system implementation and make it possible to administer the structures and to use them in applications while ignoring the details of the implementation. They provide “hooks” down into the handle system without directly manipulating the handle records directly. This not only simplifies the process but would allow the mapping to be altered at a future date without breaking existing applications.

Figure 2 shows the DOI data model from the API perspective. Here the details of the way in which DOI names, APs, and Services are connected is hidden from the application developer who is provided instead with methods for retrieving APs and Services associated with a given DOI name.

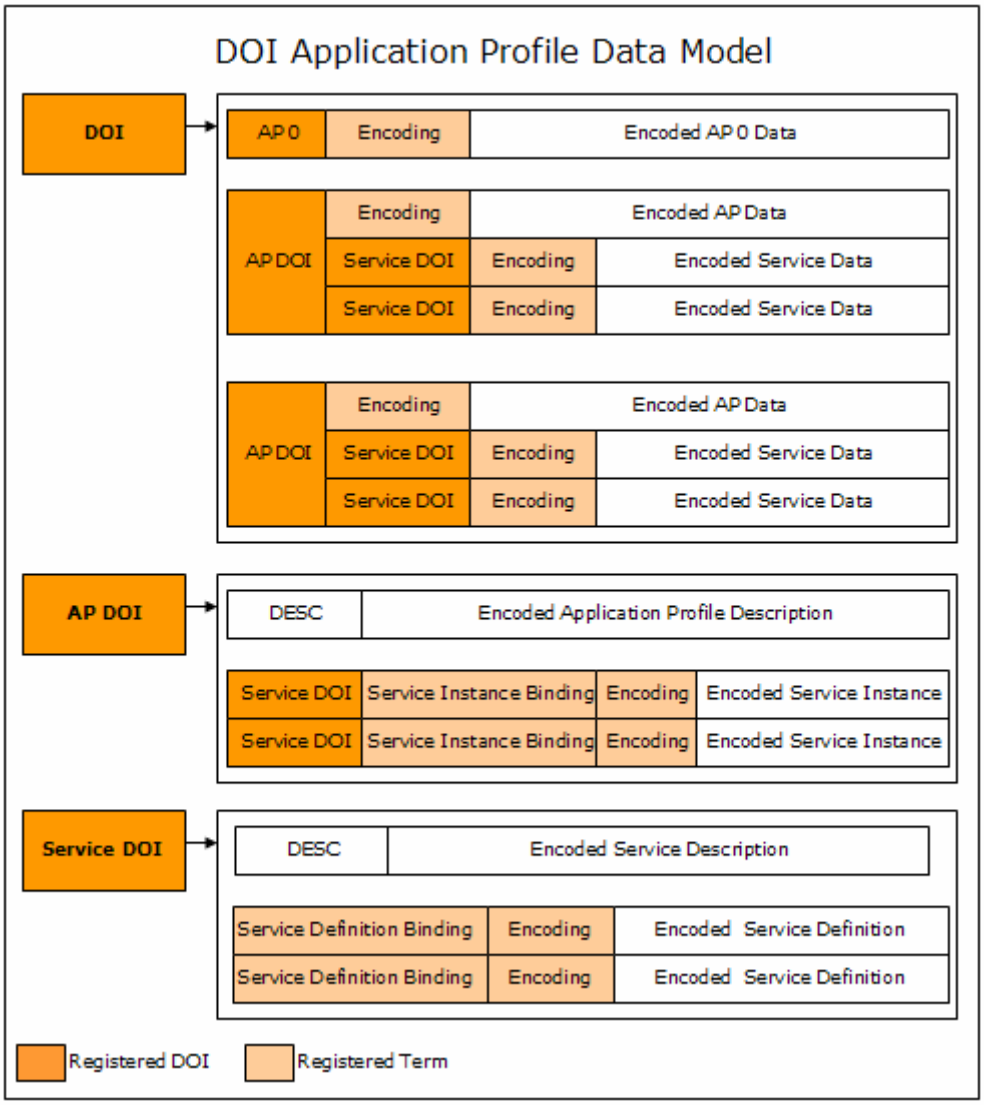


Figure 2: DOI Application Profile Data Model

The approach taken so far has been to develop a Structural API which discusses DOI names, APs and Services for (object-oriented) programmers; and a Service API which ignores APs etc and simply asks "tell me what the services are." The top box in figure 2, for example, shows all of the connections between a single DOI name and its related APs and Services. These could be derived from the handle system by starting with the DOI name and following the links to the associated APs and then following the links from the APs to the associated Services. The APIs take care of this and simply return either the APs and Services or, in the case of the Services API, cutting right to the chase and just returning available Services. Details of the API are currently available to IDF members only while it is still in an advanced prototype stage. It will be made freely available at a later date. [See http://www.cnri.reston.va.us/javadoc/DOIAPI/Main.html].

5.5 Management and administration of APs and Services

5.5.1 Registration of APs and Services

All APs and Services must be registered with, and so implicitly approved by, the IDF. The registration of both APs and Services is restricted to IDF Registration Agencies. This does not preclude third parties from providing DOI System Services or defining APs, but they have to do so through agreements with one or more RAs. Similarly, multiple RAs could collaborate to provide services associated with a given AP. The business model and specific

arrangements are left to the individual RAs to develop as they see fit and the IDF will control only the registration process.

Services, by IDF policy, should be as consistent as possible across all DOI names, e.g., kernel metadata for all DOI names should be obtained in the same manner even though the schema and network locations vary. Part of the reason for enforcing a centralized registration process is to prevent essentially the same service from being offered under multiple names, and so confusing application builders and artificially impeding potential DOI System use.

All current DOI names resolving to a single URL and nothing else are now defined as in the "Zero Application Profile", AP type 0 (10.profile/0). That single URL will be known as the Primary URL and will continue to serve its current purpose of default web location to return for the identified entity.

Any DOI name which is associated with an AP other than 0 will, by policy, be associated with at least the "Base AP", 10.profile/1. The base AP will carry three pieces of information, either directly or through indirection: the RA which registered the DOI name, the Kernel Metadata Declaration (KMD), and the Primary URL. Other than being mandated by policy, 10.profile/1 will function as a normal AP.

One of the services linked to 10.profile/1 will be provision of the Kernel metadata with as much commonality as possible across formats and protocols, such that a core group of clients will be able to retrieve and parse the Kernel metadata across all DOI names.

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 AP (of which the simplest should be the Base AP); entities *may* belong to many APs, so long as the Registrant complies with any rules that may be associated with each AP.

5.5.2 AP and Service Management

While some APs and Services may be relatively small in terms of numbers of affiliated DOI names and number of organizations involved, it is likely that some set of both APs and Services will develop "user communities" that will require governance mechanisms. The RA which registered the AP or Service may alone decide to act as the governance body for the AP, or may delegate this to a third party.

Rules of procedure for the management of the DOI-AP could include issues such as access to and exploitation of metadata, as well the implementation of applications based on the metadata. The procedural rules could also cover the question of who shall be permitted to register a DOI name and to manage the associated data, within an AP.

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

The decision to develop a new AP or Service, versus extending the use of an existing schemes, is based on functional granularity: a new AP or Service should be developed when there is a reason to distinguish it from one already in existence. The IDF will examine each proposed registration with this question in mind. The IDF's primary guideline will be to minimize the number of similar APs and Services, while maximizing the applications of the DOI System that are available to registrants and facilitating and encouraging the development of new applications. The IDF will encourage dialogue between organizations with similar requirements, to explore the potential for convergence among various APs and Services.

Mapping of AP metadata schemas, as planned in the definition of APs, to the indecs Data Dictionary will ensure interoperability between the metadata schemas of different APs, thus allowing the same entity to be registered in more than one AP without metadata conflicts. A DOI name registered in more than one AP will therefore have available to it, potentially, the metadata consisting of the union of the individual APs. The practical busi-

ness implications will however need to be discussed and a role of IDF may be for example to encourage of interaction of AP agencies or users.

An IDF registry will list and summarize APs and Services. This will allow RAs and other interested parties to check existing APs and Services before creating new APs or Services. A simple generated list of registered AP and Service DOI names and their natural language summaries will suffice in the beginning and more complex arrangements will be made as the need arises.

While there is by definition only a single RA for each DOI name, that DOI name can still be associated with services from multiple RAs. The management of this will be left to bilateral agreements not centralized through the IDF. Policy implications also follow from this: for example, if there is any cross charging between RAs for use of "their" DOI names in some service provision, this will be left to bilateral agreements not centralised through IDF.

5.5.3 Ownership of Intellectual Property in DOI APs

Any intellectual property that may exist in a DOI AP, as defined within the context of the DOI System and the iDD, is the property of the IDF. The IDF will license the *use* of those APs and Services to any Registration Agency without charge. However, in recognition of the costs involved in defining an AP and the applications which the AP enables, the IDF may consider a negotiated period of exclusivity for a Registration Agency, before others are licensed to use any particular AP.

There is no compulsion on Registration Agencies to license the use of any services that they may have developed or to license their intellectual property in any context other than that of APs.

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 AP to use in another AP.

5.6 Benefits

DOI names identify different kinds of entities with different attributes and different related services which may vary over time; and so the core DOI name resolution mechanism must be able to show those differences and point to those services, either directly or indirectly, as they appear.

The purpose in creating Application Profiles and then linking each individual DOI name to one or perhaps more of them is to allow users to determine what can be done with a given DOI name. An application is able to recognise that "this is a CrossRef Citation DOI name and so it is reasonable for me to use it in a rights clearance query or to offer some human user the option to do so". It can use all such DOI names in the same way, directly or indirectly through some service which knows how to use that structure.

DOI Application Profiles are a grouping mechanism, and through combination with resolution the AP also becomes a level of indirection for services. So if the Acme Registration Agency registers one million DOI names all of Application Profile 9 and a year later adds one more service to the three services that were available from the start, only the single 10.profile/9 record needs updating, and not the one million DOI names, which are already tagged with AP9.

Similarly, additional services may be made available to existing APs: when a new service is created it will not be necessary to change every DOI name to point to that service; nor will users have to find out about the service; the Application Profile of each DOI name is the key to making that happen: add the new facility (service) to the Application Profile record and all of the millions of DOI names associated with that AP inherit the new facility.

The evolution of this approach will include more precise definitions of both APs and Services via linkage to the iDD. This will greatly aid in the decision of whether to use an existing AP or Service or to create anew and will enable application builders to more easily understand the purpose of a given AP or Service.

6 Policy

This chapter describes the process of policy formulation within the International DOI Foundation (IDF), and summarises current key policies. Some areas of policy are discussed in detail: the data associated with a DOI® name; restricted use of DOI names; exclusivity; persistence; uniqueness; the separation of object, identifier, and metadata; use of DOI names.

6.1 Policy formulation

The DOI® System requires rules for its management to ensure that the system behaves in ways that are predictable and consistent.

The formulation of policy is in many ways more complex than the management and development of the technology. The way in which the DOI System is implemented can have significant impact on the way in which the intellectual-property based businesses that use the DOI System operate in the network environment. Policies with respect to metadata access and exploitation, for example, are designed to have minimum influence on the business models of Registration Agencies.

The detail of IDF policy is continually developed to meet the requirements of implementation. IDF members are fully involved in all aspects of policy formulation and have the opportunity democratically to affect its outcome.

Working group deliberations may result in written documents circulated to members, other 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.

6.2 Current areas of 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.

Policy development includes:

- Policies relating to the establishment and use of DOI® Application Profiles and DOI System services. (Existing policies are included in the relevant chapters on DOI® Data Model (4) and DOI System Applications (5).
- Agreement articulating the relationship of an RA to the IDF.
- The extension of DOI System applications to new sectors and applications and the policies related to social governance in a sector. There is a 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 and DOI System Services concepts allow for considerable autonomy in applications.

6.3 Current policies of the IDF

The following top-level policies have already been agreed by the IDF: refer to relevant chapters for details:

1. **Open resolution:** Once a DOI name is assigned, anyone may resolve that DOI name without charge. At least some information will always be available on resolution (see Chapter 3).

2. **A DOI name can be used to identify any intellectual property entity:** this includes both physical and digital manifestations, performances and abstract works. An entity can be identified at any arbitrary level of granularity (see Chapter 1.6). DOI names may also be assigned to any entity involved in an intellectual property transaction: e.g. parties, licences (see Chapter 1).

3. **Intellectual property focus:** The primary focus of the DOI System is on the management of intellectual property entities, but this does not preclude (for example) issuing a DOI name to an entity that is in the public domain (see Chapter 1.6).

4. **Cost recovery:** 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 (see Chapter 7).

5. **DOI names must be registered:** All DOI names must be registered in the DOI System Directory. Registrants are responsible for the maintenance of state data and metadata relating to DOI names that they have registered (see Chapter 3.7).

6. **Standard syntax:** The syntax of the DOI name follows a standardized syntax (see Chapter 2.2).

7. **Opaque (non-intelligent) syntax:** In use, the DOI name is an opaque string (dumb number); nothing at all can or should be inferred from the number in respect of its use in the DOI System (see Chapter 2.2).

8. **Variety of business models:** 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 (see Chapter 8).

9. **Application Profile:** Each entity registered for a DOI name will be assigned by its Registrant to at least one DOI Application Profile, either with no additional data or with some structured data (see Chapter 4).

10. **Kernel:** An RA must be capable of producing a Kernel Metadata Declaration (the minimum required to permit basic recognition of the entity to which the DOI name is assigned) for each DOI name issued other than the “zero” AP. Controlled values used by RAs in Kernel Metadata Declarations should be registered in the IDF’s data dictionary (the “iDD”) (see Chapter 4).

11. **No ability to search from metadata to a DOI name:** Reverse look-up (from metadata to a DOI name) 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 may offer this service by agreement with their registrants and suppliers on commercial terms, not determined by IDF.

12. **Input and service metadata:** DOI System policy places no restrictions on the form and content of an RA’s input and service metadata declarations, except insofar as input metadata must support the minimum requirements for supporting a Kernel Metadata Declaration (see Chapter 4).

13. **No sale or re-use of DOI System data by IDF:** IDF will not consolidate DOI name 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 name to the declared metadata by any user.

14. **No disclosure of usage:** Usage statistics and information about individual DOI name resolution will not be released by IDF to any party. IDF will only release statistics relating to the aggregate activity of the DOI System. Records of resolution

activity by the DOI System are not made available. However the IDF may make data patterns and registration information available for research in the interests of e.g. Handle System performance analysis and optimization. In such cases, all data will be scrambled to render them un-interpretable and un-resolvable.

15. **Escrow:** DOI name data deposited with a Registration Agency will be held in escrow under contractual terms between the Registration Agency and IDF; the data will be available to the IDF in the event of cessation of the Registration Agency.

16. **Data interchange between RAs:** Metadata exchanged between RAs supporting DOI System services should be exchanged using an agreed DOI System Resource Metadata Declaration ("RMD") using metadata elements and controlled values registered in the IDF's data dictionary (the "iDD") (see Chapter 4).

6.4 Data associated with a DOI name

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

In more sophisticated applications, a DOI name 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 DOI names which have similar characteristics, such as the same metadata schemas and business rules for DOI name assignment. Thus, discovering that a given DOI name is a member of a given AP is a shortcut to knowing what metadata elements can be found for the DOI name, for knowing who is responsible for maintaining the DOI name, and for any other characteristic that is common to the set of DOI names which are of that AP.

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

6.5 Restricted use of the DOI System

An exception to the rules on public information could be some uses of the DOI System, which are not public (either permanently or temporarily). The introduction of Restricted Application Profiles (see earlier chapters) is intended to allow this. Private use of the DOI System 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 DOI names are intended to be used by any discoverer of the DOI System. There may be instances where the benefits of a open, interoperable architecture are not appropriate for a users needs:

1. Where a wholly non-public use is envisaged, the concept of private use of DOI names 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 DOI names 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 but this alone does not confer the advantages of the DOI names structured metadata approach and rules-based governance.

2. Migration from Private to Public identifiers is a likely development which argues in favour of a single underlying system. Some DOI names 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. Hence it would be advantageous if DOI names could be allocated on exactly the same technical basis as described for public DOI names, but with a policy or technical change which allows them to be private (and potentially switched to public) if desired.

Private DOI names 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. Examples:

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

A spectrum of accessibility options could be devised in this way. This definition of "private use" of DOI names equates to "not public use" of DOI names. The concept of "private use" is something that we wish to explore further, to take advantage of the DOI System 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.

6.6 Exclusivity and restrictions

Exclusivity of DOI name 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 any 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 name registration alone, or by being granted artificial exclusivity for a wide range of activities. DOI name registration is "common infrastructure". Investors in RA businesses are naturally looking for the lowest-risk proposition possible and may feel that a grant of exclusivity conferring a monopoly is an attractive option; however, no commonly accepted models for namespace management and resolution services in general are yet established and the IDF is unwilling to cement any business models alone. In the future, globally unambiguous names will be critical, and a ubiquitous infrastructure to "resolve" those names will be essential. IDF will allow service and technology providers to overlap into adjacent domains, and leave interfaces (technical and business) open to allow experimentation with various models to happen and evolution to follow its course.

In order to provide some confidence to RA investors however, applications to become a Registration Agency will be assessed in the context of likely business implications. Where there is an overlap of the expected market or services of existing RA(s) and the applicant RA(s), each will be informed of the potential for overlap or competition and invited to address the problem in such a way as to encourage uptake of the DOI System as a whole whilst ensuring that legitimate business interests are met.

The development of DOI Application Profiles and DOI System Services allows the possibility of dealing with any necessary commercial or legal requirements for restricted access in a more precise way than granting exclusivity:

- Application Profiles may be either public (available to all RAs) or restricted: business rules of an AP may well embody a restriction (e.g. if organization X has an existing legal mandate and responsibility to publish and manage documents on be-

half of a government body on an exclusive basis, and wishes to translate this activity into the digital world, it may well be appropriate to continue by devising an AP that has as one of its business rules the same restriction). The default position, however, will be that APs are publicly declared and available for use by any RA.

- DOI System Services may well be either public, or carry a de facto restriction – for example, if a particular DOI System Service is a look up on a proprietary database, then normal commercial terms for access to that database will be a requirement of use.

6.7 Ensuring persistence

6.7.1 Persistence of the DOI System

Persistence is the consistent availability over time of useful information about a specified entity: ultimately guaranteed by social infrastructure (through *policy*) and assisted by technology such as managed *metadata* and indirection through *resolution* which allows reference to a *first class entity* to be maintained in the face of legitimate, desirable, and unavoidable changes in associated data such as organisation names, domain names, URLs, etc.

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 of identifiers is something we are familiar with in the physical world e.g. ISBNs for out of print books can still be useful. Persistent identification alone is a good enough reason to adopt identifiers such as the DOI System 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.

The DOI System 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 name 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 name, by contrast, specifies an object, not a location. Each DOI name 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 name to the new location and the object can be found once more; any external party accessing the DOI name does not need to know of the change and will be taken transparently to the object. The DOI System is designed to enable registrants 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.

Two key features of DOI names as tools that encourage persistence are:

(1) Indirection – one can change the current attributes without changing the name; this is difficult with URLs which have one attribute (domain name) wired semantically into the identifier;
(2) Separation of name from ownership. DOI names do not base the identifier on the owner, beyond the initial creation. That is, the naming authority (prefix) is used as a convenience for identifier creation but at that point becomes meaningless and

the ownership of individual DOI names can be transferred without regard for the naming authority.

'Persistence' is an imprecise term: we need to talk about persistence in a more fine-grained fashion such as persisting over what conditions, and to what end. DOI names resolve to information (metadata) about the identified object in a manner intended to persist over changes in location, ownership, description methods, and other changeable attributes but not necessarily over changes in the basic utility of the object. If the object ceases to be available, the DOI name may no longer resolve to any information.

6.7.2 Persistence of the IDF

The Foundation was created in 1998; it is a Delaware registered not-for-profit organization, controlled by a Board elected by the members of the Foundation. The by-laws ensure governance in the interests of the stated aims of the DOI System and provide for the members to determine the outcome in the event of winding up. The IDF business model and plan is a self-sustaining cost recovery model. Accounts have been approved by auditors without qualification in each year. The risk of any failure of one application is spread over many companies. In the event of the IDF ceasing to exist the Board would seek to transfer the system to other parties. The IDF has substantial assets in trademarks and the existence of many applications using the specific unique combination of open standard components under the DOI System licenses.

6.7.3 Persistence of the resolution technology (Handle System®)

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

The Handle System is an open standard, so anyone can build/use a handle service; both source code and APIs are public (www.handle.net). It relies of course on the top level Global Handle Registry to be in place somewhere (but so does the Internet Domain Name System assume 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 (these include e.g. Library of Congress, the US Dept of Defense DTIC, the IDF and its constituent members, IDF RAs and their constituents (e.g. CrossRef and publishers), etc.). The Handle System 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 HSAC.

Key Handle infrastructure is placed with a professional hosting company with resources to ensure 24x7 cover, mirroring machines to ensure against power outages etc. Further steps to make the system more persistent from an organizational point of view are under discussion, largely influenced by IDF requirements. Further steps in supporting technical infrastructure through mirror sites in Europe are under discussion.

