Digital Object Identifiers

Norman Paskin*

Keywords: Identifiers, Digital, e-Commerce, Internet, Standards

Abstract: The Digital Object Identifier (DOI®) is a system which provides a mechanism to interoperably identify and exchange intellectual property in the digital environment. It provides an extensible framework for managing intellectual content based on proven standards of digital object architecture and intellectual property management. The International DOI Foundation, a non-profit membership-based organization, manages development, policy, and licensing of the DOI system. Interoperability in a networked environment requires *resolution*, *structured metadata* and *social infrastructure* to ensure interoperability between processes, within and between organisations, through the use of open standards to offer cost savings.

Many of the issues which we have encountered in developing the Digital Object Identifier (DOI) system are common to TAGA's considerations (TAGA, 2002), especially as we move into a multi-media, open standards world. Interoperability and the solutions to it - such as separation of the problem into components, and the recognition that persistence is fundamentally a *social infrastructure* as well as a *technology infrastructure* problem - have been fundamental design principles of information identifiers that influenced DOI (Paskin, 1999) Equally, the importance of unique identification has been widely recognised by the information "content" communities (Sundt, 1997) and computer scientists (Svenonius, 2000).

1. Interoperability

As information flows become increasingly divorced from logistics, the sharing of information with partners in a value chain is becoming essential to doing business. A DOI assigned to content enhances a content producer's ability to trade electronically by providing a mechanism to increase content availability

*International DOI Foundation

and facilitate e-commerce. DOIs also allow third-party use of material with assurance that the material is securely linked to current information (version, rights, status, etc). DOI applications may use existing systems of identification and description but enable interoperation across media. The key to enabling automation across a supply chain is interoperability. When we consider how we might ensure that two items of data can be used together, one logical approach is to have a central, single, authoritative standard, or (better) place of deposit and management of such data. Yet a single central standard and a single central archive are unrealistic expectations: rather than one source and mechanism, there will be many. Therefore, the various components contributing to, or using, this architecture must interact with each other in a structured way: they must interoperate.

Interoperability in the face of legitimate change has been a theme of the DOI work. The problem of *persistence* is when the dimension of change is time: "how do we interoperate with the future?" (Paskin, 2002). This is provided in the DOI system by the process of resolution. Interoperability *across applications* – when the dimension of change is spatial or logical - needs to recognise that different applications, users and communities have differing requirements for data, differing data models, differing standards, and differing expectations. This interoperability is provided for in the DOI system by the use of well-formed metadata.

In addition to the spectrum of social interoperability considerations (Arms et al, 2002), there are at least six different types of technical interoperability:

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

We must also recognise that persistence and application interoperability are ultimately ensured not by technology, but by *social infrastructure.* This is a key aspect of the DOI system: far from being an abstract specification, it is a working system with many participants, policies, business models, and implementations.

2. Digital Object Identifiers and their role in managing content

2.1 Why unique identifiers are needed

Physical bar codes, standard book numbers ISBN, and standards for credit card numbering all demonstrate that a unique identifier is a key to integrated management of sales, stock, royalties, and so on (UCC, 2002 and Sidman, 2001). In 1997 a project was launched at the Frankfurt Book Fair to extend this principle to the digital world. The DOI (Digital Object Identifier) was proposed to identify items of intellectual property which were distributed via networks - the intangible world where bar codes and the like cannot be applied – and in so doing, to be a tool for the effective commercial use of rights and intellectual property, offering a persistent label and services built on that identifier.

The DOI system is run by the International DOI Foundation (IDF) (www.doi.org), an open membership body funded by a variety of publishing and technology companies and others, committed to integrating DOI into existing standards and business practices by developing a self-funding operation which will gradually assume responsibility for the system, enabling the "bootstrap" membership funding to gradually be removed. The DOI system is up and running – by the end of last year, well over four million DOIs had been issued, with over 200 organisations allocating DOIs; the first four DOI Registration Agencies were appointed during the year (with more planned to come), and DOI is well integrated into several related standards activities, with many applications actively under development. DOI is already a well-known term in parts of the text publishing world: scientific journal publishers were early adopters of DOI and DOI received wide coverage in the specialist press (including a full article in Business Week in the middle of last year). DOI unavoidably involves several fast-moving developments in technical matters (metadata, name resolution and services): technology is in fact the easy part - new business models, infrastructure economics, politics and even applied philosophy have all also played a role.

2.2 Overview of DOI

DOI is a tool for naming "content objects" as first class objects in their own right, with a mechanism to make these names actionable through "resolution". DOI offers persistent managed identification for any entity. But that alone is not enough: managing resources interoperably requires appropriate metadata. Creating a mechanism to provide a structured description of what is identified allows services about the object to be built. The IDF has outlined, and is developing a standard way of doing this and mapping to existing metadata standards such as ONIX (2001) (for product information), Dublin Core (2002) (for resource discovery) and so on, allowing each community to bring its own identifiers, descriptions and purposes into play. Wrapping these tools into a

social and policy framework, through a DOI Registration Agency federation, allows the development of DOIs in a consistent quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

It is not the aim of this presentation to give extensive details of the DOI system: these may be found elsewhere (DOI Handbook, 2002). The following provides:

- a brief description of the key points of DOI implementation
- more detail on the principles which DOI is built on: (1) the level of indirection (separating the name from the particular instance addressed) offered by *resolution*; and (2) the use of well-formed *metadata* to describe the objects identified and so offer appropriate hooks for third-party services; (3) the social infrastructure necessary to make these a practical implementation.
- comments on the relevance of DOI to the graphic arts communities.