IDF's agreement with CNRI allows for reversion of all DOI System data, licenses, rights etc to IDF in the unlikely event of CNRI closing or being unable to sustain its activities. IDF's agreements with RAs allow for reversion of one RAs DOI System data to IDF in the event of an RA defaulting.

DOI names are licensed implementations of the Handle System. The specific implementation, the DOI System, is protected by trademarks of the International DOI Foundation and additional specific metadata technologies and policies.

6.7.4 Persistence of the metadata technology

The DOI Data Model is based on the principles outlined in the indecs (interoperability of data in e commerce systems) project and its subsequent development. The fundamental approach of an ontology and dictionary is now published as ISO standard, MPEG-21 RDD, and is freely available for use. The implementation of this standard in an actionable and

maintained dictionary is shared with other users, initially the ONIX metadata set for publisher product information, and again uses open standards, especially XML. The right to use indecs material is held jointly by IDF and EDItEUR under a perpetual assignment agreement which specifically commits the assignees to develop it "consistent with the aims and principles set out in the indecs final report" and in particular to promote its use and make the indecs Analysis freely available and to ensure that a third party does not acquire any exclusive right or license in respect of the indecs Analysis. In the event of failure of one or both, the reversionary rightsholders include International Federation of the Phonographic Industries, rights organisations (CAL, Kopiosto), etc, all of whom have an interest in the open use of the technology. The specific development of indecs into MPEG 21 was done by IDF in collaboration with major organisations such as the Recording Industry Association of America and Motion Picture Association – the MPEG work is available as an open published standard, and the intention of the consortium is to promote and make widely available tools to aid in the use of the dictionary. All the stakeholders in the metadata tools have a shared aim of common and widespread deployment.

6.7.5 Acceptance by users

Since the Handle System is built on open Internet standards, it does not need any formal "acceptance" to be used. As a demonstration: search Google using the terms ["Mona Lisa" Corbis]. The result will (a) show a DOI name as the first hit; and (b) you can then click on that link and use full DOI System functionality. Neither Google nor your browser need to "accept" any DOI System technology. Fundamentally DOI names are simply URIs and therefore any Internet or web technology compliant with standards can use them. If you wish to enhance applications, some tools are available, but these are optional not essential (see http://www.doi.org/tools.html.) More will be added as they become available.

Acceptance by intermediaries is not something that is a necessary step to implementation, and hence the IDF does not need to know which vendors are using DOI names and does not maintain any central list. DOI System policy is that DOI names are "free at the point of consumption"; there is no charge or barrier to resolving a DOI name. The use of DOI names in an information chain is demonstrated in e.g. http://www.crossref.org/. That system has "the DOI names inside", yet it may not be obvious to outsiders that it uses DOI names. Secondary information databases such as IEEE "support" DOI names in that they are simply an entry in the database, as any search on Google under "IEEE and DOI name" will show; etc.

6.7.6 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 name 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 name registrant is free to have the DOI name resolve to a response screen indicating that the identified entity is no longer available. This 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 DOI names, whereby DOI names 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 DOI names. 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 name 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 name, including commercial issues. However this is an issue where we welcome active contributions and suggestions.

6.8 Ensuring uniqueness

The DOI System will not accept duplicate DOI names on registration. As each prefix is unique to a publisher, no two DOI names from different publishers can ever share the same prefix.

As far as is reasonably possible each DOI name should be associated with a unique item of intellectual property. RAs operating with the same or related Application Profiles may agree a method for registration agencies to check if an item of intellectual property already has a DOI name assigned to it. 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.

6.9 Community autonomy

The requirement that the DOI System be extensible across any medium and type of content requires the ability for decentralised application building. For a managed system, this implies subsidiarity: decisions about changes should be taken at the lowest level compatible with maintaining integrity of the system. The consequence for DOI System deployment is that individual communities of interest should be empowered to use the system with a great deal of autonomy. Registration Agencies, based on market models like physical bar codes, effectively hold a “franchise” on the DOI System. 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 name registration and charging fees for participation. Since DOI names are designed to serve a wide range of communities, it would be senseless for the IDF to impose any specific business model, or to analyse a specific community’s detailed needs. However, it would equally be senseless if communities pretended the problems of managing digital information were unique to them and had nothing in common with others. If a community becomes an RA, or endorses an entity as an RA, it is free to set up any service and business model it wishes.

There is a fee for participation in the system to all RAs, based on number of DOI names assigned; nonetheless, the IDF has no say whatsoever in how RAs generate that fee (and so DOI names can be issued free or as part of a fee-bearing service). As an example, CrossRef is a service of PILA (Publishers International Linking Association), itself a non-profit consortium of some 200 publishers – both for-profit and not-for profit. PILA has its own Board, business plan, governance, etc. The only mandatory connection it has with IDF is a formal agreement for the use of the DOI System, through which it obtains licenses to the Handle System and <*indecs*> *Data Dictionary*, DOI System implementations, and the benefits of common technology. PILA uses the DOI System as one part of what CrossRef builds; the rest of the technology, and its use, is entirely up to PILA. More importantly perhaps, the influence is in fact the other way around: rather than IDF exerting control on RAs, RAs exert an influence on IDF. CrossRef is on the IDF Board, and they have representation on the IDF working groups. Ultimately, the RAs will wholly control IDF as a federation. This structure means that every RA has a say in managing the common infrastructure of many applications. The more RAs, the more valuable such collaboration will be.

A new RA is also able to take advantage of existing DOI System work and common infrastructure to save time and money and ensure future interoperability. An RA has far more influence than does an organization which developed its own “island of interoperability” by creating a separate “X-namespace” with its own rules and mechanisms: that would work within the island of X, but have no influence on what others were doing; therefore, such an organization would need gateways to Z, Y, etc.

Communities may also pool resources, developing several RAs yet using one RA as a common back-office, and this is being actively pursued by some DOI System participants at present. To encourage communities to work on relevant applications, the IDF has encouraged the formation of working groups that can focus on specific areas and may build prototypes or construct full applications: for example, CrossRef was inspired by an early IDF working group.

6.10 Separation of metadata, identifier, and entity

6.10.1 Separation of metadata and identifier

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

The DOI name is an opaque string. No definitive information can or should be interpreted from the number in use. In particular, the fact that the DOI name has a prefix issued by a particular organization should not be used to identify the owner of any given intellectual property – the DOI name 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 name 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 name may, if the registrant wishes, be an existing identifier (e.g. an ISBN, or a SICI) which contains some intelligence in that other system. 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 the DOI System): 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 System has used as a basis for its metadata principles.

6.10.2 Separation of identifier from object

In certain digital objects, an identifier such as the DOI name may be part of the binary bit stream, which makes up the object. However the DOI name could be stripped from the bitstream and the intellectual property could be used without the identifier. The DOI System 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 name 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 System is to identify materials for management purposes including legal management. DOI names 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 System.

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

- The Handle System allows ad hoc trusted resolution across the entire distributed system. There is some overhead involved in utilizing this feature.
- The Handle System supports distributed trusted administration. The permissions are at the handle (DOI name) level, not at the server level. Again, this is especially useful for a distributed system. If you have administrative privileges on a given DOI name, 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 System entities by providing a readily interoperable platform and set of policies.

6.11 Use of DOI names 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": the DOI System 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 name assigned to an entity cannot, of itself, certify anything about the copyright status of that entity. DOI names may be *assigned* by third parties; and most certainly DOI names 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 name 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. DOI names 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, the DOI System 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 DOI names 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 name, I need to locate an appropriate source to which I have access rights, or on some other basis (this too could have a DOI name).

- DOI names 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 name, which has a relationship to the original work. The key to describing this relationship between two entities is metadata. The indecs Data Dictionary offers formal mechanisms for expressing these relationships as part of Application Profiles.

- 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 DOI names and their customers. DOI names can of course resolve to other DOI names or to an internal service organisation identifier.

7 The International DOI Foundation

This chapter discusses the organization which governs the DOI® System. Business model and structural principles; the role of the IDF in relation to its Registration Agencies and the underlying technologies of the DOI System; the role of the IDF as the DOI System maintenance agency and development arm; and IDF's relationship to other organizations are described. Details of membership may be found here.

7.1 Origins and status

The International DOI Foundation is an organization established to develop and manage the *DOI System*. The DOI System evolved from a project of the Association of American Publishers in 1996, which joined forces with the International Publishers Association and the International Association of Scientific Technical and Medical Publishers to launch the Foundation in 1998. The International DOI Foundation (IDF) supports the needs of the intellectual property community in the digital environment, by the development and promotion of the DOI System as a common infrastructure for content management. The Foundation is international in its membership and activities.

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, registered on 10 October 1997 – registration number 2807134 8100. The Foundation is controlled by a Board elected by the members of the Foundation. 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. The IDF is also 100% owner of a UK company, IDF (UK) Ltd.

7.2 Business and organizational principles

7.2.1 Cost recovery and its protection

Persistence is a function of organisations not technology. Hence in building a persistent identifier system, we needed to design a model for a persistent organization. The principle concern of a persistent organization is of continuing funding; hence the model selected for a long-term position was a body that is not reliant on external sources, such as funding or membership, but a self-funding system that can be supported in perpetuity from its own resources.

The implementation of the DOI System adds value, but necessarily incurs some resource costs: in data management, infrastructure provision and governance, all of which contribute to persistence. The mechanism chosen to recoup those costs in a self-funding business model, as used by the physical bar code UCC/EAN system, and other proven systems, is a fee for allocation of an identifier but not for its use once issued.

To make such a system work effectively requires protection of the assets within the system (1) from illicit exploitation, and (2) for assured quality control. Illicit exploitation would include someone calling something a DOI® name when it is not part of the official system; this could be damaging to one or both the financial health (avoiding payment of an issuing fee) or the quality of the system (poor quality data). To prevent this requires the availability of legal remedies: specifically, the DOI System relies on copyright and trademark law to protect the DOI System brand and reputation. The DOI System is not a patented system; the IDF has not developed any patent claims on the DOI System and does not rely on patent law for remedy.

The underlying technologies used by the DOI System also have similar considerations. The Handle System is used by IDF under licence from the Corporation for National Research Initiatives, who have certain intellectual property claims to protect the misuse of the Handle System; indecs intellectual property (IP) is assigned to, jointly and solely, IDF and

EDItEUR and made available freely but under stated terms to others (an example being the indecs RDD work contributed to MPEG 21).

7.2.2 The creation of Registration Agencies

When DOI names were launched, no third party Registration Agencies existed. Early users of DOI names 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 than IDF, are the face of the DOI System as far as the end customer is concerned.

Registration Agencies are now being appointed, in a gradual process of migration to this franchise model. The policies and procedures related to DOI name allocation are moving from IDF alone to this network or federation of Registration Agencies.

7.3 DOI System business model

At the outset of the DOI name development, a simple model was introduced whereby a prefix assignment was purchased for a one-off fee. The fee was introduced not to cover actual costs, but to recognize the fact that some charging for DOI names 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 DOI names to be constructed using that prefix, and entitling the registrant to an infinite number of suffixes. The IDF reserve the right to vary this at a future date. There is no limitation placed on the number of DOI name 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.

DOI names allocated using these prefixes purchased directly from IDF were registered without structured metadata: they are now denoted as in “zero Application Profile”. The disadvantage of using the direct prefix purchase 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 now only an option in exceptional circumstances where an appropriate Registration Agency doesn’t yet exist or if the assigning of DOI names is for experimental purposes. All future DOI names will be registered through one of many Registration Agencies, each of which will use one or more defined DOI® Application Profiles, and each of which is empowered to offer 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 DOI names allocated, numbers of DOI names 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 places minimal constraints on the business models offered by RAs, and enters into discussion on practical implementation of any of these.

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 name – 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 name costs”.

RAs may find it beneficial to develop new DOI Application Profiles and services for their customers, or to the same market segment, in order to widen the potential for use and income stream from their DOI System 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.

The IDF encourages the appropriate use of DOI name 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 limits this.

While the migration to more sophisticated business models is under way, and market development is being undertaken by Registration Agencies, the IDF deliberately offers 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.

7.4 DOI System Oorganisation model

7.4.1 Initial model

The IDF organisation was set up on a similar model to the World Wide Web Consortium (W3C) with funding from Members, and no direct funding from the operational DOI System. The member-based IDF subsidized all the operational running of the system. 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 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.

7.4.2 Future model

The IDF aims to migrate to a different sort of organisation as an established standards deployment activity. This is modeled not on W3C but on similar activities such as ISBN, EAN, VISA. The EAN model is perhaps the closest to likely mature DOI System yet identified. In such a model, the operating entities deploying the standard (in the case of the DOI System, Registration Agencies (RAs)) form a federation structure. The Operating Federation is run by the Agencies via an agreed structure and with an appointed managing agent.

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. The economic model is not yet fully established but we are working towards it. The business model of each RA is determined by the RA itself. The RA enters into agreements with its customers (DOI name 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 man-

agement, 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”.

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 or dynamically by agreeing on a pre-set formula (e.g. a function of number of prefixes, number of DOI names, number of resolutions, etc).

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 DOI names 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.

A variation on this model will eventually 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 a proportion of the seats on the controlling board of the Foundation. It is the intention that this percentage be increased approximately in proportion to the percentage of revenue derived from RA operating activities.

7.5 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 Agencies 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.

7.6 The IDF as the DOI System Maintenance Agency

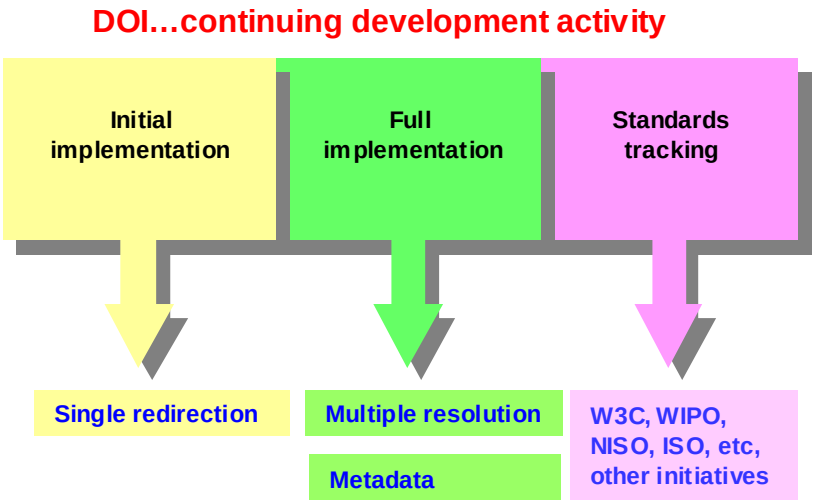
DOI System reliability and predictability can only be delivered in an automated environment, if the DOI System operates in conformance with a flexible and extensible framework of standards; the framework itself can remain unchanging, while specific market-driven developments can be incorporated and managed by extending the framework. For example, new DOI-APs can be readily developed to meet specific community needs, and new services added. There is an increasing community of interest in the DOI System – Registration Agencies, Registrants, users, and the members of IDF. Each of these groups needs to have a voice in the development of DOI System technical and procedural standards, to ensure that they are genuinely market driven.

However, there must ultimately be one organization that arbitrates and decides what should or should not be developed into a standard for the DOI System. The IDF is the Maintenance Agency both for those aspects of the DOI System that are put through external standardization procedures, 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.

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. With respect to internal DOI System standards, the IDF acts as final arbiter.

7.7 The IDF and DOI System development

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



The development of the initial implementation – the resolution of a DOI name to a single URL, or “single redirection” – was the urgent first task of the Foundation.

As the initial implementation was being developed, work was in hand to develop the other two strands of the DOI activity: the full implementation that is currently 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.

The initial implementation of the DOI System was an essential, but limited, first step. Several million DOI names 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. DOI names already registered have the option to migrate into the "Zero Application Profile" or into another appropriate DOI-AP. A DOI name in the Zero AP will have no associated metadata, and will therefore have very limited functionality.

The full implementation offers the necessary infrastructure to provide for the declaration of metadata alongside the DOI name, and 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 market demand. The IDF works 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.

7.7.1 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 to allow participation in a working group at a reduced fee, without a commitment to full membership in IDF. 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 Agencies 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.

7.7.2 TWG (Technical Working Group)

The DOI System Technical Working Group (DOI-TWG) is the technical sub-committee of the RAWG (Registration Agencies Working Group). The group is restricted to IDF members; it comprises representatives of RA members; non-RA IDF members may also participate if they wish. The DOI-TWG is a forum for general technical discussions. Examples include further developments related to infrastructure resilience, multiple resolution, the DOI-AP Framework, operations of the proxy server, distributing the DOI System proxy server and mirroring of DOI System LHS. The DOI-TWG reports to the RAWG, which determines related policy and business issues.

The DOI-TWG was originally a list open to both IDF members and non-members and played an important role in many early DOI System developments. The archive of the original list from January 2002 – July 2004 is available at:
http://www.doi.org/mailman/listinfo/doi-twg/.

DOI-TWG@doi.org is the discussion list for the DOI-TWG and the list requires approval to join. The list home page is http://www.doi.org/mailman/listinfo/doi-twg/ That page has a

link to the list archive. The subscriber list has also been posted to the RAWG web site at http://www.doi.org/rawg-members/TWG_subscribers.html. Please contact Julian Sowa (julian.sowa@vnuinc.co.uk) who is the current chair, if you wish to participate in this working group.

7.7.3 RAWG (Registration Agencies Working Group).

The IDF Board of Directors established the RAWG as a forum for discussion of technical, operational and policy issues of concern to Registration Agencies (RAs). The RAWG is responsible to the Board, which retains the ultimate responsibility for policy decisions.

The RAWG will elect a Chair who is then a sitting member of the board of directors. The RAWG Chair will report to the board and represent RAs.

Each approved RA will have one voting representative. The IDF Managing Agent will be a member of the RAWG. The RAWG Chair will invite other attendees to RAWG meetings as appropriate.

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 is the primary focus for discussions on the development and implementation of the DOI System infrastructure (Application Profiles, DOI® API, DOI® Data Model policy). The goal of the RAWG will be to develop consensus opinions and make recommendations to the IDF board of directors and staff on technical, operational and policy priorities and matters directly involved in IDF governance.

The DOI Technical Working Group (TWG), which is open to all IDF members, will continue to be the main forum for detailed technical discussions. There will be overlap in some instances between the TWG and the RAWG and some discussions will need to be sent to both the TWG and the RAWG.

7.7.4 Application Working Groups

The IDF encourages working groups for 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 DOI names 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 System 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 System 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
- 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.

7.7.5 Governance of Working Groups

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 coordinator, to the IDF Managing Agent 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.

7.7.6 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 System, 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.

The purpose of the IDF is to promote the distribution and use of the DOI System; 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 DOI names" 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. The 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 a profit motive inconsistent with its primary objective or charter.

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.

7.7.7 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, can prevent competitors or potential competitors from being able to profit from disclosure of their own proprietary information by 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.

7.7.8 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 available. Non-Disclosure Agreements are available between IDF and any parties. 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.

7.7.9 Third party development

DOI System 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.

Prospective third party developers are invited 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.

7.8 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 System 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. Of particular note are:

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.

DOI-TWG
Mailing list for the DOI System Technical Working Group.

RAWG
Discussion list for DOI System Registration Agencies and other interested members.

7.9 DOI System and Handle System® development

7.9.1 The Handle System Advisory Committee

IDF participates in the Handle System 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.

7.9.2 Further development of related technology

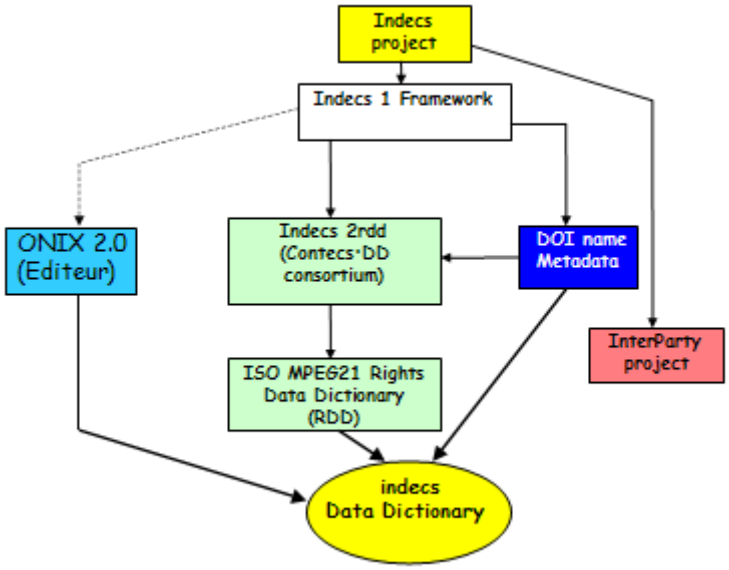
CNRI continues to improve the Handle System. Feedback from users, including the IDF, as well as developments in related network protocols, have resulted in significant evolution of the system. Complete technical specifications and other details can be found at the Handle System web site at http://www.handle.net. 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.

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"*). 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.

7.10 DOI System and indecs development

IDF shares with EDItEUR the role of safeguarding and developing the work resulting from the indecs project, which has since been developed further in a number of ways shown schematically here:



7.10.1 The initial indecs project (Dec 1998-June 2000)

The indecs (interoperability of data in e-commerce systems project) project was established at the end of 1998, with support from the European Commission's 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 initial indecs project developed the indecs metadata framework, a reference model. At the end of the project (June 2000) the intellectual property resulting was entrusted jointly to the International DOI Foundation (developing the DOI System) and EDItEUR (developing ONIX International 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, already collaborating closely.

7.10.2 DOI System-specific development (2001)

The IDF took the indecs analysis and integrated this 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, IDF began to develop a functioning "DOI System namespace" metadata dictionary to support interoperability between DOI Application Profiles, and potentially with other metadata schemes (e.g. Dublin Core, ONIX, MARC, PRISM etc). During the course of 2001, interest in the development of such schemes widened, and an opportunity arose to combine further DOI System work with a new phase of development of the indecs analysis with other organizations.

7.10.3 indecs as input to MPEG (2001-2003)

In April 2001, the IDF funded a feasibility study for a consortium (later known as CONTECS:DD) which would fund the development of standard rights terms to enable the exchange of key information between content industries for e-commerce trading of intellectual property rights. Recognition of the urgent need for a content-industry RDD from activities especially in 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 organizations including the Motion Picture Association of America (MPA), the Recording Industry Association of America (RIAA), the International Federation of the Phonographic Industry (IFPI) and others in a consortium to develop a dictionary which was adopted as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard (see http://www.doi.org/news/020114-DRM.html#current).

7.10.4 The indecs Data Dictionary (2003 onwards)

In 2003, IDF further developed the concepts of indecs by carrying out a proof-of-concept exercise aimed at providing support for all current and future DOI System metadata requirements in collaboration with ONIX. Terms from the DOI System metadata set, ONIX for Books Release 2.0, Crossref metadata and relevant portions of the draft MPEG-21 Rights Data Dictionary (RDD) were mapped together using the "contextual" methodology developed through the indecs framework and the MPEG-21 RDD standardization process. The single operational dictionary will support both DOI Application Profiles and the various ONIX messages. ONIX and the DOI System will continue with independent but now complementing development of their work. This development offered the possibility of re-uniting indecs developments under a common banner with renewed visibility. Since the first indecs work was completed in 2000, there have been several strands of indecs development: the establishment of these terms in a common indecs Data Dictionary re-unites these efforts under one well-established name (see News – http://www.doi.org/announce.html)

7.10.5 InterParty (2002-2003)

InterParty (www.interparty.org) is a project funded under the European Commission's Information Society Technologies Programme (IST), to design and specify a network to support interoperability of party identification (for both natural and corporate names) across different domains. InterParty builds on the work of the indecs project, one of whose deliverables was a specification for a Directory of Parties. InterParty is not proposed as a replacement for existing schemes for the identification of participants in the intellectual property domain (e.g. national library name authority files or systems oriented towards the needs of rights licensing) but as a means of effecting their interoperation. Because of the

close relationship with the interoperability function of the DOI System and the acceptance of common principles, IDF is a participant in the project.

7.11 Alliances and liaisons with other organizations

7.11.1 Alliances and liaisons

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 System 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 DOI System is 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. Many communities are seeing the need to develop an identification scheme for their material. Joining the International DOI Foundation community allows them to instantly leverage years of intelligent policies, standards development and other value-adds, yet not limit in any way autonomy or ability to organize/create own activities. The IDF framework is open enough to leave tremendous room for a community’s own autonomy, activities, and control. Working together rather than dividing efforts is a sensible way forward, and so we welcome discussion with any community.

7.11.2 Formal relationships

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 Members’ 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 System to potential implementations. This list is expanding and we welcome expressions of interest from organizations who wish to establish such a relationship.

ALPSP
Association of Learned and Professional Society Publishers serves, represents and strengthens the community of not-for-profit publishers, demonstrating their essential role in the future of international academic and professional communication.

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 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 in Japan. 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. The DOI System 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 System offers a tool for this purpose and we have advocated its use in this project.

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.

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 and further developed. The indecs trademark and intellectual property is managed jointly by IDF and EDItEUR.

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 the DOI System 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 System development; the DOI® syntax is a NISO standard (ANSI/NISO Z39.84-2000) which was published in May 2000.

OeBF
The Open eBook Forum (OeBF) is a trade and standards organization for the electronic publishing industry – http://www.openebook.org/. Members consist of hardware and software companies publishers, accessibility advocates, authors, users of electronic books, and related organizations whose common goals are to establish specifications and standards and to advance the competitiveness of the electronic publishing industry. The Forum's work fosters the development of applications and products that will benefit creators of content, makers of reading systems and consumers.

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).

7.12 Governance of the DOI System

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 Managing Agent responsible for carrying out policy formulated by the Foundation. The Board is responsible for all aspects of management of the DOI System, 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 Managing Agent, Dr Norman Paskin, represents the IDF in many different forums worldwide and is responsible for the implementation of Board policies and management of all aspects of the affairs of the IDF.

7.13 Membership of the International DOI Foundation

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. IDF has a membership of committed companies and organizations that participate in the development of the system and its applications and welcome 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 hence the high-quality operation of the DOI System, which will be integral to furthering the interests of the Foundation's members. 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. For a list of current members see http://www.doi.org/idf-member-list.html.

7.13.1 Benefits of membership

Participating in an international effort to standardize Internet-related technology is perhaps one of the most important steps that 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 other forum does what the International DOI Foundation does.

Why collaborate? To inter-operate with others outside your business; supply chain, customers, and competitors that are adopting the key standards designed for interoperability. Collaborating through the International DOI Foundation will reduce costs and prevent mistakes and dead-ends when trying to advance as a single company. In our working groups you have the opportunity to take account of intelligent reviews of others' activities. 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.

On joining the IDF, a free DOI name prefix will be available for experimental purposes should you require it. Please contact the IDF at contact@doi.org to request this. The number of DOI names allocated with this experimental prefix will be subject to review and may be limited at the discretion of the IDF. If further prefixes are required for non-experimental purposes, you should work with one of our Registration Agencies. Registration Agencies (RAs) are established to provide services on behalf of specific user communities. For further information on RAs, please see Chapter 8 and the Registration Agencies page on the DOI.ORG website (http://www.doi.org/).

7.13.2 Classes of membership

Members may be either for-profit or not-for-profit organizations. Membership is normally not open to individuals, but the Board reserves the right to allow this in exceptional cases.

There are four classes of membership: General, Charter, Registration Agency, and Affiliate. Charter, General and RA Members are entitled to vote in annual IDF elections within their own category of membership (Affiliate Members do not have voting rights).

All IDF members can participate in IDF working groups however Affiliate membership allows participation in one Working Group only (i.e. participation in two groups requires two Affiliate fees). All IDF members receive material distributed to IDF members only, either from IDF or other parties. They also may attend members meetings, receive information and advice from IDF staff and contribute to the direction and management of the DOI System.

- **General Membership** is offered to any organizations with an interest in electronic publishing and its related enabling technologies, who wish to support the development of the DOI System but are not active Registration Agencies.

- **Charter Membership** is only offered to organizations whose main activities are in the creation or production and dissemination of intellectual property, who wish to support the development of the DOI System but are not active Registration Agencies. 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. The primary (and minimum) role of Registration Agencies is to provide services to Registrants – allocating prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data.

- **Affiliate Membership** is a paid membership without voting rights or other full membership rights in the IDF, which is designed to allow 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).

The benefits of membership participation flow downward to subsidiaries of members as follows:

- For organisations which themselves have member organisations, 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).
- 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 co-ordinated 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.

A membership application can be found at http://www.doi.org/membership/membership-form.html

7.13.3 Membership fees

Membership fees fall due annually on the anniversary of joining, – US members will be invoiced their member fees in US$ and non-US members will be invoiced their member fees in GBP£. The pound sterling value will be calculated each year based on the US exchange

rate set at the start of each year, for 2006 this is $1.86:£1. Eligibility for Membership may be re-evaluated at each point the Member Agreement falls due for renewal.

The current annual fee for General Members is $US 35,000 (GBP£18,817).

The current annual fee for Charter Members is $US 70,000 (GBP£37,634), reduced to $40,000 (GBP£21,505) if the Charter Member is also a lender to the foundation.

The current annual fee for Members may be reduced at the sole discretion of the Board, subject to a minimum fee of $US 11,500 (GBP£6,183) 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 current annual fee for Registration Agency Members is $US 35,000 (GBP£18,817).

The current annual fee for Affiliate Members is $US 5,000 (GBP£2,688) per working group.

7.14 Working with user communities

Joining the International DOI Foundation community allows you to instantly leverage years of intelligent policies, standards development and other value-adds, yet not limit in any way your own autonomy or ability to organise/create your own activities. Our framework is open enough to leave tremendous room for a community's own autonomy, activities, and control. Working together rather than dividing efforts is a sensible way forward. We welcome discussion with any community.

Each DOI System user organisations can do whatever it wishes. The DOI System allows for DOI names to be assigned by Registration Agencies (RAs); each RA is autonomous in its business model – the usual analogy is that the RAs collectively set the rules of the road re resolution etc but do not specify the route to be taken, the vehicle to be driven on the road, or the date, time or purpose of the journey of each RA. There's an equal, small, fee for participation in the system to all users; we have no say whatsoever in how they generate that fee. Rather than IDF exerting control on RAs, RAs exert an influence on IDF: RAs may have seats on the Board of the IDF (whereas IDF is not on their board) and representation on the working groups of the IDF. Ultimately the RAs will wholly control IDF as a federation. This structure means that every RA has a say in managing the common infrastructure of many applications; which is our aim. The more RAs, the more valuable such collaboration will be.

If a community develops an RA, or endorses a separate entity as an RA, it is free to set up any service and business model it wishes. It will also be able to take advantage of the existing DOI System work and common infrastructure to save time and money and ensure future interoperability. Finally, it would have far more influence than if it developed its own "island of interoperability" by developing a separate scheme which would require gateways to other systems.

8 Registration Agencies

This chapter explains how obtain a DOI® name prefix and select a Registration Agency (RA), the role of Registration Agencies) in the DOI® System, their relationship and obligations with the IDF, and the business and technical models that they can adopt to implement the DOI System.

8.1 How to obtain a DOI Prefix

To obtain a DOI name prefix, apply to a DOI System Registration Agency.

8.1.1 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.

For a list of our current Registration Agencies and their area of coverage, please see the DOI® website at http://www.doi.org/registration_agencies.html

8.1.2 Registering a DOI® name with the IDF

The IDF itself registers DOI names for internal identification of Registration Agencies, Application profiles, etc. The IDF may also assign DOI prefixes directly: however as DOI names are used in the context of applications, we actively encourage users to assign DOI names for a specific application, through a Registration Agency, or to develop their own RA (with others if needed) if they feel their requirements aren't met by existing RAs. RAs are able to offer support for metadata allocation, social infrastructure and other support. DOI prefixes direct from IDF are restricted for experimental rather than production use.

In order to obtain a DOI prefix for experimental use only, contact the IDF (contact@doi.org). Any prefixes that may be issued directly by the IDF may be charged at $1,000 per prefix and will be issued purely at the discretion of the IDF.

8.2 The role of Registration Agencies

8.2.1 Overview

At the outset of the DOI System 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 DOI names would be the intention. IDF used a simple initial economic model: a charge of $1000 for allocation of a prefix (a one-off charge though with the right to vary this at a future date) allowing unlimited number of DOI names to be constructed using that prefix. It was recognized at the outset that this fee structure was a starting point but would be insufficiently flexible for the long term.

DOI names allocated using these prefixes purchased directly from IDF are registered without structured metadata: they are now defined as being in the zero Application Profile. The disadvantage of using direct prefix purchase is that IDF cannot offer the level of metadata support and social infrastructure support of the type which can be given by a Registration Agency. Prefixes obtained directly from IDF may however be useful if you wish to experi-

ment or consider developing your own applications. Prefixes will now only be issued through this direct route at the discretion of the IDF Managing Agent.

We now use a wide variety of existing or potential business models, using third party Registration Agencies, in recognition of the fact that such a simple model is not a "one size fits all" solution. All future DOI names will be registered through one of many Registration Agencies, each of which will use one or more defined DOI® Application Profiles 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 DOI names allocated, numbers of DOI names 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.

The primary (and minimum) role of Registration Agencies is to provide services to Registrants – allocating DOI prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data. This is expected to encompass quality assurance measures, 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. All Registration Agencies will be expected to support registration of at least one DOI® AP (as well as the Base AP). 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. Registration Agencies are expected actively to promote the widespread adoption of the DOI System, and to co-operate with the IDF in the development of the DOI System as a whole.

8.2.2 Development of Registration Agency role

The IDF is moving to full deployment of the DOI System 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.

8.2.3 RA technical requirements

- Assign prefixes to new registrants in accordance with IDF standard terms; liaise with the resolution system provider to register prefixes in system directories.
- Ensure that DOI names 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

8.2.4 RA information management requirements

- Ensure that appropriate minimal supporting metadata for each DOI name (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 name 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.

8.2.5 RA general requirements

- Implement mechanisms for quality control of DOI name 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.

8.2.6 Summary of functions

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 DOI names 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 System 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 name resolution service
- ensure resolution service integrity and performance
- maintain Data Type Registries (fields within handle specifying different multiple resolution results)
- provide DOI System-pecific 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 System (e.g. use of Data Model)
- role in governing body (operating federation)

8.3 Business models for Registration Agencies

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

- Direct charging based on prefix allocation, numbers of DOI names allocated, numbers of DOI names 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.

Our aim is to encourage the appropriate use of DOI name 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 prefixes as part of their overall package of services to Registrants. The charge and process for obtaining a 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 Agencies and these are detailed above.

Registration Agencies *may* choose to provide other DOI System-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. Registration Agencies may (but are not obliged to) establish their own local handle service.

8.4 Criteria for becoming a 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 prefixes to Registrants. Registrants can be any individual or organization that wishes to uniquely identify intellectual property entities using the DOI System.

Other than in exceptional circumstances, candidates for RA status will be expected to first become a Member of the IDF in a non-RA category. This will enable the organisation to construct a formal application with the benefit of experience and advice from the IDF, its technical partners, other RAs, the IDF Technical Working Group interactions, and the RA Working Group (on which only RAs are represented but whose conclusions are made known to all members) and have access to the member-only information provided by IDF.

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 DOI names 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 System 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 System 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 name 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 the Handle System and Data Model 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 Managing Agent of the IDF, the applicant will provide a preliminary assessment of the issues they will face in adopting Handle System technology. Specifically some thought should be given to the potential scale of the implementation, reflecting the anticipated volume requirements of DOI name 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.

8.5 Formalizing the relationship between an RA and the IDF

8.5.1 Current procedures

If an organization wishes to be involved in determining RA status or applying to be an RA, an initial commitment must be made 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 US$35,000 (GBP£18,817) and this amount must be paid in full before a new Registration Agency begins to operate. US members will be invoiced their member fees in US$ and non-US members will be invoiced their member fees in GBP£. The pound sterling value will be calculated each year based on the US exchange rate set at the start of each year, for 2005 this is $1.86:£1.

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, contact@doi.org). The letter of intent also commits the Registration Agency to developing the document on Terms and Conditions for DOI Registration Agencies into a formal IDF/RA Agreement.

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.

8.5.2 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. 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.

2. A definitive Registration Agency Agreement implementing and expanding upon the terms of existing letters of intent and incorporating IDF's policies is evolving. Any individual rights granted to a particular RA will be reflected in the definitive Registration Agency Agreement executed with that 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.

8.6 Registration Agencies and IDF governance and policy

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 DOI names.

Each Registration Agency will offer services to 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 registrants and the users of the system.

Each Registration Agency will provide its community with a mutually acceptable framework for discussing matters of shared concern and interest. Registrants will be contractually related to a Registration Agency. So long as a 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.

Registration Agencies are expected actively to promote the widespread adoption of the DOI System, 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 intended 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.

8.7 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 "DOI" trademark 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" trademark. 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 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.

8.8 Fee structure for Registration Agencies

A fee is paid by RAs to the IDF in recognition of their participation in, and their ability to build a business using, the DOI System. RAs are free to establish their own business model and fee structure with their own customers.

The 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.

The fee structure for a Registration Agency is:

- The Registration Agency must be a current member of the IDF, the annual membership fee for which is US$35,000 (GBP£18,817). There will be no discretionary reductions in the membership fee for any members who subscribe within the Registration Agency category. US members will be invoiced their member fees in US$ and non-US members will be invoiced their member fees in GBP£. The pound sterling value will be calculated each year based on the US exchange rate set at the start of each year, for 2005 this is $1.86:£1.

- A franchise fee will be charged. The franchise fee will be calculated as US 4 cents per DOI name registered, with a minimum and maximum as noted below. Accounting will be made on 1st January and 1st July each year. Each Registration Agency will guarantee a *minimum* payment of currently $20,000 each year irrespective of the number of DOI names allocated, this is split over the January and July invoicing periods.

The *minimum* element of the "per DOI name" fee will be invoiced in advance of each six-month period. When an RA registers more than the minimum number of DOI names in the preceding six-month period, the difference between the minimum already previously paid will be invoiced.

The franchise fee will be renegotiated if an RA registers more than 10 million DOI names during any six-month period.

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

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

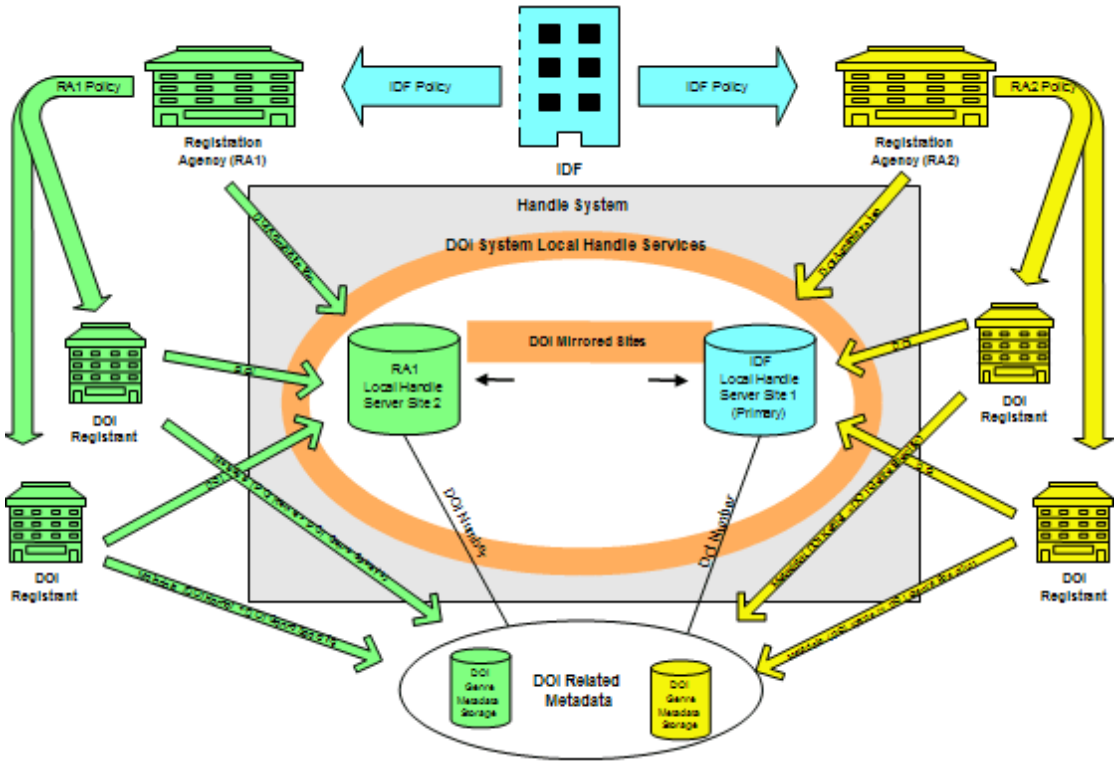
At the beginning of each year, the number of DOI names previously registered will be calculated to the end of the previous year. The annual maintenance fee will be calculated based on the above levels and will be invoiced in two parts, in January and July of that year along with the relevant franchise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

The RA membership fees are invoiced directly by the IDF. US RA members will be invoiced their member fees in US$ and non-US RA members will be invoiced their member fees in GBP£. The pound sterling value will be calculated each year based on the US exchange rate set at the start of each year.

The franchise and maintenance are invoiced by Corporation for National Research Initiatives (CNRI), as an agent of the IDF, and will be invoiced in US$ for all RA members.

8.9 Operational and technical requirements for Registration Agencies

This diagram 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 name Registrants. For the purpose of illustrating different possible scenarios, RA1 is operating and hosting a mirrored DOI System Local Handle Service, whilst RA2 does not host any Handle System infrastructure. All DOI System Local Handle Services are automatically synchronized by the Handle System.

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 names to the Handle System along with the DOI name resolution information.
- At the same time, the Registrant submits DOI System 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 name provides the key to linking the DOI System related metadata with the registered DOI name.