3. What and how: a basic guide to Digital Object Identifiers

3.1 What is a DOI?

A Digital Object Identifier is a *digital identifier* for any *object of intellectual property*. A DOI provides a means of persistently identifying a piece of intellectual property on a digital network and associating it with related current data. On digital networks, all intellectual property is simply a string of bits; a DOI can apply to any form of intellectual property in any digital environment. DOIs have been called "the bar code for intellectual property": like the physical bar code, they are enabling tools for use all through the supply chain to add value and save cost.

A DOI is different from commonly used internet pointers to material such as the URL -- Uniform Resource Locator, the usual means of referring to World Wide Web material -- because it identifies an object as a first-class entity, not simply the place where the object is located.

A DOI is also different from commonly used identifiers of intellectual property like standard bibliographic and related identifiers (ISBN, ISRC, etc) because it is associated with defined services and is immediately "actionable" on a network.

A DOI is an implementation of the Internet concepts of Uniform Resource Name and Universal Resource Identifier. But a DOI is different from abstract naming specifications such as URN in that it is a defined *implementation* complete with social and technical infrastructure, ready to use.

DOIs can be used by anyone, independent of the applications which may have been originally devised by the registrant. DOI users can be at any point in an information chain – intermediary, retailer, user, producer, agent, etc, in the same way as the physical bar code is useful to (and used by) a range of retailers, logistics companies, re-sellers etc even though the code is originally assigned by a manufacturer.

3.2 How is a DOI assigned?

A DOI prefix (e.g. 10.1000/) enables a registrant to assign many DOIs by building on the prefix to construct a range of unique identifiers (10.1000/abc, etc). To obtain a DOI Prefix, you need to work either with a DOI Registration Agency or with the International DOI Foundation directly. Working with a Registration Agency is preferred, as this brings with it the advantages of participation in a defined DOI application with others. Some DOI Registration Agencies have been appointed, and additional DOI Registration Agencies will be coming on stream during 2002. DOIs allocated using prefixes purchased directly from IDF have less social infrastructure support than can be given by a Registration Agency, but may be useful if you wish to experiment or consider developing your own applications.

The DOI suffix can be any alphanumeric string that the Registrant chooses. This can simply be a sequential number, or it can make use of an existing (legacy) identifier; the latter may often be administratively convenient for the Registrant and can confer interoperability quite simply: for example, in the CrossRef implementation of DOI, publishers do not need to re-allocate legacy identifiers but are able to incorporate whatever they are using: if one publisher uses Publisher Item Identifiers (PII) and another uses Serial Item and Contribution Identifier (SICI) [each are legitimate standards in that community, with slightly different aims], the resulting DOIs can be used interoperably to create links in the CrossRef database.

3.3 How much does it cost to assign a DOI?

A DOI prefix obtained directly from the IDF costs $1000. This one-off charge allows an unlimited number of DOIs to be constructed using that prefix. There is no limitation placed on the number of DOI prefixes that any organization may choose to apply for. If you work with a Registration Agency, that agency is free to set its fees independently of the IDF. This allows more flexible pricing and a wide variety of potential business models using third part registration agencies, in recognition of the fact that a simple model is not a "one size fits all" solution

3.4 Why select a Registration Agency to assign DOIs?

Registration Agencies (RAs) are established to provide services on behalf of specific user communities. Choosing an appropriate RA will give you access to DOI services and implementations offered by the RA for that community.

There will be a growing number of RAs with sectoral specialisms which may have global application. At the same time, there may be 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 use the full range of services that are offered.

If an appropriate Registration Agency able to meet specific needs does not exist, IDF can help to form one. The IDF will act as a "default" Registration Agency for the foreseeable future, to host registration of such DOIs until an appropriate Registration Agency can take over. IDF can also form working groups of like-minded organizations that may wish to establish a collaborative activity to form an RA, and stimulate the development of business opportunities. It will not compete with RAs that have an established market position.

3.5 How is DOI governed?

The IDF governs the system, to ensure that all applications follow common rules. The system itself has several components: the technology is based on open agreed standards, while the infrastructure is defined by agreements between the various organisations which run the system, such as the Registration Agencies and the technology providers.

The Foundation was created in 1998 and supports the development and promotion of the Digital Object Identifier system as a common infrastructure for content management. The Foundation is controlled by a Board elected by the members of the Foundation, with an appointed full-time Director. Membership is open to all organizations with an interest in electronic publishing and related enabling technologies.

3.6 What is the relationship between a DOI and other standards?

The Digital Object Identifier (DOI) uses open standards such as the Handle system and indecs framework, and can integrate with existing identifiers (they can be incorporated as a suffix into a DOI) and with other network services. DOI builds on open Internet standards and works with information industry bodies wherever possible to ensure compatibility and interoperability.