An authorized Registration Agency issues prefixes to registrants and requests the resolution system provider to register such new prefixes in the Handle System. 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 Service to mirror the services provided by the Global Handle Registry® as illustrated by RA1 in the above diagram.

Currently, CNRI implement and maintain the Default DOI System Local Handle Service 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 Services, should they wish. 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.

8.10 Registration Agency Local Handle Service (LHS) Operation

IDF operates a Central DOI® Directory as a master directory of all assigned DOI names, 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 DOI names 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.

8.10.1 Advantages of using a Local Handle Service at an RA

- 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 DOI names, 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 DOI names, increased amount of information per DOI name or increased frequency of change of DOI name would necessitate some corresponding increase in mirror resources).

8.10.2 DOI® Handle Service configuration

Currently, the IDF manages one local handle service known as the DOI® Service or DOIDirectory. This Service consists of multiple servers and contains the DOI names 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 name 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 DOI names that start with, for example, 10.1234/ to necessarily be co-located on the same handle service as DOI names, which start with 10.1234.1/. It is, however, important for all DOI names that start with any single prefix to all be co-located on the same handle service. While it would be theoretically possible to have DOI names 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 name in hand. There is no limit to the number of prefixes for which a single handle service may be responsible.

In the case of a prefix holder needing to transfer DOI names 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 DOI names 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 DOI names 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 name 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 name to a new DOI name (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.

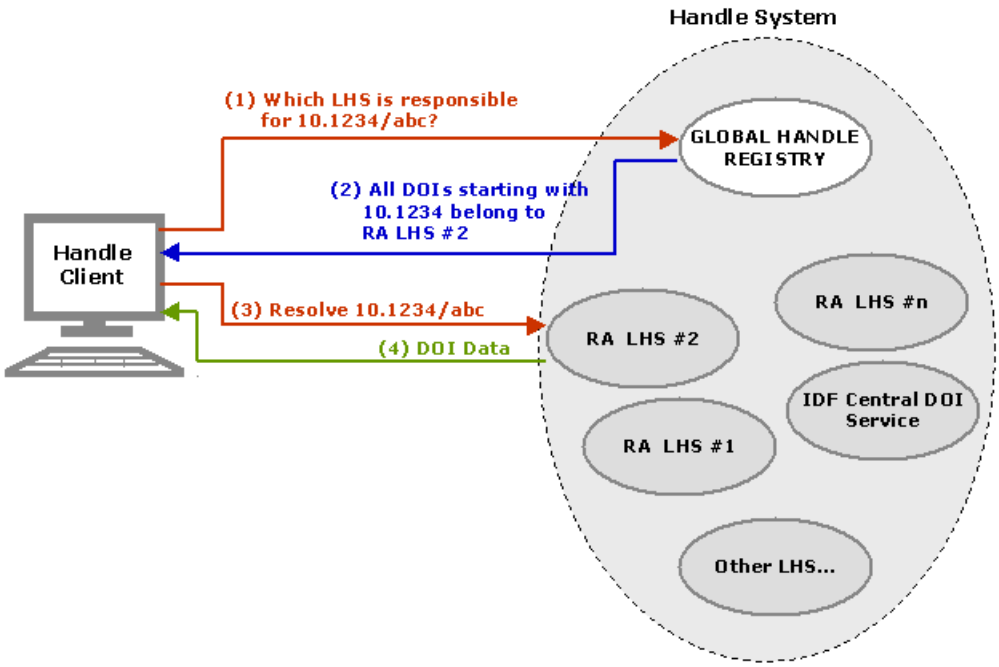


Figure 1

8.10.3 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 DOI names. 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.

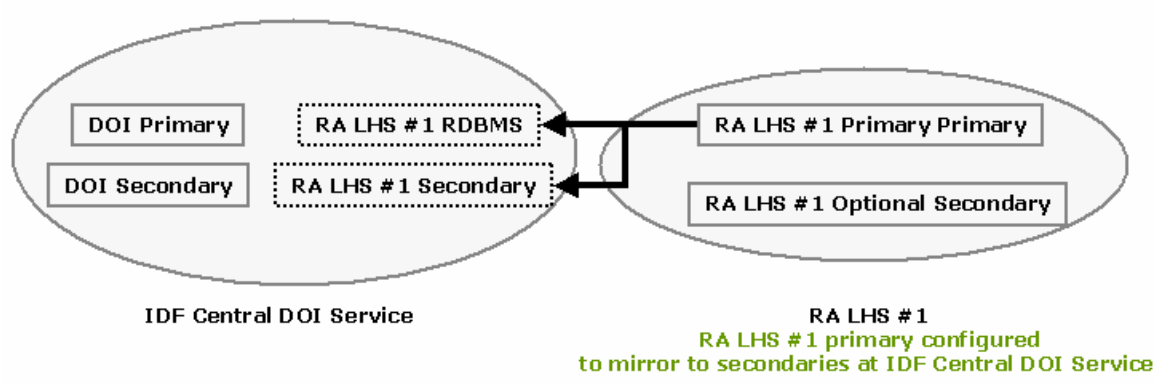


Figure 2

8.10.4 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 DOI names will require the use of an admin DOI name. Each admin DOI name 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.

Appendix 1 Excerpt from ANSI/NISO Z39.84-2005 Syntax for the Digital Object Identifier

This excerpt from ANSI/NISO Z39.84-2005 Syntax for the Digital Object Identifier is reprinted here with permission of NISO Press. Some text has been omitted from the standard as originally published. Since the publication of this standard, some URLs have been updated in this version for accuracy. The full standard can be downloaded as a PDF file for no charge from the NISO website www.niso.org by clicking on the NISO Press icon. DOI® and DOI.ORG® are registered trademarks and the DOI> logo is a trademark of The International DOI Foundation.

ANSI/NISO Z39.84-2005 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
The Unicode Consortium. *The Unicode Standard*, Version 4.0.1, defined by: The Unicode Standard, Version 4.0 (Reading, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1), as amended by Unicode 4.0.1 (http://www.unicode.org/versions/Unicode4.0.1/).

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 a Registration Agency to a Registrant for use 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.

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 practical limit on the length of a DOI string, or any of its components (the Handle system allows strings of up to 4 GB; under UTF-8 encoding each ASCII character takes one byte, hence in ASCII encoding a DOI may be approx 4 billion characters).

Characters ‘a’ - ‘z’ and ‘A’ - ‘Z’ in the DOI string are case insensitive (e.g. 10.123/ABC is identical to 10.123/AbC). These characters in the DOI string are converted to upper case upon registration and resolution. If a DOI were registered as 10.123/ABC, then 10.123/abc would resolve it and a later attempt to register 10.123/AbC would be rejected with an error message stating that the DOI was already in existence. Comparison of two DOIs (to decide if they match or not) should be done by first converting all characters ‘a’ - ‘z’ in DOI strings to upper case, followed by octet-by-octet comparison of the entire DOI string..

4.1 DOI Character Set

Legal characters are the legal graphic characters of Unicode. This specifically excludes the control character ranges 0x00-0x1F and 0x80-0x9F, which are therefore not valid characters for DOI strings, and will never be present in DOI conformant systems. 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. The Directory Code is numeric; currently the only valid value is <DIR>=10.

<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 the DOI Character Set as defined above.

<DIR> and <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 the DOI Character Set as defined above, with the exception that the Suffix cannot start with */ where * is any single character. This is reserved for future use. The DSS is case insensitive.

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 *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only.)

This appendix provides information on aspects of the DOI system syntax implementation 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

<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://www.ietf.org/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 Unicode 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 *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. 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

Dr Norman Paskin
Director
International DOI Foundation
5, Linkside Avenue

Oxford
OX2 8HY
UK
Tel: (+44) 1865 559070

APPENDIX C Examples of Digital Object Identifiers

(This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only, and may include editorial updates and corrections.)

DOI registrants can use a variety of strings for the DSS including private identifiers and existing standards such as SICI (Serial Item and Contribution Identifier). The syntax of the identifier numbering scheme is such that any existing identifier syntax string can be expressed in a form suitable for use with the DOI system.

The following are examples of Digital Object Identifiers:

DOI (incorporating a SICI) from an article in the Journal of the American Society for Information Science, published by John Wiley & Sons:

10.1002/(SICI)1097-4571(199806)49:8<693::AID-ASI4>3.0.CO:2-0

DOI for an article from JAMA, the Journal of the American Medical Association:

10.1001/PUBS.JAMA(278)3,JOC7055-ABSY:

DOI for the article "ABO Blood Group System" from Encyclopedia of Immunology Online, 2nd edition, published by Academic Press:

10.1006/rwei.1999.0001

APPENDIX D Related Standards and References

(This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only, and may include editorial updates and corrections.)

The standard cited in Section 2 is required for the construction of the DOI syntax. This appendix includes references to other standards and citations that may be useful with DOIs or which provide additional information on the DOI.

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

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.

DOI Handbook: DOI 10.1000/182, http://www.doi.org/hb.html

DOI factsheets (DOI and Handle; DOI and Numbering Schemes; DOI and Data Dictionaries; DOI and Internet Identifier Specifications; DOI Applications; Value added by the DOI System: http://www.doi.org/factsheets.html

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

Sun, Sam; Lannom, Larry; Boesch, Brian. "Handle system Overview". RFC 3650, November 2003. http://www.handle.net/rfc/rfc3650.html

Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle system Namespace and Service Definition". RFC 3651, November 2003. http://www.handle.net/rfc/rfc3651.html

Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". RFC 3652, November 2003. http://www.handle.net/rfc/rfc3652.html

UTF-*, A Transform Format for Unicode and ISO 10646", RFC 2044, October 1996, Yergeau, Francois - http://www.ietf.org/rfc/rfc2044.txt

---

APPENDIX E Application Issues

(This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84-2005. 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. The default encoding of Unicode is that each character consists of 16 bits (2 octets). UTF-8 is a variation of the Unicode 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 example, the Japanese word “nihongo” is written as:

The Unicode sequence representing the Han characters for “nihongo” is: 65E5 672C 8A9E These may be encoded in UTF-8 as follows: E6 97 A5 E6 9C AC E8 AA 9E 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 Web browser technology requires additional functionality to allow the browser to make full use of DOIs: additional browser features are necessary. It is anticipated that features supporting resolution will commonly be built into browsers in the future.

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.

Alternatively, without the need to extend the Web browsers’ capability, DOIs may be structured to use the default public DOI proxy server (http://dx.doi.org). The resolution of the DOI in this case depends on the use of URL syntax. For example, “doi:10.123/456” would be written as http://dx.doi.org/10.123/456.

DOIs are also primarily used in HTML pages. The DOI 10.1006/rwei.1999”.0001as 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 The Handle System®

This appendix provides an overview of CNRI's Handle System® which is used as the resolution component of the DOI® System.

For more detailed information, visit the Handle Web site at http://www.handle.net.

A2.1 Handle System overview

A2.1.1 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.

A2.1.2 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.ORG® website (http://www.doi.org/) also provide 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.

A2.1.3 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.

A2.2 Handle Syntax

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

<Handle> ::= <Handle Prefix> "/"<Handle Suffix>

For example, "10.1045/january03-paskin" is a handle for an article published in ***D-Lib Magazine***. It is defined under the prefix (naming authority) "10.1045", and its suffix (local name) is "january03-paskin".

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 among 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.
By default, handles are case sensitive. However, any handle service, including the global service, may define its namespace such that all ASCII characters within any handle are case insensitive.

The handle namespace can be considered as a superset of many local namespaces, with each local namespace having its own unique prefix. The prefix identifies the administrative unit of creation, although not necessarily continuing administration, of the associated handles. Each prefix is guaranteed to be globally unique within the Handle System. Any existing local namespace can join the global handle namespace by obtaining a unique prefix, with the resulting handles being a combination of prefix and local name as shown above.

Each prefix may have "sub" or derived prefixes. For example, once the prefix 12345 has been created, 12345.1 can be created. Handle 12345.1 is therefore defined under prefix 12345. The syntax can be represented as "string.substring".

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

A2.3 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 which make up each service, and there are no limits on the number of servers which 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.

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.

A2.4 Handle System scalability

Scalability was a critical design criteria for the Handle System. The problem can be divided into storage and performance. That is, is there some limit to the number of identifiers (handles) that can be added? And, does performance go down, or do some functions simply break with increased numbers of identifiers, 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 which the basic Handle System design does not and cannot address.

A2.4.1 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 1 illustrates two possible configurations.

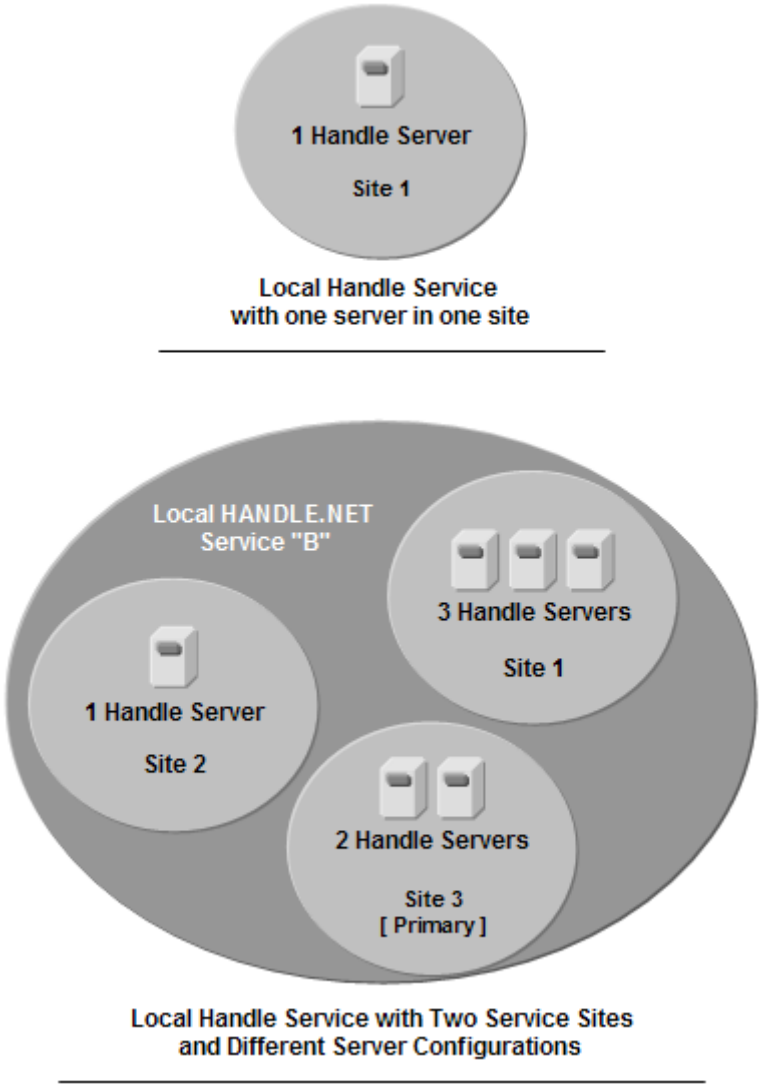


Figure 1 - Example Handle Site Configurations

Identifiers 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 which constitute the Handle System, there are no design limits on the number of sites which make up each service, and there are no limits on the number of servers which 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 identifiers, each site may allocate that set of identifiers across a different number of servers. Thus increased numbers of identifiers 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 ser-

vice can also have as many sites with as many servers as needed. The result is that the number of identifiers that can be accommodated in the current system is limited only by the number of computers available.

A2.4.2 Performance

Constant performance across increasing numbers of identifiers 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 identifiers 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 identifier is located, regardless of the number of identifiers or the number of servers. Each server within a site is responsible for a subset of identifiers managed by that site. Given a specific identifier and knowledge of the service responsible for that identifier, a handle client selects a site within that service and can perform a single computation on the identifier to determine which server within the site contains the identifier. 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 identifiers belong to which servers. The computation is independent of the number of servers and identifiers, and it will not take a client any longer to locate and query the correct server for an identifier within a service that contains billions of identifiers and hundreds of servers, than for a service that contains only millions of identifiers and only a few servers.

The connection between a given identifier and the responsible handle service is determined by prefix. Prefix 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 identifiers 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 handle 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.

Handle clients may optimize performance across parallel service sites and, given a choice of multiple sites, will largely ignore sites which are slow or completely unresponsive, either because of server problems or because of network problems. Any given handle service can thus be made more robust both in terms of performance and reliability, through the addition of servers and collections of 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.

A2.5 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.
Currently available client side software includes:

- **Client Library (ver. 6) -- JAVA™ Version**, a library of Java classes which understands the handle protocol and would form the foundation for Java-based custom client software development.

- **Client Library (ver. 5) -- C Version**, a library of C functions which understands the handle protocol and would form the foundation for custom client software development.

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:

- **HANDLE.NET 6.2** , 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
- **Proxy Servlet (ver. 2) -- Java™ Version**, proxy servlet code for developers who want to set up their own proxy server for handle resolution.

For information on related tools developed for the DOI® System, see "Tools" on the DOI.ORG web site.

A2.6 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.

A 2.7 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://www.ietf.org/rfc/rfc1034.txt
[5] P. Mockapetris, "Domain Names - Implementation and Specification", RFC1035, November 1987.
http://www.ietf.org/rfc/rfc1035.txt
[6] Berners-Lee, T., Masinter, L., McCahill, M., et al., "Uniform Resource Locators (URL)", RFC1738, December 1994. http://www.ietf.org/rfc/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://www.ietf.org/rfc/rfc2044.txt

Appendix 3 DOI® Data Model and sector applications

This appendix comments on some sectors which are using DOI® names in conjunction with existing metadata schemes or where existing schemes are likely to be of particular relevance. Please refer to a relevant Registration Agency for further information.

A3.1 DOI Data Model and text applications

The ONIX (Online Information Exchange) metadata specification (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® System 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/index.cfm?flashplugin=1&fuseaction=scormabt), and potentially PRISM (www.prismstandard.org) etc. EDItEUR and the International DOI Foundation (IDF) are collaborating to ensure that users of ONIX and the DOI System will be able to easily achieve interoperability.

EDItEUR is committed to managing ONIX in such a way that any DOI System 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 DOI names will be consistent wherever they overlap. The International DOI Foundation is committed to ensuring that ONIX specifications can be easily used in developing DOI System Applications. The joint intention is to produce a combined common data dictionary, the indecs Data Dictionary, for use by both ONIX and IDF.

The International DOI Foundation (IDF) and EDItEUR (the International Group for electronic commerce in the book and serial sectors) announce their intention of continuing the work announced last year* towards ONIX and DOI Data Model harmonisation, through the development of a common data dictionary.

IDF has carried out a proof-of-concept exercise aimed at providing support for all current and future DOI Data Model requirements through use of this dictionary. Terms from the DOI Data Model, ONIX for Books Release 2.0, CrossRef DOI System applications and relevant portions of the draft MPEG21 Rights Data Dictionary (RDD) have been mapped together using the “contextual” methodology developed through the original indecs framework and subsequent work including the MPEG21 Rights Data Dictionary standardization process**. The ONIX mappings are now under review, completion and approval by EDItEUR. On approval, an essential first step will have been taken towards the establishment of a single operational dictionary to support both DOI® Application Profiles and the various ONIX messages.

The single operational dictionary of ONIX and IDF will be known as the indecs Data Dictionary in recognition of the fact that this work is fundamentally grounded in the work of the indecs project (www.indecs.org) and subsequent developments of the principles and framework of indecs through activities such as the ISO MPEG-21 Rights Data Dictionary development work co-funded by IDF. The indecs Data Dictionary forms a core component of the DOI Data Model and is discussed in detail elsewhere in this Handbook.

A common dictionary, and processes to support its maintenance and ongoing development, allow:

- DOI Registration Agency (RA) and ONIX Terms to be part of a single dictionary.
- Each Term in the dictionary to have a unique XML tag and URI (DOI name), which could be explicitly cross-referenced in any DTD or Schema for an RA or ONIX format in which it appears. The Term may have a different name and written definition in each different scheme, but as long as it is recognised as having the same meaning it could be linked to a common dictionary tag and URI.

- The dictionary supports mapping different RA and ONIX schemes using the methodologies developed in the indecs framework and MPEG21 RDD.
- A broad and continually expanding dictionary of well-formed metadata Terms is available to RAs to help in developing their own metadata declarations.

EDItEUR and IDF will continue to work together on the next stages of this development, and intend to do everything possible to ensure that DOI Data Model policies and ONIX formats are consistent in those areas where interoperability is increasingly required. This leaves open the possibility of other bodies participating or joining in the use of this dictionary: IDF and EDItEUR also welcome collaboration with other parties wishing to adopt a similar open and interoperable approach to metadata use.

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

A3.2 DOI Data Model and learning object applications

Some DOI System users are 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/index.cfm?flashplugin=1&fuseaction=scormabt

SCORM has an element ("general.identifier") to carry a unique, permanent identifier for a content object. A DOI name 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 the DOI Data Model. 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."

A3.3 DOI Data Model and MPEG applications

The IDF has taken an active role in introducing the concept of the DOI System 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 standard MPEG-21 "Multimedia Framework" started in June 2000.

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 IDF calls a "Digital Object"; hence DOI names 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, a DOI name is specified as a URI and therefore can be used in MPEG-21. DOI names 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).

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 MPEG-21 Rights Data Dictionary, based on indecs principles, provides a key part of the architecture required to deliver interoperability between develop a digital rights management (DRM) standard systems. Since the DOI System 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 DOI names with structured metadata and MPEG-21 identification and description.

Appendix 4 indecs Data Dictionary

This appendix describes the indecs Data Dictionary which forms a key part of the DOI® Data Model. The indecs Data Dictionary (iDD) contains all Terms used in DOI® AP Metadata Declarations, ONIX messages and other schemes, and formal mappings of the relationships between them. A detailed understanding of the iDD is not required by DOI® name assigners or developers; mapping and related services may be made available through IDF.

A4.1 Introduction

The indecs Data Dictionary (iDD) is an ontology which exists to support semantic analysis and metadata interoperability for all DOI-AP and ONIX metadata schemes. It is the repository of all Terms used in DOI® Kernel Metadata Declarations and Resource Metadata Declarations (RMDs), and other Terms that are required to establish the semantic relationships between them. The iDD is managed on behalf of IDF by a Maintenance Agency.

The iDD is jointly controlled by the IDF and EDItEUR, managers of the ONIX message formats which are the generally used standards in the international publishing industry. It is also fully integrated with the MPEG-21 Rights Data Dictionary (due to become a full International Standard in 2004).

All Terms used in DOI Kernel and Resource Metadata Declarations must 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.

The iDD is the IDF's tool for *semantic analysis*, both to assist RAs in the development and validation of their Metadata Declarations, and to complete the mapping of those Declarations into the iDD itself to create semantic compatibility to enable interoperability with other schemes.

The iDD will be viewable online by RAs, and semantic analysis and mapping is managed by the iDD Maintenance Agency (Ontologyx).

The iDD has developed concepts which are also being worked on elsewhere (e.g. ABC, CIDOC, MetaNet etc). We believe that the semantic mapping of some of these other developments isn't rich or contextual enough to deliver full interoperability: the vocabulary base is too limited and growing it is a non-trivial matter as we know: one-to-one schema mapping of fairly simple terms will start to break down as they become multi-schema and contextualized, unless there is a comprehensive underlying ontology to map onto. The "thesaurus" approach isn't robust enough. Approaches such as the ABC model are geared to describing resources in an event-based way, rather than modelling contexts per se, which limits its scope.

A4.2 iDD Structure

The architecture of the iDD has been developed from the indecs framework analysis. IDF has been closely involved in its development since the beginning of the first indecs project in 1998. IDF was a partner in the Contecs:DD consortium which backed the second stage of the framework's development as the basis of the International Standard MPEG-21 Rights Data Dictionary in 2001-3. Through this process the point was reached in 2003 where an implementation of an operational indecs dictionary supporting lookup, mapping and data transformation was realized although the transformation tools remain under development by Ontologyx.

iDD is an ontology based on the indecs ***ContextModel,*** a simple but powerful data model which takes *activity* as the basis of its semantics. Fundamental terms of the iDD – *Verb, Context, Agent, Resource, Time, Place, Property, Quality and Relator*– provide a framework

by which, through the process of Subtyping, Terms of any type or complexity can be defined and related.

The iDD is housed in an SQL database and its Terms are available for online lookup by RAs. It has a highly generic table structure and the bulk of its knowledge is contained in subject-predicate-object "triples" (compatible with those used in the emerging Semantic Web standards RDF and OWL) which can be grouped in chains and sets of any complexity and granularity (the iDD is expressible in RDF and OWL).

Mapping of Terms into iDD is, of necessity, a process requiring human analysis and validation (see ***A4.6 Semantic Analysis and Mapping of Terms***). However, once mapped, the structure of iDD is capable of supporting metadata queries and transformations to a high level of complexity, including the generation of scheme-to-scheme maps. Though rich, the iDD is equally at home with simple or complex schemes and structures. It explicitly establishes all the relationships required for *semantic engineering* of Metadata Declarations, enabling Declarations to be designed and mapped at the level of granularity appropriate to an AP.

Currently the IDF and EDItEUR community are considering the appropriate format to use for iids (indecs identifiers, the unique identifiers of each element in the dictionary). No final decision has yet been reached, but the discussion is straightforward. The URIs will take the form of DOI names. A DOI name consists of a prefix and a suffix separated by a slash, e.g. 10.1234/1234567

In order to easily manage DOI names for elements as opposed to those for first class content objects, IDF uses straightforward numeric prefixes for all content objects (such as 10.1234/), and reserved specific prefixes for administrative purposes. These will take the form of 10.ap/???? for Application Profiles, 10.ra/???? for Registration Agencies, and so on (where ???? indicates the variable suffix string). Therefore a likely prefix for a metadata element in the indecs Data Dictionary would be 10.iid/

The string would be published as a URI available to those who have access to the iDD, as a DOI name.

In future, these identifiers would be used for establishing interoperability between different IDF and ONIX schemas (and third parties), enabling data transformation and integration between them.

A4.3 iDD Terms: Attributes

Each Term in the iDD has Attributes as shown below. A Term itself is not the TermName but the underlying abstract meaning. Each Term has a unique identifier within the iDD and a public "iDDtag" by which it may be referenced in XML and other schemas. A DOI AP will be established for iDD Terms so that DOI names can be used for lookup and other functions.

A Term may have any number of different names (in different Languages where appropriate) for use in different APs and schemes. For example, a *Writer* in one AP may be the same as an *Author* in another, and both these names can be maintained in iDD under their respective *Authorities.*

A Term may also have any number of expressions of its Definition, also in different Languages if required, appropriate to each RA.

A Term is mapped in the iDD by groups of "triple" Relationships which define its parents and other "family" relationships, and the specific constraints under which it operates. Attributes of an iDD Term are shown in the table below:

Attribute	**occ**	**Example**
TermName	1-n	idd:agentIdentifier

	Type	1	Headword
	Authority	1-n	Idd
	Language	1	En
Idd-id		1	doi:10.idd/idd:agentIdentifier
MeaningType		1	Partly Derived
Definition		1-n	An identifier of an agent.
	Authority	1-n	idd
	Language	1	en
Comment		1-n	In the DOI Kernel Metadata Declaration this element is contextualized to refer only to primaryAgents.
	Authority	1-n	idd
	Language	1	en
Relationships		1-n	1. idd:agentIdentifier idd:isSubClassOf idd:identifier 2. idd:agentIdentifier idd:isIdentifierOf idd:agent

A4.4 iDD Terms: Scope

The iDD contains various TermSets:

TermSet	Description
DOI Kernel Terms	The terms and allowed values of the Kernel Metadata Declaration, from all DOI® RAs using this Declaration.
DOI® RMD Terms	The terms and allowed values of Resource Metadata Declarations (RMDs), from all DOI RAs using these Declarations.
ONIX Terms	All terms used in ONIX messages.
RDD	All terms of the MPEG-21 Rights Data Dictionary (RDD).
Other standard terms	Certain standard terms recommended for use. These include: (a) ISO 3166 Territory Codes (b) ISO 4217 Currency Codes (c) ISO 639 Language Codes (d) ISO 8601 standard formats for descriptions of Time.

All these Terms are available for use by any RA in any combination, along with new Terms which the RA has mapped into iDD. Other Termsets will be added from time to time as required.

A4.5 iDD Maintenance Agency

The iDD is maintained for IDF by an appointed Maintenance Agency. Its functions are:

- To provide semantic analysis for RAs to assist in the creation and amendment of Metadata Declarations.
- To map new Terms into the iDD as required by IDF and DOI System RAs.
- To maintain up-to-date versions of ISO and other standard Terms required by IDF.
- To maintain an updated version of the iDD available for online lookup by RAs.
- To provide services for data querying and transformation between RA, ONIX, RDD and other mapped Metadata Schemes.

The relationship between IDF and the Maintenance Agency is to be managed through a Service Contract. The costs of semantic analysis, mapping and data querying and transformation services will be borne directly by users of the system.

A4.6 Semantic Analysis and Mapping of Terms

All Terms used in Kernel Metadata and Resource Metadata Declarations should be mapped into the iDD. This mapping establishes the relationship between a Term and all other

Terms used in these formats, and is the way in which *semantic integrity* of the DOI® System is achieved.

The unusual aspect of mapping to the iDD is that a mapped term ***becomes a part of the Dictionary itself.*** 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 through Type hierarchies and RelationshipSets.

Mapping is a ***consensual*** exercise. It requires agreement between the organization responsible for managing the AP and the Maintenance Agency of the iDD that a given mapping is a correct interpretation of the meaning of a Term. This consent is registered as authorizations on a Term-by-Term basis in the iDD. Mappings cannot be added arbitrarily to the iDD by either the Maintenance Agency or an RA. Explicit authority is essential because third parties such as other RAs and DOI System users will be reliant on iDD mapping (it follows the third principle of the indecs framework – "*Designated Authority"*).

Mapping is a formal exercise with several required steps for each Term. However, Terms are not mapped in isolation. The first step is that the structure of the whole scheme to be mapped is drawn roughly into a "tree" or set of "trees" whose branches descend from the high-level elements of the iDD (including Context, Verb, Agent, Resource, Time, Place, Quality and Relator) so that its own hierarchy can be understood and Terms mapped in the most appropriate order (starting at the top of each tree).

Each Term is allocated a unique identity as it is "plugged into" the iDD using one or more "triple" relationships with existing iDD Terms. The set of these Relationships is called the Term's ***Genealogy***. The Genealogy identifies those Relationships by which the Term inherits its meaning, or acquires global constraints upon its inherited meaning. The most important of these are the equivalence Relationships (for example, ***cal:journal*** *IsSameAs* ***idd:Journal***) and parent-child Relationships (for example ***onix:Author*** *IsSubClassOf* ***idd:Creator***). A global constraint on this might be ***onix:Author*** *IsCreatorOf* ***idd:Words***. Note that although these examples are shown here for convenience with Names, within the iDD structure each of these Terms is represented not by a Name but by a unique Term Identifier in the form of a DOI name so the iDD is in essense a complex network of DOI names.

Many Terms have *contextual* as well as global constraints, and these are then described in sets of Relationships known as ***ContextualConstraint Sets***. These allow dependencies to be shown. For example, a Term such as ***onix:replacesISBN*** in the ONIX for Books message set refers to the "*International Standard Book Number of a former product which the current product replaces".* Semantic analysis against the iDD deconstructs this into its component elements, and represents them in a formal chain of Relationships like this.

#1 IsA onix:replacesISBN
#2 IsA onix:product
#1 IsAttributeOf #2
#2 IsReplacementOf #3
#3 HasISBN #1

In these Relationships the numbers #1, #2 and #3 represent values of the Terms in any particular instance. These "arbitrary values" re-appear in different relationships, providing the contextual links with which the chain is built. The chain thus described provides all the necessary logic for queries or transformations based on knowledge of the *actual* value in any circumstances of any two of the three values (#1, #2 or #3).

A MeaningType of ***Derived*** or ***PartlyDerived*** is given to the Term. A ***Derived*** Term is one whose semantics are fully explained by its axiomatic Relationships (such as ***onix:replacesISBN***). A ***PartlyDerived*** Term is one which relies to some extent on an external meaning defined in language.

Names, Definitions and ***Comments*** are added to the Term as required.

This is a painstaking process, but it is 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 to legacy data).

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

A4.7 Benefits of mapping through iDD

The level of granularity described above is unnecessary if only two or three schemes are being mapped to one another. However, the fundamental assumption underlying the DOI Data Model 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 data interchange or more complex searching and processing. Semantic integrity on such a scale is 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 richer underlying model in which to precisely locate the others' terms. To give a very simple example, if one scheme has a Term "Author" (meaning "a creator of words") and the other has a Term "Composer" (meaning "a creator of music"), there is no direct relationship between the two: they are not equivalent, and neither is a subtype of the other. To establish a relationship, both need to be mapped to another element (say, "Creator"), of which they are both subtypes. Then to distinguish them, the elements "Words" and "Music" need also to be identified. A common underlying model or ontology is needed, in which the new Terms already exist or can be added, to establish the explicit relationships which the individual schemes lack the richness to express.

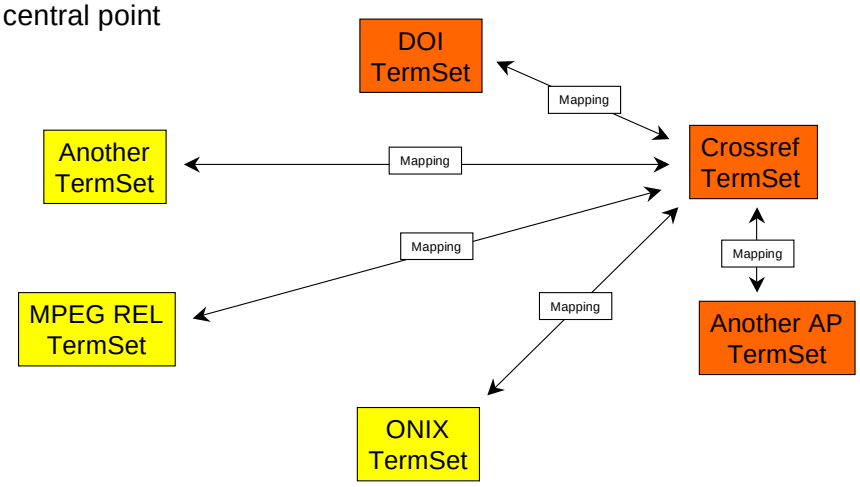
In general, schemes adopt data models which are tailored to meet their own particular requirements, and these are normally not rich enough to support unambiguous mapping. For example, a trial mapping recently between ONIX and a major metadata scheme from the educational world showed many approximate, unresolvable and ambiguous relationships. This is not because of any failings of either scheme, but because they were being tested beyond 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 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.

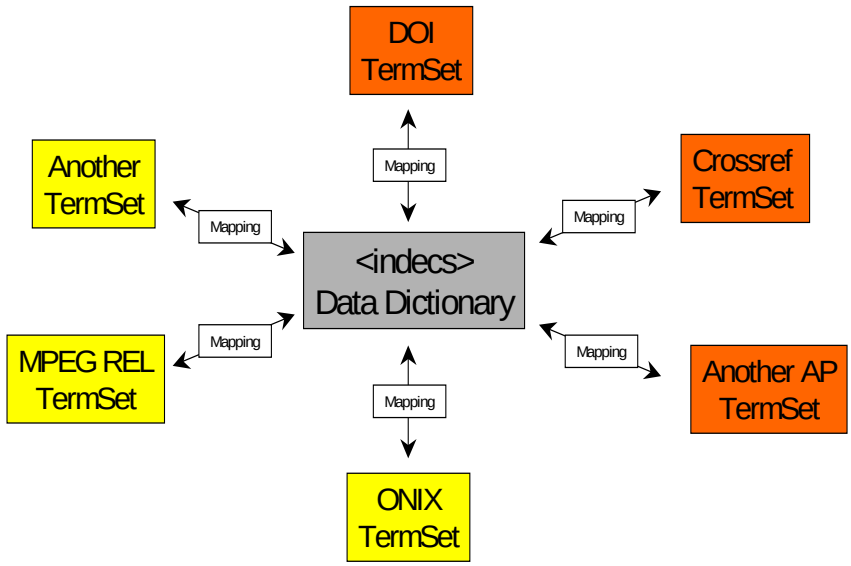
The diagram below illustrates what happens if six schemes need to map to each other. Each scheme must do five one-to-one mappings: a total of fifteen mappings, probably with very mixed results. Any further scheme which joins this community then has to be mapped to each other scheme; and the task grows by arithmetic progression. When there are n schemes, there are (n/2)x(n-1) one-to-one mappings needed. With twenty-five schemes, that is 300 possible one-to-one maps.

The next diagram 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.

Many-to-many mapping with no central point



Many-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 RA are themselves ambiguous or imprecise.* iDD can accurately describe the ambiguity and leave the resolution to users. iDD can achieve accurate mapping *as far as the source data allows,* producing 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 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.

Appendix 5 DOI® Resource Metadata Declaration

This appendix describes the DOI® Resource Metadata Declaration (RMD) a non-mandatory model with comprehensive semantics and an XML schema for declaring metadata of any resource, of any complexity. This RMD is based on the indecs framework and development. Its principal function is for metadata interchange between RAs, but it may be used for any messaging purpose required by an RA (for example, for general metadata capture or delivery). A detailed understanding of the RMD is not required by DOI® name assigners or developers; mapping and related services may be made available through IDF or specialist services.

The DOI® Data Model (DDM) offers RAs a standard form of extended Metadata Declaration, the ***DOI Resource Metadata Declaration (RMD),*** whose primary purpose is for metadata interchange between RAs. This Declaration contains a comprehensive set of attribute semantics (drawn from the iDD), variant methods of attribution and syntax in the form of an XML Schema to provide a comprehensive method for making interoperable Resource description metadata to any level of complexity. The RMD is designed for the description of Resources: that is, creations which are identified with a DOI® name. However, the attribute structure can be easily adapted to provide a framework for standard Metadata Declarations for other kinds of entity (for example, Agents, Events or Places) if required in future.

The structure and Terms of the RMD are derived from the work of the indecs framework, the Contecs:DD consortium (the latter through the development of the MPEG-21 Rights Data Dictionary), and Ontologyx's OntologyX, and reflect the fundamental Terms of the iDD.

A5.1 Basic Attributes of the RMD

The RMD describes ten high-level types of Attributes for any Resource:

Basic Attributes of the RMD

	Attribute Type	Description	Structure of basic Attribute Value
Dependent Attributes (existing only in relation to the Resource)			
1	**Identifier**	A unique label which makes the Resource referable.	Uncontrolled alphanumeric string. Structure may be governed by rules.
2	**Name**	A non-unique label which makes the Resource referable.	Uncontrolled alphanumeric string. Structure may be governed by rules.
3	**Quantity**	A numeric quantification of the Resource.	Number which may be supported by a UnitOfMeasure.
4	**Annotation**	A textual description of, or note or comment about the Resource.	Uncontrolled text string.
Independent Attributes (existing independently of the Resource)			
5	**Category**	A classification or type to which the Resource belongs, or a quality which it has.	Allowed value Code or Name.
6	**Context**	An event or situation affecting the Resource.	Identifier or Name of a related Context.
7	**Agent**	An entity acting in relation to the Resource.	Identifier or Name of a related Agent.
8	**Time**	A Time in relation to which some-	Identifier or Name of a related

		thing happened to the Resource.	Time.
9	**Place**	A Place in relation to which something happened to the Resource.	Identifier or Name of a related Place.
10	**Relative**	Another Resource related to the Resource.	Identifier or Name of a related Resource.

The basic RMD attributes may be simply illustrated like this:

Basic Resource Metadata Declaration model

Resource
Name [0-n] [Value]
Identifier [1-n]* [Value]
Quantity [0-n] [Value]
Annotation [0-n][Value]
Category [0-n] [Value]
Context [0-n] [Value]
Agent [0-n] [Value]
Time [0-n] [Value]
Place [0-n] [Value]
Resource [0-n] [Value]

*The only mandatory attribute of a DOI RMD is (self-evidently) a DOI name, which is a Subtype of Identifier.

The RMD is based on *Subtypes* and combinations of the ten basic attributes. Using these ten attributes, and subtypes from the iDD, the RMD model enables the building of Metadata Declarations (including the Kernel) 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.

A5.2 Semantic engineering of RMDs

RMDs are developed through applying a number of basic processes to the ten basic attributes of a Term. These are listed in the table below, described in the rest of this section. Together they provide a set of tools for what might be termed the *semantic engineering* of metadata declarations. Note that these processes are not schema-dependent: they rely only on the semantic relationships explicitly mapped in the iDD. They may be represented in different ways according to the syntax and naming conventions of any given schema.

Processes of Semantic engineering for RMDs

	Semantic Process	**Description**
1	**Subtyping Attributes**	To generate new Attributes as specialized types of basic Attributes.
2	**Substituting Types**	To simplify a declaration by substituting a Type for one of its Archetypes (parents).
3	**Chaining Attributes**	To give Attributes to an Attribute.
4	**Transferring Attributes**	To create a new Attribute by compressing a Chain.
5	**Reciprocating Attributes**	To reverse the relationship of Entity and Attribute.
6	**Compositing Attributes**	To group two or more distinct elements as a single Attribute.
7	**Allowed Values and Codes**	To provide controlled Values and Codes for an Attribute.

A5.2.1 Subtyping Attributes

Basic RMD attributes are developed through ***Subtypes:*** (that is, by *specializing* some aspect of an attribute of a Term to give it a more precise meaning. For example, a ***Title*** is a Subtype of Name which has been specialized so that it is only valid for a Creation (and not, for example, to a plant or animal). In similar ways a ***DOI*** name is a Subtype of Identifier, ***Height*** is a Subtype of Measure, ***Description*** is a Subtype of Annotation, ***Creator*** is a Subtype of Agent, ***Translation*** is a Subtype of Resource, and so on. Each Subtype is more limited in its scope than its parent or *Archetype*. Subtypes can be specialized to any level.

All Subtypes in the iDD may be used in RMDs, and others may be added to the iDD through the mapping of new Terms from other schemes. Subtyping is the most important mechanism of the iDD for defining new Terms.

Subtypes can be represented in the RMD as shown in this example:

RMD Subtyped Attributes – example

Resource
- Identifier ["10.9999/12345"]
  - Subtype [DOI name]
- Name ["My picture"]
  - Subtype [Title]
- Quantity ["54"]
  - Subtype [DigitalCapacity]
    - Subtype [SizeInKB]
- Category [jpeg]
  - Subtype [MimeType]

In this example one *extended* Subtype hierarchy has also been shown: Quantity *hasSubtype* DigitalCapacity *hasSubtype* SizeInKB. Subtype hierarchies can go on to any required level of granularity, with each Type being progressively more specialized than its parent.

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

The use of Subtypes can be demonstrated in the DOI® Kernel in figure below. Each of the Kernel elements is a Subtype of one of the RMD Basic Attributes (Kernel elements in **bold**):

RMD Subtyped Resource Attributes in the Kernel Metdata Declaration

Resource
- Identifier [Value] [0-n]
  - Subtype [**resourceIdentifier**]
- Identifier [Value] [1]
  - Subtype [**DOI name**]
- Name [Value] [0-n]
  - Subtype [**resourceName**]
- Category [Value] [1]
  - Subtype [**structuralType**]
- Category [Value] [1-n]
  - Subtype [**resourceType**]
- Category [Value] [1-3]
  - Subtype [**mode**]

```
        Agent [Value] [1-n]
                Subtype [primaryAgent]
```

A5.2.2 Substituting Subtypes

Declarations can be simplified radically by substituting the name of the Subtype directly for its parent (or *Supertype*) within the RMD. This can be done from any point in the Subtype hierarchy to any point above it. In the RMD, Subtypes are substituted directly for one of the ten basic attributes. The example given in A5.2.1 can therefore be written much more simply as follows:

RMD Substituted Types – example

```
Resource
        DOI name ["10.9999/12345"]
        Title ["My picture"]
        SizeInKB ["54"]
        MimeType [jpeg]
```

Because the Subtype hierarchies are all explicitly recorded in the iDD, the substitution of Subtypes for their Supertypes or (vice versa) can be managed in an automated fashion to support the expansion or contraction of attributes from one form of Declaration to another, and thereby transformation between schemes.

The figure below shows the Kernel elements with Subtypes (from A5.2.1) substituted as direct attributes of the Resource. In this figure (and those which follow) the parent basic attributes are given in italics to show how they fit into the now-hidden underlying "ten element" structure.

RMD Substituted Subtypes in the Kernel

```
Resource
[Identifier]
        resourceIdentifier [Value] [0-n]
        DOI name [Value] [1]
[Name]
        resourceName [Value] [0-n]
[Category]
        resourceType [Value] [1-n]
        structuralType [Value] [1]
        mode [Value] [1-3]
[Agent]
        primaryAgent [Value] [1-n]
```

A5.2.3 Chaining Attributes

Each of the basic attributes may be an entity in its own right with attributes drawn from the same group 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.* This process may go on to any level of granularity, creating *Attribute chains*. For example, the Resource used in the example above may have Agent details in a simple Attribute chain as follows:

RMD Attribute Chain – example

```
Resource
        DOI name ["10.9999/12345"]
        Title ["My picture"]
        SizeInKB ["54"]
        MimeType [Jpeg]
```

Creator [“John Smith”]
Relative [“Photographers Inc”]
Subtype [“AffiliatedOrganization”]

This can be simplified in turn by substituting Subtypes:

RMD Attribute Chain with Substituted Subtype – example

Resource
DOI name [“10.9999/12345”]
Title [“My picture”]
SizeInKB [“54”]
MimeType [Jpeg]
Creator [“John Smith”]
AffiliatedOrganization [“Photographers Inc”]

The eighth Kernel element (***AgentRole***) is an example of a chained attribute, as shown when it is added in the figure below:

RMD Attribute Chain in the Kernel

Resource
[Identifier]
resourceIdentifier [Value] [0-n]
DOI name [Value] [1]
[Name]
resourceName [Value] [0-n]
[Category]
resourceType [Value] [1-n]
structuralType [Value] [1]
mode [Value] [1-3]
[Agent]
primaryAgent [Value] [1-n]
Category [Value] [1-n]
Subtype [**AgentRole**]

This chain is then simplified by Substituting the Type for its parent, as shown in the figure below:

RMD Attribute Chain with Substituted Type in the Kernel

Resource
[Identifier]
resourceIdentifier [Value] [0-n]
DOI name [Value] [1]
[Name]
resourceName [Value] [0-n]
[Category]
resourceType [Value] [1-n]
structuralType [Value] [1]
mode [Value] [1-3]
[Agent]
primaryAgent [Value] [1-n]
[Category]

Some of the other Kernel attributes will need further attributes, depending on how they are to be referenced. Some will require labels: the ***primaryAgent*** will be identified either by a ***name*** or ***identifier***, as in the figure below

Further RMD Attribute Chain in the Kernel

Resource
[Identifier]
    resourceIdentifier [Value] [0-n]
    DOI name [Value] [1]
[Name]

    resourceName [Value] [0-n]
[Category]
    resourceType [Value] [1-n]
    structuralType [Value] [1]
    mode [Value] [1-3]
[Agent]
    primaryAgent [Value] [1-n]
    [Category]
        agentRole [Value] [1-n]
    [Name]
        agentName [Value] [0-n]
    [Identifier]
        agentIdentifier [Value] [0-n]

Some attributes (including those just added) may have their own attributes in the form of Subtypes, as shown in the figure below:

Further RMD Attribute Chain in the Kernel

Resource
[Identifier]
    resourceIdentifier [Value] [0-n]
    DOI name [Value] [1]
[Name]
    resourceName [Value] [0-n]
[Category]
    resourceType [Value] [1-n]
    structuralType [Value] [1]
    mode [Value] [1-3]
[Agent]
    primaryAgent [Value] [1-n]
    [Category]
        agentRole [Value] [1-n]
    [Name]
        agentName [Value] [0-n]
            Subtype [Value] [0-1]
    [Identifier]
        agentIdentifier [Value] [0-n]
            Subtype [Value] [0-1]

A5.2.4 Transferring Attributes

Another very common form of simplification of Declarations is the compression of Attribute chains into a single ***Transferred Attribute***. In this process, an attribute of one thing is transferred to another in order to “flatten out” the hierarchical structure. In the illustration above, the fact that “John Smith” is affiliated to “Photographers Inc” may be shown by compressing the chain into a single Attribute “CreatorAffiliatedOrganization”, which becomes an Attribute of the Resource. Attribution is thereby “displaced” from the Creator to the Resource. By the combination of substituting Types and Transferred Attributes, the “flattest” practical form of Declaration can be produced. For example:

RMD Transferred Attibute - example

```
Resource
        DOI name ["10.9999/12345"]
        Title ["My picture"]
        SizeInKB ["54"]
        MimeType [Jpeg]
        Creator ["John Smith"]
        CreatorAffiliatedOrganization ["Photographers Inc"]
```

No semantic value is lost as the iDD can add the mapping of the new Displaced Attribute to its chain. However, there are clear limitations to the compression of attribute chains. For example, if in this illustration there were more than one Creator and more than one AffiliatedOrganization, it would not be clear which one was Affiliated to which organization. Wherever a Resource's attributes have many-to-many Relationships among themselves, Transferred Attributes will have limited value.

A5.2.5 Reciprocating Attributes

The ten basic attributes all describe "Has…" relationships, although for simplicity they are shown here as (for example) "Title" rather than "hasTitle" etc. Sometimes it is necessary to show the reverse: that the Resource (or one of its attributes) being described is itself an attribute of something else: in other words, that something else "Has…" the Resource. This is called a *Reciprocal Attribute*. In the illustration above, for example, it may be that we wish to state that "John Smith" is a Member of an Organization called "Photographers Union" – in other words that "Photographers Union" *hasMember* "John Smith". Now, it is technically possible to do this within the RMD and iDD by giving "John Smith" an attribute such as *OrganizationOfWhichEntityIsMember*, but this is cumbersome and unintuitive. Instead, all entity-attribute relationships can be reversed by using a reciprocal Term from the iDD, which normally adds *Is…Of* to the attribute name. The reciprocal of *Member* is therefore *IsMemberOf*. In our example:

RMD Reciprocated Attribute – example

```
Resource
        DOI name ["10.9999/12345"]
        Title ["My picture"]
        SizeInKB ["54"]
        MimeType [Jpeg]
        Creator ["John Smith"]
                IsMemberOf ["Photographers Union"]
        CreatorAffiliatedOrganization ["Photographers Inc"]
```

A5.2.6 Compositing Attributes

Attributes are often represented by two or more elements in a structured group, known in RMD (as in ONIX) as a ***Composite Attribute*** or *Composite* for short. Composite Attributes are defined as Terms in their own right in the iDD. RAs may define their own Composites which can be mapped to iDD like any other Term (for example, the CrossRef AP has Composites for ***doi name_data***, ***article***, ***journal*** and ***author*** among others). ONIX employs a large number of Composites. Composites may also contain other Composites.

For example, a MeasureComposite may be described in a Composite comprised of these basic attributes and Subtypes, as shown in the figure below:

Composite Attribute with unsubstituted Subtypes

```
MeasureComposite
        Quantity [Value] [1]
        Subtype [AllowedValue] [1]
        Category [AllowedValue] [1]
                Subtype [UnitOfMeasure]
        Category [AllowedValue] [1]
                Subtype [Precision]
```

which after substituting Subtypes looks like this:

Composite Attribute with Substituted Subtypes

MeasureComposite
[Quantity]
    Quantity [Value] [1]
[Category]
    Subtype [AllowedValue] [1]
    UnitOfMeasure [AllowedValue] [1]
    Precision [AllowedValue] [1]

For example:

Composite Attribute example

MeasureComposite
[Quantity]
    Quantity [“54”]
[Category]
    Type [DigitalCapacity]
    UnitOfMeasure [KB]
    Precision [Exactly]

This can be shown in place within the example given in A5.2.5:

RMD Composite Attribute - example

Resource
    DOI name [“10.9999/12345”]
    Title [“My picture”]
    QuantityComposite
        Quantity [“54”]
        Subtype [DigitalCapacity]
        UnitOfMeasure [KB]
    MimeType [Jpeg]
    Creator [“John Smith”]
        IsMemberOf [“Photographers Union”]
    CreatorAffiliatedOrganization [“Photographers Inc”]

A5.2.7 Allowed Values and Codes

Some attribute values are drawn from sets of *allowed* or *controlled* values, usually with representative *Codes.* This is always true for Subtypes, and usually for Categories, but is not uncommon for Places and Agents. All AllowedValues are mapped as Terms in iDD. Allowed values (often presented 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.

Allowed values are shown in the RMD using the iDD Term *AllowedValue*. In turn these Values will be drawn from a set of Terms which is known as an *AllowedValueSet* or (when they have codes) a *CodeSet.* Again this can be illustrated in the Kernel, which can now be shown in full as an RMD in the figure below, with Subtypes substituted as appropriate. Note that in some cases an RA may specify any AllowedValueSet, while in others (*structuralType* and *mode*) the AllowedValues are mandated in the Kernel and provided from iDD.

Complete Kernel, showing addition of AllowedValues and AllowedValueSets

Resource
[Identifier] **resourceIdentifier** [Value] [0-n] Subtype [AllowedValue] [1] AllowedValueSet [Value] [1] **DOI name** [Value] [1] *[Name]* **resourceName** [Value] [0-n] *[Category]* **resourceType** [Value] [1-n] **structuralType** [AllowedValue] [1] **mode** [AllowedValue] [1-3] *[Agent]* **primaryAgent** [Value] [1-n] *[Category]* **agentRole** [Value] [1-n] *[Name]* **agentName** [Value] [0-n] Subtype [AllowedValue] [0-1] *[Identifier]* **agentIdentifier** [Value] [0-n] Subtype [AllowedValue] [0-1]

The standard iDD AllowedValues for the Kernel are as follows:

AllowedValues for **structuralType:**
PhysicalFixation
DigitalFixation
Performance
Abstraction
RestrictedType

AllowedValues for **mode**:
Audio
Visual
Audiovisual
Abstract
RestrictedMode

A5.3 Contexts as Attributes

Four of the ten basic attributes shown have a special relationship: three of them (***Agent, Time, Place***) are attributes of the first (***Context***), as is ***Resource*** itself, according to the <indecs> Context Model, as shown in the figure below:

Context Attributes

Context [Value]
Resource [Value] [0-n]
Agent [Value] [0-n]
Time [Value] [1-n]
Place [Value] [1-n]

(Note that Agent is actually a Subtype of Resource in iDD: it has been substituted here).
The Context is at the heart of the indecs Data Dictionary model where it is the key to defining and mapping many Terms. This section explains and illustrates how Contextual attributes may be used in the RMD.

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 Context of a ***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 may be 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 Subtype hierarchy is *not* shown but is shown in footnotes. First, an example where the publication Event is made explicit[1]:

Resource
PublishingEvent ["#1"]
Publisher ["IDG Books Worldwide, Inc"]
Date ["1996"]
Place ["California, USA"]

Secondly, where the Event is implicit[2]:

Resource
Publisher ["IDG Books Worldwide, Inc"]
DateOfPublication ["1996"]
PlaceOfPublication ["California, USA"]

Either of these approaches is supported by the RMD.

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 (shown here as "#1"). 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 Contextual 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
Publisher ["IDG Books Worldwide, Inc"]

---

[1] Showing the Subtype hierarchy, this would look as follows:
Resource
Context ["#1"]
Subtype [PublishingEvent]
[Agent]
Publisher ["IDG Books Worldwide, Inc"]
[Time]
Date ["1996"]
Place ["California, USA"]

[2] Showing the Subtype hierarchy, this would look as follows:
Resource
[HasAgent]
Publisher ["IDG Books Worldwide, Inc"]
[HasTime]
DateOfPublication ["1996"]
[HasPlace]
PlaceOfPublication ["California, USA"]

DateOfPublication [“1996”]
PlaceOfPublication [“California, USA”]
DateOfPublication [“1998”]
PlaceOfPublication [“Paris, France”]
DateOfPublication [“1999”]
PlaceOfPublication [“Moscow, Russia”]

(Note that in this example the *PlaceOfPublication* is an attribute of the *DateOfPublication,* but it could just as easily have been done the other way round).

Our example now contains a level of indirect attribution. Then suppose that the book has different publishers in one of the countries:

Resource
Publisher [“IDG Books Worldwide, Inc”]
DateOfPublication [“1996”]
PlaceOfPublication [“California, USA”]
DateOfPublication [“1998”]
PlaceOfPublication [“Paris, France”]
Publisher [“XYZ Books”]
DateOfPublication [“1999”]
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
PublishingEvent [“#1”]
Publisher [“IDG Books Worldwide, Inc”]
DateOfPublication [“1996”]
PlaceOfPublication [“California, USA”]
PublishingEvent [“#2”]
Publisher [“IDG Books Worldwide, Inc”]
DateOfPublication [“1998”]
PlaceOfPublication [“Paris, France”]
PublishingEvent [“#3”]
Publisher [“XYZ Books”]
DateOfPublication [“1999”]
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 RMD 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.

A5.4 Journal-RMD XML schema

The first RMD to be expressed in the form of an XML schema, called Journal-RMD and dealing with Journal metadata, is under development.

Appendix 6 DOI® Kernel Metadata Declaration: XML Schema

This appendix defines how to express Kernel Metadata Declaration, the mandatory minimum public metadata applicable to general intellectual property resources identified by a DOI® name, in XML.

A6.1 The DOI® Resource Metadata Kernel Declaration

A 6.1.1 Introduction

The DOI® Kernel (or *DOI Resource Metadata Kernel Declaration* to give it its full name) is the IDF's standard for a minimal public declaration of metadata to enable the simple discovery and disambiguation of created resources identified with a DOI name. Each resource identified with a DOI name must have a supporting Kernel Declaration, produced as an XML document according to the XML schema *resourceKernel.xsd* (see A6.5) and its supporting specification (see A6.2).

This document includes a summary of the data elements and allowed values for DOI Kernel declarations, details of which are given in full in the *resourceKernelEnumerations.xsd* (see A6.6) and its supporting specification (see A6.3).

A6.1.2 Allowed values for Kernel elements

Eight of the Kernel elements have controlled lists of allowed values which must be used. These lists are registered in the indecs Data Dictionary (iDD), the IDF's ontology. They are published in tabular form below and included in the *resourceKernelEnumerations.xsd*.

Allowed value lists for two elements (*structuralType* and *mode*) are "closed", and cannot be added to except by change in IDF policy. Lists for the other six elements are "open" and may be added to at any time by a Registration Authority by registering new values with the iDD. These values are then mapped to the DOI® ontology as a whole to ensure full interoperability within the DOI® RA network.

A6.1.3 Use of "Restricted" values

Where a DOI name is issued under a Restricted Application Profile – that is, for private use only – it may still be necessary to issue a Kernel Declaration for system audit purposes, to show that the DOI name is registered and valid. However, the term "Restricted" may be used for all allowed values of mandatory elements, and optional elements left blank.

A6.1.4 Non-Kernel metadata declarations

The Kernel provides a simple, fully interoperable metadata declaration but is not intended to meet the full metadata requirements of a DOI® Application Profile. Metadata declarations of any other kind may be issued in support of DOI names in addition to the Kernel.

A6.1.5 Kernels for other entities

At the time of the issue of this document, DOI names are only being issued to identify created *Resources* (or *Creations*), and not for *People, Places* or *Events*. When, as is likely, DOI Application Profiles for such entities emerge, separate Kernel specifications will be produced.

A6.2 Kernel Metadata Declaration: specification of data elements

XML element [attribute]	Occ	Datatype	Definition	Allowed Values, examples, other constraints
registrationAgency	1	DOI name	The DOI® Registration Agency responsible for issuing a Kernel Declaration.	The DOI name of the RA issuing the Kernel Declaration.
issueDate	1	Date	The Date on which the Kernel Declaration was made.	ISO 8601 format (YYYY-MM-DD).
issueNumber	1	Integer	The sequence number of this Declaration in the series of Kernel Declarations for this DOI name.	First issue=1. If a Kernel Declaration for a specific DOI name is re-issued, this number should be incremented by one.
resource	1-n	Element Group	A group of elements describing a resource identified by a DOI name.	
DOI name	1	DOI name	A DOI name that identifies a resource.	Any DOI name.
resourceIdentifier	0-n	String	An identifier other than the DOI name applied to a resource.	Any alphanumeric string which is unique within its domain of issue: for example, an ID from a legacy scheme or from the internal database of the resource's publisher. A **resourceIdentifier** is to be declared if one exists: some resources may have no Identifier other than a DOI name. If a Resource has, in error, more than one DOI name allocated to it, the alternative DOI name can be shown as a resourceIdentifier.
[type]	1	iDD Term	The type of the resourceIdentifier.	See **Allowed iDD Values for Kernel elements.** This is an open list so new types may be registered at any time by an RA.
structuralType	1	iDD Term	The primary structural type of a resource.	See **Allowed iDD Values for Kernel elements**. The four mutually exclusive **structuralTypes** (Physical, Digital, Performance and Abstraction) allow resources to be simply classified according to their overall form. Note that the structuralTypes may be contained within one another: for example a CD (Physical) may contain files (Digital) which contain recordings of Performances of songs (Abstractions). However the structuralType is always defined by the overall form, so a CD is Physical. Elements of its content can be classified if necessary under **resourceType.**
mode	1-3	iDD Term	The principal sensory	See **Allowed iDD Values for**

			mode(s) in which a resource is intended to be perceived.	**Kernel elements**. **mode** identifies only the principal *intended* modes of percerption of a resource, recognizing that most physical resources are perceivable with all five senses, but that some of these perceptions may be trivial: for example, an ordinary CD or printed book may be touched or smelled, but these are supplementary or incidental to its intended function as a content carrier. For a Braille book, however, touch is a principal mode, as is smell for a perfume.
resourceType	1-n	iDD Term	The general type of a resource.	See **Allowed iDD Values for Kernel elements.** This is an open list so new types may be registered at any time by an RA. **resourceType** typically describes the abstract nature of the content of a resource, irrespective of its structural type. For example, a resource may have the resourceType “JournalArticle” and a structural type of Abstraction, Digital or Physical. resourceType may, however, be extended to include format and genre elements: for example “ClassicalAudioCD” or “MedicalJournalArticlePDF” would be acceptable resourceType values if an RA wished to define and register them. resourceTypes may be defined at any level of granularity that is useful to an RA.
resourceName	0-n	String	A name or title by which a resource is known.	Any. A **resourceName** is to be declared if one exists (some resources may have no name other than a DOI name or other unique identifier). Note that the only distinction between a **resourceName** and **resourceIdentifier i**s that the latter is unique within its domain of issue.
[type]	1	iDD Term	The type of the resourceName.	See **Allowed iDD Values for Kernel elements.** This is an open list so new types may be registered at any time by an RA.
[primaryLanguage]	0-1	ISO code	The primary language of the resourceName.	ISO 639-1 code values.
principalAgent	1-n	Element Group	The principal agent(s) responsible for a resource, normally but not necessarily its creator(s).	The specification of what constitutes a **principalAgent** for any given AP must be declared, but is at the discretion of an RA: for example, a specific AP may require only the Publisher of a resource to be declared, while another may require only the identity of one of its Creators.

agentName	0-n	String	A name by which the principal agent is known.	At least one Name *or* one Identifier of a **principalAgent** must be declared.
[type]	1	iDD Term	The type of the agentName.	See **Allowed iDD Values for Kernel elements.** This is an open list so new types may be registered at any time by an RA.
agentIdentifier	0-n	String	An identifier of the primary agent.	At least one Name *or* one Identifier of a **principalAgent** must be declared.
[type]	1	iDD Term	The type of the agentIdentifier.	See **Allowed iDD Values for Kernel elements.** This is an open list so new types may be registered at any time by an RA.
agentRole	1-n	iDD Term	The role(s) played by the primary agent.	See **Allowed iDD Values for Kernel elements.** This is an open list so new agentRoles may be registered at any time by an RA.

A6.3 Allowed iDD values for Kernel elements

This table shows the Allowed Values for the Kernel Elements which require them, as documented in the *resourceKernelEnumerations.xsd* (see A6.6).

A **Closed list** of values is one that cannot be added to except by a change in IDF policy.

An **Open list** of values is one that can be added to at any time by an RA registering a new value with the iDD. This value will be mapped to an iDD Term (either an existing Term or one created for the purpose) and that Term will then be added to the allowed values in the *resourceKernelEnumerations.xsd* schema.

Element iDD Allowed Values	**Status of value list** Definition	**Constraints and comments**
structuralType	**Closed list**	
Abstraction	Of a resource which exists as a concept.	
Performance	Of a resource which is expressed in transient form.	
Digital	Of a resource which is fixed in digital form.	
Physical	Of a resource which is fixed in physical form.	
Restricted	Of a resource which is described by a restricted DOI Application Profile.	The Restricted structuralType may only be used for resources belonging to RestrictedApplicationProfiles and for which there metadata interoperability is not a requirement.
mode	**Closed list**	
Abstract	Of a resource which is not directly perceivable with any of the five human senses.	
Audio	Of a resource which is principally intended to be listened to.	
Visual	Of a resource which is principally intended to be looked at.	

AudioVisual	Of a resource which is principally intended to be listened to and looked at simultaneously.	
Tangible	Of a resource which is principally intended to be touched.	
Restricted	Of a resource which is described by a restricted DOI Application Profile.	The Restricted mode is used for resources belonging to Restricted ApplicationProfiles and for which there metadata interoperability is not a requirement.
resourceNameType	**Open list**	
Name	A name of a resource.	A resource may have any number of Names.
Title	A name of a created resource.	A resource may have any number of Titles. If a preference is indicated between them then the PrincipalTitle and AlternativeTitle values should be used.
PrincipalTitle	The Title by which a resource is mainly known.	A PrincipalTitle may be supplemented by a Subtitle. The use of PrincipalTitle is contextual, as a resource may be mainly known by different Titles in different contexts. For example, a resource available in different countries may be known by Titles in different languages. By default, the PrincipalTitle means the Title by which the resource is mainly known by the Registrant at the time of issue of the Kernel Declaration.
AlternativeTitle	A Title other than a PrincipalTitle	See comments on PrincipalTitle.
AbbreviatedTitle	A Title which is a shortened form of another Title.	The Title of which this is an abbreviation may or may not be included in the Kernel Declaration.
Subtitle	A Title which is supplementary to another Title.	A resource cannot have a Subtitle without a Title to which it is supplementary. There may be more than one “pair” of Title/Subtitle for any given resource (especially if TranslatedTitles are included). By default a Subtitle is assumed to be supplementary to the PrincipalTitle, or the Title in the absence of a PrincipalTitle.
FormerTitle	A Title by which a resource was once known.	The use of FormerTitle is contextual: it means that at the issueDate of the Kernel declaration, this Title is no longer being actively used in newly created metadata, but may exist in legacy systems.
TranslatedTitle	A Title which is a translation of another Title.	Only a resource’s Title is translated for a TranslatedTitle: it does *not* refer to the Title of a Translation of the resource.
resourceIdentifierType	**Open list**	
DOI name	A Digital Object Identifier issued under the authority of the International DOI Foundation.	

ISBN	An International Standard Book Number, the ISO Standard Identifier for books.	
ISSN	An International Standard Serial Number, the ISO Standard Identifier for published Serials.	
ISTC	An International Standard Text Code, the ISO Standard Identifier for textual works.	
PII	Publisher Item Identifier.	
UPC	A Universal Product Code.	
EditionNumber	An identifier of a resource which denotes its version.	
PartIdentifier	An identifier of a resource in relation to another resource of which it is a part.	
ProprietaryIdentifier	An identifier which is not in public use.	For example, a publisher's own catalogue or stock number.
agentNameType	**Open list**	
Name	A name by which an agent is known.	
PrincipalName	The Name by which an agent is mainly known.	The use of PrincipalName is contextual, as an agent may be mainly known by different Names in different contexts. By default, the PrincipalName means the Name by which the agent is mainly known by the Registrant at the time of issue of the Kernel Declaration.
AlternativeName	A Name other than a PrincipalName by which an agent is also known.	See comments on PrincipalName.
FormerName	A Name by which an agent was once known.	The use of FormerName is contextual: it means that at the issueDate of the Kernel declaration, this Name is no longer being actively used in newly created metadata, but may exist in legacy systems.
IndividualName	A Name of an agent that is a person.	
OrganizationName	A Name of an agent that is an organization.	
agentIdentifierType	**Open list**	
ProprietaryIdentifier	An identifier of an agent issued by an organization for proprietary purposes.	For example, a publisher's internal author identifier.
CaeNumber	The CISAC Composer, Author and Publisher (*Compositeur, Auteur, Editeur*) identifier.	CaeNumber
resourceType	**Open list**	
No values registered yet.		
agentRole	**Open list**	
No values registered yet.		

A6.4 Namespace

All elements within the **KernelMetadata** tags belong 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/.

A6.5 DOI Resource Metadata Kernel XML schema – version 0.9

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://www.doi.org/2004/DOISchema"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:doi="http://www.doi.org/2004/DOISchema" elementFormDefault="qualified" ver-
sion="0.9">
    <xs:include schemaLocation="resourceKernelEnumerations.xsd"/>
    <xs:complexType name="resource">
        <xs:annotation>
            <xs:documentation>DOI Kernel Resource Metadata Version
0.9</xs:documentation>
        </xs:annotation>
        <xs:sequence>
            <xs:element name="DOI name" type="doi:doi">
                <xs:annotation>
                    <xs:documentation>A DOI name that identifies a re-
source.</xs:documentation>
                    <xs:documentation>Any DOI name.</xs:documentation>
                </xs:annotation>
            </xs:element>
            <xs:element name="structuralType" type="doi:structuralType">
                <xs:annotation>
                    <xs:documentation>The primary structural type of a re-
source.</xs:documentation>
                </xs:annotation>
            </xs:element>
            <xs:element name="modes">
                <xs:annotation>
                    <xs:documentation>The principal sensory mode(s) in which a resource is
intended to be perceived.</xs:documentation>
                </xs:annotation>
                <xs:complexType>
                    <xs:sequence>
                        <xs:element name="mode" type="doi:mode" maxOccurs="3">
                            <xs:annotation>
                                <xs:documentation>A principal sensory mode in which a re-
source is intended to be perceived.</xs:documentation>
                            </xs:annotation>
                        </xs:element>
                    </xs:sequence>
                </xs:complexType>
            </xs:element>
            <xs:element name="resourceTypes">
                <xs:annotation>
                    <xs:documentation>The general type(s) of a re-
source.</xs:documentation>
                </xs:annotation>
                <xs:complexType>
                    <xs:sequence>
                        <xs:element name="resourceType" type="doi:term" maxOc-
curs="unbounded">
                            <xs:annotation>
```

```
                        <xs:documentation>The general type of a re-
source.</xs:documentation>
                    </xs:annotation>
                </xs:element>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name=“principalAgents”>
        <xs:annotation>
            <xs:documentation>The principal agent(s) responsible for a resource,
normally but not necessaily its creator(s).</xs:documentation>
        </xs:annotation>
        <xs:complexType>
            <xs:sequence>
                <xs:element name=“principalAgent” maxOccurs=“unbounded”>
                    <xs:complexType>
                        <xs:complexContent>
                            <xs:extension base=“doi:agentType”/>
                        </xs:complexContent>
                    </xs:complexType>
                </xs:element>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name=“resourceNames” minOccurs=“0”>
        <xs:complexType>
            <xs:sequence>
                <xs:element name=“resourceName” maxOccurs=“unbounded”>
                    <xs:annotation>
                        <xs:documentation>A name or title by which a resource is
known.</xs:documentation>
                    </xs:annotation>
                    <xs:complexType>
                        <xs:simpleContent>
                            <xs:extension base=“xs:string”>
                                <xs:attribute name=“type”
type=“doi:kernelResourceNameType” use=“required”/>
                                <xs:attribute name=“primaryLanguage”
type=“xs:language” use=“optional”/>
                            </xs:extension>
                        </xs:simpleContent>
                    </xs:complexType>
                </xs:element>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name=“resourceIdentifiers” minOccurs=“0”>
        <xs:complexType>
            <xs:sequence>
                <xs:element name=“resourceIdentifier” maxOccurs=“unbounded”>
                    <xs:annotation>
                        <xs:documentation>An identifier other than the pimary DOI
name applied to a resource.</xs:documentation>
                    </xs:annotation>
                    <xs:complexType>
                        <xs:simpleContent>
                            <xs:extension base=“xs:string”>
                                <xs:attribute name=“type”
type=“doi:kernelResourceIdentifierType” use=“required”/>
                            </xs:extension>
                        </xs:simpleContent>
                    </xs:complexType>
                </xs:element>
```

```
            </xs:sequence>
         </xs:complexType>
      </xs:element>
   </xs:sequence>
   <xs:attribute name="version" type="xs:string" use="optional"/>
</xs:complexType>
<xs:complexType name="agentType">
   <xs:sequence>
      <xs:choice>
         <xs:sequence>
            <xs:element name="agentNames">
               <xs:complexType>
                  <xs:sequence>
                     <xs:element name="agentName" maxOccurs="unbounded">
                        <xs:annotation>
                           <xs:documentation>A name by which an agent is
known.</xs:documentation>
                        </xs:annotation>
                        <xs:complexType>
                           <xs:simpleContent>
                              <xs:extension base="xs:string">
                                 <xs:attribute name="type"
type="doi:kernelAgentNameType" use="required"/>
                              </xs:extension>
                           </xs:simpleContent>
                        </xs:complexType>
                     </xs:element>
                  </xs:sequence>
               </xs:complexType>
            </xs:element>
            <xs:element name="agentIdentifiers" minOccurs="0">
               <xs:complexType>
                  <xs:sequence>
                     <xs:element name="agentIdentifier" maxOc-
curs="unbounded">
                        <xs:annotation>
                           <xs:documentation>An identifier of an
agent.</xs:documentation>
                        </xs:annotation>
                        <xs:complexType>
                           <xs:simpleContent>
                              <xs:extension base="xs:string">
                                 <xs:attribute name="type"
type="doi:kernelAgentIdentifierType" use="required"/>
                              </xs:extension>
                           </xs:simpleContent>
                        </xs:complexType>
                     </xs:element>
                  </xs:sequence>
               </xs:complexType>
            </xs:element>
         </xs:sequence>
         <xs:sequence>
            <xs:element name="agentIdentifiers">
               <xs:complexType>
                  <xs:sequence>
                     <xs:element name="agentIdentifier" maxOc-
curs="unbounded">
                        <xs:annotation>
                           <xs:documentation>An identifier of an
agent.</xs:documentation>
                        </xs:annotation>
                        <xs:complexType>
```

```
                                    <xs:simpleContent>
                                        <xs:extension base="xs:string">
                                            <xs:attribute name="type"
type="doi:kernelAgentIdentifierType" use="required"/>
                                        </xs:extension>
                                    </xs:simpleContent>
                                </xs:complexType>
                            </xs:element>
                        </xs:sequence>
                    </xs:complexType>
                </xs:element>
            </xs:sequence>
        </xs:choice>
        <xs:element name="agentRoles">
            <xs:annotation>
                <xs:documentation>The role(s) played an agent.</xs:documentation>
            </xs:annotation>
            <xs:complexType>
                <xs:sequence>
                    <xs:element name="agentRole" type="doi:term" maxOc-
curs="unbounded">
                        <xs:annotation>
                            <xs:documentation>A role played by an
agent.</xs:documentation>
                        </xs:annotation>
                    </xs:element>
                </xs:sequence>
            </xs:complexType>
        </xs:element>
    </xs:sequence>
</xs:complexType>
<xs:element name="kernelMetadata">
    <xs:annotation>
        <xs:documentation>DOI Kernel Metadata Declaration 0.9</xs:documentation>
    </xs:annotation>
    <xs:complexType>
        <xs:sequence>
            <xs:element name="registrationAgency" type="doi:doi">
                <xs:annotation>
                    <xs:documentation>The DOI Registration Agency responsible for issu-
ing a Kernel Declaration.</xs:documentation>
                </xs:annotation>
            </xs:element>
            <xs:element name="issueDate" type="xs:date">
                <xs:annotation>
                    <xs:documentation>The Date on which the Kernel Declaration was
made.</xs:documentation>
                </xs:annotation>
            </xs:element>
            <xs:element name="issueNumber" type="xs:unsignedInt">
                <xs:annotation>
                    <xs:documentation>The sequence number of this Declaration in the
series of Kernel Declarations for this DOI</xs:documentation>
                    <xs:documentation>First issue=1. If a Kernel Declaration for a specific
DOI name is re-issued, this number should be incremented by one.</xs:documentation>
                </xs:annotation>
            </xs:element>
            <xs:element name="resources">
                <xs:complexType>
                    <xs:sequence>
                        <xs:element name="resource" type="doi:resource" maxOc-
curs="unbounded"/>
                    </xs:sequence>
```

```
            </xs:complexType>
         </xs:element>
      </xs:sequence>
   </xs:complexType>
</xs:element>
</xs:schema>
```

A6.6 DOI Resource Metadata Kernel XML schema enumerations – version 0.9

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://www.doi.org/2004/DOISchema"
xmlns:doi="http://www.doi.org/2004/DOISchema"
xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified" ver-
sion="0.9">
   <xs:simpleType name="DOI name">
      <xs:annotation>
         <xs:documentation>A DOI name issued under the authority of the International
DOI Foundation.</xs:documentation>
      </xs:annotation>
      <xs:restriction base="xs:string">
         <xs:pattern value="[^\./]+(\.[^\./]+)*/.+"/>
      </xs:restriction>
   </xs:simpleType>
   <xs:simpleType name="term">
      <xs:annotation>
         <xs:documentation>A Term in the iDD ontology.</xs:documentation>
      </xs:annotation>
      <xs:restriction base="xs:NMTOKEN"/>
   </xs:simpleType>
   <xs:simpleType name="structuralType">
      <xs:annotation>
         <xs:documentation>Structural types of a resource.</xs:documentation>
      </xs:annotation>
      <xs:restriction base="doi:term">
         <xs:enumeration value="Abstraction">
            <xs:annotation>
               <xs:documentation>Of a resource which exists as a con-
cept.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="Performance">
            <xs:annotation>
               <xs:documentation>Of a resource which is expressed in a transient
form.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="Digital">
            <xs:annotation>
               <xs:documentation>Of a resource which is expressed in digital
form.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="Physical">
            <xs:annotation>
               <xs:documentation>Of a resource which is expressed in physical
form</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="Restricted">
            <xs:annotation>
```

```
                <xs:documentation>Of a resouce which is described by a restricted DOI
Application profile.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="mode">
    <xs:annotation>
        <xs:documentation>Modes in which a resource may be per-
ceived.</xs:documentation>
    </xs:annotation>
    <xs:restriction base="doi:term">
        <xs:enumeration value="Abstract">
            <xs:annotation>
                <xs:documentation>Of a resource which is not directly perceivable with
any of the five human senses.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Audio">
            <xs:annotation>
                <xs:documentation>Of a resource which is principally intended to be lis-
tened to.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Visual">
            <xs:annotation>
                <xs:documentation>Of a resource which is principally intended to be
looked at.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="AudioVisual">
            <xs:annotation>
                <xs:documentation>Of a resource which is principally intended to be
looked at and listened to simultaneously.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Tangible">
            <xs:annotation>
                <xs:documentation>Of a resource which is principally intended to be
touched.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Restricted">
            <xs:annotation>
                <xs:documentation>Of a resource which is described by a restricted DOI
Application Profile.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="kernelResourceNameType">
    <xs:annotation>
        <xs:documentation>Types (e.g. PrimaryTitle, OriginalTitle, TranslatedTitle,
ChapterTitle), may be registered in the iDD by any RA.</xs:documentation>
    </xs:annotation>
    <xs:restriction base="doi:term">
        <xs:enumeration value="Name">
            <xs:annotation>
                <xs:documentation>A name of a resource.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Title">
            <xs:annotation>
```

```
                <xs:documentation>A name of a created resource.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="PrincipalTitle">
            <xs:annotation>
                <xs:documentation>The Title by which a resource is mainly
known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="AlternativeTitle">
            <xs:annotation>
                <xs:documentation>A Title other than the PrincipalTi-
tle.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="AbbreviatedTitle">
            <xs:annotation>
                <xs:documentation>A Title which is a shortened form of another Ti-
tle.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="Subtitle">
            <xs:annotation>
                <xs:documentation>A Title which is supplementary to another Ti-
tle.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="FormerTitle">
            <xs:annotation>
                <xs:documentation>A Title by which a resource was once
known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="TranslatedTitle">
            <xs:annotation>
                <xs:documentation>A Title which is a translation of another Ti-
tle.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="kernelResourceIdentifierType">
    <xs:restriction base="doi:term">
        <xs:enumeration value="DOI name">
            <xs:annotation>
                <xs:documentation>A DOI name issued under the authority of the Inter-
national DOI Foundation.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="ISBN">
            <xs:annotation>
                <xs:documentation>An International Standard Book Number, the ISO
Standard Identifier for books.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="ISSN">
            <xs:annotation>
                <xs:documentation>An International Standard Serial Number, the ISO
Standard Identifier for published Serials.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="ISTC">
            <xs:annotation>
```

```
                <xs:documentation>An International Standard Text Code, the ISO Stan-
dard Identifier for textual works.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="PII"/>
        <xs:enumeration value="EAN13">
            <xs:annotation>
                <xs:documentation>A European Article Number.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="UPC">
            <xs:annotation>
                <xs:documentation>A Universal Product Code.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="EditionNumber">
            <xs:annotation>
                <xs:documentation>An identifier of a resource which denotes its ver-
sion.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="PartIdentifier">
            <xs:annotation>
                <xs:documentation>An identifier of a resource in relation to another re-
source of which it is a part.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="ProprietaryIdentiifier">
            <xs:annotation>
                <xs:documentation>An identifier which is not in public
use.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="kernelAgentNameType">
    <xs:restriction base="doi:term">
        <xs:enumeration value="Name">
            <xs:annotation>
                <xs:documentation>A name by which an agent is
known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="PrincipalName">
            <xs:annotation>
                <xs:documentation>The Name by which an agent is mainly
known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="AlternativeName">
            <xs:annotation>
                <xs:documentation>A Name, other than the PrincipalName, by which an
agent is known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="FormerName">
            <xs:annotation>
                <xs:documentation>A Name by which an agent was once
known.</xs:documentation>
            </xs:annotation>
        </xs:enumeration>
        <xs:enumeration value="IndividualName">
            <xs:annotation>
```

```
                <xs:documentation>A Name of an agent that is is a per-
son.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="OrganizationName">
            <xs:annotation>
               <xs:documentation>A Name of an agent that is an organiza-
tion.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
      </xs:restriction>
   </xs:simpleType>
   <xs:simpleType name="kernelAgentIdentifierType">
      <xs:restriction base="doi:term">
         <xs:enumeration value="ProprietaryIdentifier">
            <xs:annotation>
               <xs:documentation>An identifier of an agent issued by an organization
for proprietary purposes.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
         <xs:enumeration value="CaeNumber">
            <xs:annotation>
               <xs:documentation>The CISAC Composer, Author and Publisher (Compo-
siteur, Auteur, Editeur) identifier.</xs:documentation>
            </xs:annotation>
         </xs:enumeration>
      </xs:restriction>
   </xs:simpleType>
</xs:schema>
```

A6.7 Example DOI Resource Metadata Kernel Declaration

```
<?xml version="1.0" encoding="UTF-8"?>
<kernelMetadata xmlns="http://www.doi.org/2004/DOISchema">
   <registrationAgency>10.1025/aDoiRegistrationAgency</registrationAgency>
   <issueDate>2004-03-13</issueDate>
   <issueNumber>1</issueNumber>
   <resources>
      <resource>
         <DOI>10.1025/abio.4372.9898</DOI>
         <structuralType>Digital</structuralType>
         <modes>
            <mode>Visual</mode>
         </modes>
         <resourceTypes>
            <resourceType>JournalArticle</resourceType>
         </resourceTypes>
         <principalAgents>
            <principalAgent>
               <agentNames>
                  <agentName type="IndividualName">Lincoln Smith</agentName>
               </agentNames>
               <agentRoles>
                  <agentRole>author</agentRole>
               </agentRoles>
            </principalAgent>
            <principalAgent>
               <agentNames>
                  <agentName type="IndividualName">Marsha
Grossman</agentName>
               </agentNames>
               <agentRoles>
```

```
                    <agentRole>illustrator</agentRole>
                </agentRoles>
            </principalAgent>
        </principalAgents>
        <resourceNames>
            <resourceName type="Title" primaryLanguage="en">DRM in Streaming Me-
dia</resourceName>
        </resourceNames>
        <resourceIdentifiers>
            <resourceIdentifier type="PII">S1031-5806(95)00403-
9</resourceIdentifier>
        </resourceIdentifiers>
    </resource>
  </resources>
</kernelMetadata>
```

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® System specific articles, papers and reports

- ANSI/NISO Z39.84-2005 Syntax for the Digital Object Identifier
  http://www.niso.org/standards/standard_detail.cfm?std_id=480

- Attanasio, Piero. "The use of DOI in eContent value chain. The case of Casalini and mEDRA", Feb 04
  http://dx.doi.org/10.1392/BC1.0

- Cover Pages, Technology Reports; "Digital Object Identifier (DOI) System"; March 2003 http://xml.coverpages.org/doi.html

- DeCandido, GraceAnne A.; "DOI: The Persistence of Memory"; Public Library Association, April 2000
  http://www.ala.org/ala/pla/plapubs/technotes/digitalobject.htm

- Dempsey, Nick. Caccavale, Margherita; "DOI in 2004: What are the tipping points"; EPS Focus Report; January 2004 (Subscription necessary to view report)
  http://www.epsltd.com/clients/briefingReports.asp

- Erickson John S, "Digital Object Identifier," McGraw-Hill Yearbook of Science & Technology 2003. Copyright (c) 2003 by The McGraw-Hill Companies, Inc. All rights reserved. Electronic version graciously provided by The McGraw-Hill Companies and appears by permission.
  http://www.doi.org/topics/je-mh-doi-030970.pdf

- Paskin, Norman; "Digital Object Identifier: implementing a standard digital identifier as key to effective digital rights management"; International DOI Foundation; April 2000.
  http://www.doi.org/doi_presentations/aprilpaper.pdf

- Paskin, Norman; "DOI: a 2003 Progress Report", *D-Lib Magazine*, Vol. 9, No. 6, June 2003. [A summary of DOI progress as of mid 2003]
  http://www.dlib.org/dlib/june03/paskin/06paskin.html
  There is also a Korean translation of this report published in the National Library of Korea Autumn 2003 publication –
  http://www.nl.go.kr/upload/publish/autumn2003.pdf

- 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]
  http://www.dlib.org/dlib/may99/05paskin.html

- 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]
  http://www.doi.org/one2many.pdf

- Paskin, Norman. "Digital Object Identifiers for scientific data". Data Science Journal, Vol. 4, pp12-20, 28 April 2005.
  http://www.doi.org/topics/050428CODATAarticleDSJ.pdf.
  Also available at http://www.datasciencejournal.org/

- 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

- Seybold, Patricia B.; "Protecting your digital assets"; Patricia Seybold Group; Feb 2003
  http://www.doi.org/topics/Protect_Digital_Asset.pdf

- The Seybold Report. "DOI: Slow But Steady Wins the Race? Vol 4. No.18; Dec 22 2004; http://www.seyboldreports.com/TSR/0418page1.html (Available by subscription). Local reprint available here by permission http://www.doi.org/topics/Seybold-vol4no18-22dec04.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.)

- Sieck, Steve. "Economic Benefits of Digital Object Identifier Applications in Content Marketing " EPS White Paper Series. *Part I. Using the DOI to Improve Profitability In Publishers' E-Commerce Operations*, EPS-USA, LLC, 16 December 2003.
  http://dx.doi.org/10.1220/eps1

- 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/estore/download/doi-01.pdf

- Tenopir, Carol; "Thinking about Linking"; Library Journal, July 15 2006
  http://www.libraryjournal.com/article/CA6349053.html

- TSO; "Digital Object Identifiers for Publishers and the e-Learning Community": a Report for JISC from TSO; September 2003.
  http://www.jisc.ac.uk/index.cfm?name=project_tso

- Wildstrom, Stephen H.: "A Library to End All Libraries". Business Week, July 23 2001
  http://www.businessweek.com/magazine/content/01_30/b3742031.htm

- Worlock, David; "DOI: Starting a New Generation", kindly reproduced from EPS Update; 8 May 2003
  http://www.doi.org/news/EPS080503-DOIs.pdf

Identifiers

- Caldwell, Tracey; "*Resolution to Persistence*". Information World Review. October 2005, Issue 217, pp. 20-21. Includes section on cross-publisher linking mentioning CrossRef and DOI. [Available via free subscription from IWR – http://www.iwr.co.uk.]

- Causton, Laurie; "Identifying and describing Web resources"; *Luxembourg: Office for the Official Publications of the European Communities;* 1999 55pp; ISBN 92-824-1516-3

- CENDI Persistent Identifiers Task Group. "Persistent Identification: A Key Component of an E-government Infrastructure". White Paper (PDF) 19 Mar 04.
  [http://www.doi.org/topics/PersistentIDWhitepaper-CENDI.pdf]

- 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, 1997*
  http://www.bic.org.uk/pdf/uniquid.pdf

- Hock, Dee. Birth of the Chaordic Age. Berrett-Koehler Publishers Inc. 1999 [development and social infrastructure of the VISA credit card identifier system]

- Hodgson, Cynthia. “Understanding the OpenURL Framework”. Information Standards Quarterly, Vol. 17, No. 3, pp. 1-4. Includes section “Integration with DOI”. [Available via subscription from NISO – http://www.niso.org/standards/std_resources.html]

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

- IETF URN Working Group. “Uniform Resource Names (urn)”. Modified 31 July 2001. Concluded 22 October 2002.
  http://www.ietf.org/html.charters/OLD/urn-charter.html

- 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]

- MakeAShorterLink webpage
  http://masl.to/about.php

- NISO: Digital Identifiers Roundtable: report on Recommendations and Findings. http://www.niso.org/news/events_workshops/ID-workshop-Report2006725.pdf

- Paskin, Norman; Toward Unique Identifiers; *Proceedings of the IEEE;* 87 (No.7) July 1999; pp. 1208-1227
  http://ieeexplore.ieee.org/search/freesearchresult.jsp?hidden=yes&queryText=Toward+Unique+Identifiers&Go.x=11&Go.y=11

- Paskin, Norman. “E-citations: actionable identifiers and scholarly referencing.”, Learned Publishing, Vol 13, No 3, July 2000, [doi:10.1087/09531510050145308]

- Paskin, Norman: “Components of DRM Systems: Identification and Metadata”; in E. Becker et al (eds), “Digital Rights Management: Technical, Economical, Juridicial, and Political Aspects in the European Union” in the series *Lecture Notes in Computer Science* (Springer-Verlag, 2003), pp 26-61.
  http://www.doi.org/topics/drm_paskin_20030113_b1.pdf

- Powell, A.; Young, J. and Hickey, T. “The PURL-based Object Identifier (POI)”, UKOLN, February 2004
  http://www.ukoln.ac.uk/distributed-systems/poi/

- Saltzer, J.H. “Naming and Binding of Objects” in R. Bayer, R. M. Graham, and G. Seegmüller (eds.)”Operating Systems – An Advanced Course”. *Lecture Notes in Computer Science,* 60, Springer-Verlag, 1978, pp. 99-208
  http://web.mit.edu/Saltzer/www/publications/nbo/nbo.pdf

- Stone, L. Competitive Evaluation of PURLs, March 2000
  http://web.mit.edu/handle/www/purl-eval.html

- Sun, Sam X; "Establishing Persistent Identity Using the Handle System"; *Tenth International World Wide Web Conference*; Hong Kong, May 2001; http://www10.org/cdrom/posters/1136.pdf

- GS1: [Physical bar code identifiers current information] http://www.uc-council.org

- Vitiello, Giuseppe; "Identifiers and Identification Systems: An Informational Look at Policies and Roles from a Library Perspective," *D-Lib Magazine*, Vol. 10, No. 1, January 2004. Available at: doi:10.1045/january2004-vitiello http://www.dlib.org/dlib/january04/vitiello/01vitiello.html

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]

- Bide, Mark. "Rights metadata – standards at last?". imi insights, EPS Ltd, May 2005, pp1-4 http://www.epsltd.com

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

- Bulterman, Dick C.A: "Is it time for a moratorium on metadata?" IEEE Multimedia, October-December 2004, pp 10-17 http://doi.ieeecomputersociety.org/10.1109/MMUL.2004.29

- 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

- Hunter, Jane; "MetaNet – A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains"; *Journal of Digital Information*; Vol1, Issue 8; Feb 2001 http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/

- Lagoze, Carl; "Keeping Dublin Core Simple: Cross-Domain Discovery or Resource Description?"; *D-Lib Magazine*; January 2001. http://www.dlib.org/dlib/january01/lagoze/01lagoze.html

- 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, K.G ; "Functional Requirements for Bibliographic Records". IFLA Study Group on the 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 http://www.epsltd.com/

- Paskin, Norman: "Components of DRM Systems: Identification and Metadata"; in E. Becker et al (eds), "Digital Rights Management: Technical, Economical, Juridicial, and Political Aspects in the European Union" in the series *Lecture Notes in Computer Science* (Springer-Verlag, 2003), pp 26-61. 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

- Rust, G; Bide, M;” The <indecs> Metadata Framework: Principles, model and data dictionary”; 2000. 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,* July 98 http://www.dlib.org/dlib/july98/rust/07rust.html

- SCORM: Sharable Content Object Reference Model http://www.adlnet.org/

- Svenonius, Elaine; The Intellectual Foundation of Information Organization; MIT Press; 2000

- Sowa, John F; “Knowledge Representation: Logical, Philosophical and Computational Foundations”; Brooks/Cole; 2000 http://www.jfsowa.com/krbook/index.htm

- Worlock, David; “PRISM: Magazine publishers moving towards digital delivery standards”; EPS Insights, Oct 2003 http://www.epsltd.com (subscription service)

Technology

- 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, Robert E. & Wilensky, Robert. “A Framework for Distributed Digital Object Services”. *International Journal on Digital Libraries*, (2006) 6(2): 115-123. [doi:10.1007/s00799-005-0128-x]. (First published by the authors May 13, 1995, “A Framework for Distributed Digital Object Services”, [hdl:4263537/5001]).

- McCown, Frank et al; “The availability and persistence of web references in D-Lib Magazine”. A paper presented at the 5th International Web Archiving Workshop held in conjunction with the 8th European Conference on Research and Advanced Technologies for Digital Libraries (ECDL 2005), September 2005. http://www.iwaw.net/05/papers/iwaw05-mccown1.pdf

- Rey, Catherine A; “A System for Persistent Names”; *Web Techniques;* June 2001. http://www.webtechniques.com/archives/2001/06/rey/

- Sun, Sam; Lannom, Larry; Boesch, Brian. “Handle System Overview”. RFC 3650, November 2003. http://hdl.handle.net/4263537/4069

- Sun, Sam; Reilly, Sean; Lannom, Larry. “Handle System Namespace and Service Definition”. RFC 3651, November 2003. http://hdl.handle.net/4263537/4068

- Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. “Handle System Protocol (Ver 2.1) Specification”. RFC 3652, November 2003.

http://hdl.handle.net/4263537/4086

- Zaret, Elliot. "Internet Pioneer Urges Overhaul", Interview with Robert E. Kahn, CNRI. MSNBC, November 21, 2000. Reprinted with permission from MSNBC Interactive News L.L.C.
  http://www.doi.org/news/msnbc_rkahn_interview.pdf

Infrastructure

- Atkinson, Ran; Floyd, Sally; "IAB Concerns & Recommendations Regarding Internet Research & Evolution"; Internet Architecture Board, June 2003
  http://www.iab.org/drafts/draft-iab-research-funding-01.txt

- Biddle, P et al: "The Darknet and the Future of Content Protection"; in E. Becker et al (eds), "Digital Rights Management: Technological, Economic, Legal and Political Aspects in the European Union" in the series *Lecture Notes in Computer Science* (Springer-Verlag, 2003), pp 344-365
  http://crypto.stanford.edu/DRM2002/darknet5.doc

- Blanchi, Christophe; "Distributed Interoperable Metadata Registry", D-Lib Magazine, Vol 7 Issue 12, Dec 2001
  http://www.dlib.org/dlib/december01/blanchi/12blanchi.html

- Byfield, Ted; "DNS – A Short History and a Short Future" reprinted in http://www.firstmonday.dk/issues/issue4_3/byfield/index.html

- Byfield, Ted; "Issues off the map" prepared for the Open Society institute, January 2003 http://a.parsons.edu/~byfield/pdf/OSI-030126.pdf

- Cerf, Vincent G.; Kahn, Robert E. "Lecture: Assessing the Internet: Lessons Learned, Strategies for Evolution, and Future Possibilities" co-recipients of the ACM 2004 Turing Award, presented at ACM Sigcomm 2005 22 August 2005
  http://www.sigcomm.org/sigcomm2005/webcast.html
  link direct to archived version (RealPlayer) at http://beansidhe.isc-net.upenn.edu:8080/ramgen/seas/Turing.htm

- De Soto, Hernando; "The Mystery of Capital: Why Capitalism triumphs in the west and fails everywhere else"; Basic Books; 2000.

- Dyson, Esther; "Digital Object Architecture and the Handle System", In *Online Registries: The DNS and Beyond...*, *Release 1.0*, Esther Dyson's Monthly Report, September 2003, pp. 22-28. [http://doi.contentdirections.com/reprints/dyson_excerpt.pdf]

- Gupta, A; Stahl, D; Whinston, A; "The Economics of Network Management*"; Communications of the ACM*; Sept 1999; Vol 42, No.9
  http://crec.mccombs.utexas.edu/works/articles/cacm99.pdf

- Kahn, Robert; Lyons, Patrice; "Representing Value as Digital Objects", D-Lib Magazine; May 2001
  http://www.dlib.org/dlib/may01/kahn/05kahn.html

- 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 [Chapter 6 – The Power of Infrastructure]

- Paskin, Norman. "Two practical examples of issues in internet governance". Paper presented 6 May 2005 at the Internet Governance Forum, Oxford Internet Institute.
  http://www.doi.org/topics/050504OIIDiscussionForum.pdf

- Payette, Sandra; Blanchi, Christophe; Lagoze, Carl; Overly, Edward; "Interoperability for Digital Objects and Repositories", D-Lib Magazine, Vol 5, Issue 5, May 1999 http://www.dlib.org/dlib/may99/payette/05payette.html

- 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

- Wray, Richard; "EU says Internet could fall apart". The Guardian, Oct 12 2005 http://www.guardian.co.uk/business/story/0,3604,1589902,00.html

- Zittrain, Jonathan: "What's in a name?". Harvard Law School Public Law Research Paper 047. 2002, Federal Communications Law Journal. Volume 55 pp. 155-167 http://www.law.indiana.edu/fclj/pubs/v55/no1/zittrain.pdf

Standards development issues

- Bell, Gordon: "A Time and Place for Standards". ACM Queue vol. 2, no. 6 – September 2004 http://acmqueue.com/modules.php?name=Content&pa=showpage&pid=210

- Evans P and Wurster T (2000). "Blown to bits" ("The explosion in connectivity and the advent of universal standards" : pp 30-37)

- Clark D et al: New Arch: Future Generation Internet Architecture http://www.isi.edu/newarch/iDOCS/final.finalreport.pdf

- Lemley, Mark A (2002) Intellectual Property Rights and Standard-Setting Organizations http://repositories.cdlib.org/boaltwp/24/

Applications: Digital journals and reference linking

- Arms, William A; "Digital Libraries"; MIT Press; 1999

- 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. http://www.dlib.org/dlib/february00/02risher.html

- 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". *D-Lib Magazine,* Sept 2001. http://www.dlib.org/dlib/september01/caplan/09caplan.html

- Paskin, Norman. "E-citations: actionable identifiers and scholarly referencing.", 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

- Vogt, Sjoerd; "Investigative Report; Resolving the Links"; Information Today, April 2003, Vol 20, Issue 4, p.25

Applications: Learning Objects

- Barefoot, T; Flynn, T; Secure Internet Delivery of High-Value Content; Learning Objects Network; 2001

http://www.learningobjectsnetwork.com/resources/LONWhitePaper_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/ltimagazine/article/articleDetail.jsp?id=6787

- Degen, B; Strategic Benefits of Standard Learning Objects; Learning Objects Network; 2001 http://www.learningobjectsnetwork.com/resources/LONWhitePaper_StrategicBenefitsOfStandardLearningObjects.

- Kraan, Wilbert; **"**Identifiers, migrating metadata and orphaned objects"; CETIS; Dec 2003 http://www.cetis.ac.uk/content2/20031216172927

Applications: eBooks

- Walter, Mark. "E-book Project Highlights Role of DOI in Selling Digital Content", *The Seybold Report,* Analyzing Publishing Technologies, June 18, 2001. http://www.seyboldreports.com/TSR/subs/0106/bea.pdf

Applications: Data

- Bourne, Philip; "Will a Biological Database Be Different from a Biological Journal?".PloS Computational Biology. August 26 2005 http://dx.doi.org/10.1371/journal.pcbi.0010034

- CODATA – the Committee on Data for Science and Technology, is an interdisciplinary Scientific Committee of the International Council for Science (ICSU). http://www.codata.org/

- E-BioSci – a European platform for access and retrieval of full text and factual information in the Life Sciences http://www.e-biosci.org/about.html

- Garrity, G. M., C. Lyons; Future-proofing biological nomenclature; 2003; OMICS; 7:31-33. 10.1089/153623103322006562 [this is available on subscription only]. This was presented at: Workshop on Data Management for Molecular and Cell Biology Feb. 2-3, 2003 Lister Hill Center, NLM, NIH Campus, Bethesda, MD. The workshop web page is at http://www.eecs.umich.edu/~jag/wdmbio/. The version of the paper as presented is at the workshop website: http://www.eecs.umich.edu/~jag/wdmbio/garrity.htm

- Harris, Jerald D. (2004): "Published Works" in the electronic age: recommended amendments to Articles 8 and 9 of the Code". Bulletin of Zoological Nomenclature 61(3) September 2004, pp.138-148

- 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

- Lautenschlager, M. and Sens, I. (2003) "Konzept zur Zitierfahigkeit wissenschaftlicher Primierdaten" Information – Wissenschaft und Praxis 54 (2003) 463-466

- 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

- Paskin, Norman. “Digital Object Identifiers for scientific data”. Paper presented at 19th International CODATA Conference, Berlin, 10 November 2004. (See also the accompanying presentation – http://www.doi.org/041110CODATADOI.ppt)
  http://www.doi.org/topics/041110CODATAarticleDOI.pdf

- Paskin, Norman. “Digital Object Identifiers for scientific data”. Data Science Journal, Vol. 4, pp12-20, 28 April 2005.
  http://www.doi.org/topics/050428CODATAarticleDSJ.pdf.
  Also available at http://www.datasciencejournal.org/

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

- 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

- Presentation given by Norman Paskin at the BIC/NISO Meeting: Digital Preservation on 5 Dec 2002
  http://www.doi.org/doi_presentations/021205DigPres.ppt

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

- 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

- 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

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

- Rosenblatt, Bill; Trippe, Bill; Mooney, Steve: “Digital Rights Management: Business and Technology”; Hungry Minds; ISBN 0764548891, 2002

- WIPO (World Intellectual Property Organisation). “Primer on Electronic Commerce and Intellectual Property Issues”. May 2000.
  http://www.wipo.int/copyright/ecommerce/en/ip_survey/introduction.html

Briefings & Presentations from IDF

Please refer to the HTML version of the DOI® handbook for links to the material cited.

- Detailed DOI® System Slide Presentation - A detailed look at the DOI® System policies and operations.
  http://dx.doi.org/10.1000/237

- Presentations from DOI® workshops:
    - Interoperability
      http://www.doi.org/doi_presentations/members_meeting_2004/040623_Interoperability_workshop_Rust.ppt
    - Metadata
      http://www.doi.org/doi_presentations/dec2000/metadata/Workshop_Metadata.pdf
    - Resolution
      http://www.doi.org/doi_presentations/dec2000/resolution/Workshop_resolution.pdf
    - Registration Agencies
      http://www.doi.org/doi_presentations/dec2000/RAs/workshop_RAs.pdf
    - Overview of Handle System Architecture
      http://www.doi.org/topics/HS_overview_slides.ppt

- Handle System Developers' Workshop, agenda and links to presentations, CNRI, Reston, Virginia, held 20 Sep 02