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.

In addition the IDF has a number of other relationships with significant development and standards activities in many areas of intellectual property and technology. Some of these are specific to particular application areas, and are undertaken in order to seed activities and outreach from the DOI to potential implementations.

3.7 What is the relationship between the DOI System and other technologies?

The DOI system is an application of the Handle System (a resolution system) to intellectual property. It adds to the Handle System an approach based on structured associated metadata, policies, procedures, business models and application tools. Initial implementations are now being supplemented by increasingly sophisticated value-added tools for metadata management and content management, which will use the Handle System multiple resolution function.

DOI is also an implementation of the indecs metadata framework. The IDF is one of the organisations which developed the original indecs framework and now developing it further.

DOIs are designed for use in any digital networks, not just the World Wide Web, which is only one recent aspect of the evolution of digital networks and the use of digital objects within them. DOIs can be used in open or proprietary digital networks in broadcasting, multimedia systems, or indeed any conceptual framework. DOIs are an abstract specification, which have a reference implementation in current Internet technologies. Even on the Web, only some aspects such as single redirection can be accomplished with some existing technologies. Developing concepts such as Web Services promise to make available other tools. Metadata tools such as RDF may eventually be readily usable to describe indecs relationships. We welcome these as synergistic efforts. However no other current technologies offer the same packaged combination of multiple resolution; well-formed metadata; social infrastructure; and non-proprietary non-commercial operation supported by a wide range of content and technology providers.

3.8 How do we participate in DOI development?

Options include: *obtaining a DOI prefix* and assigning DOIs on an experimental basis; *joining an IDF working group* to work with others in a defined problem area; or *joining the IDF as a full member*, with rights to participate in all working groups.

Applications can range from simply DOIs being a persistent redirection to a single URL (which is easily accomplished) to advanced applications and services. Frameworks for such services are now being developed to ensure interoperability; the starting point for such advanced applications is the registration of a set of metadata appropriate to the particular community use being conceived. This is an Application Profile, which is built in a structured way. DOI does not mandate a single metadata standard; you may use any existing metadata standard; it does however require that for full interoperability the metadata set be mapped to the DOI Namespace (data dictionary).

DOIs can currently be applied to any piece of intellectual property (creation), but not to entities such as people and agreements. However we may well extend the concept if appropriate at a later date.

Registration Agencies and registrants abide by rules of the system, which are intended solely to maintain a level playing field. These mandate for example that no consolidated data about use of a specific DOI is made public or available to other than the registrant. They also mandate rules as to syntax and services. Rules are still being defined as the system evolves.

3.9 Do customers using DOI have to do anything special?

Applications using DOI can be constructed on a web site with full functionality behind the scenes. For some applications, users may find it helpful to load a small free plug-in if the browser they are using does not support URN resolution. You may put a DOI anywhere you like. A DOI may be printed or made explicit within a digital object; or it may be hidden by e.g. underlying a hyperlink. Therefore it can either appear as a DOI, or the user may never know that a DOI has been used to "power" her transaction.

4. How and why DOI confers interoperability

DOI has used as reference principles and implementations the Handle System (CNRI, 2002) for resolution and the indecs framework and data dictionary for metadata. It is possible that other approaches could be substituted for these but this would not alter the fundamental concepts.

4.1 Resolution

A name (or unique identifier) for a digital object enables that name to be resolved to one (or many) of several different pieces of data which may be associated with the digital object. Such pieces of data can be locations of the object, or services about the object, or any other defined piece of data. Resolution enables a single name (the identifier) to be used persistently to manage the object, even if any of those pieces of data (like location) change.

Resolution therefore (a) enables persistence and (b) enables multiple services to be directly associated with the DOI.

A DOI is a name (identifier) for an entity in a network environment. Entities identified by a DOI may be of any form, including both tangible entities ("manifestations") and abstractions (sometimes called "works"). Resolution is the process of submitting an identifier of an entity to a network service and receiving in return one or more pieces of current information related to the identified entity. In the case of the Domain Name System (DNS), as an example, the resolution is from domain name, e.g., www.doi.org, to a single IP address, e.g., 132.151.1.146, which is then used to communicate with that Internet host. In the case of the DOI, using the Handle System as a reference implementation, the resolution is from a DOI, e.g., 10.1000/140, to one or more pieces of typed data: e.g. URLs representing instances of (manifestations of) the object, or services such as e mail; the resolution of one identifier to multiple data is a "multiple resolution" mechanism. Resolution can be considered as a mechanism for maintaining a relationship between two data entities; an item of metadata is a relationship that someone claims exists between two entities: therefore, such metadata relationships between entities may be articulated and automated by resolution. Using multiple resolution, a DOI can be resolved to an arbitrary number of different associated values: multiple URLs, other DOIs, or other data types representing items of metadata. Resolution requests may return all associated values of current information, or all values of one data type; these returned values might then be further processed in a specific "client" software application. At its simplest, the user may be provided with a list of options; more sophisticated automated processes would allow for the automated choice of an appropriate value for further processing.

Resolution provides a mechanism for persistence of URLs, by interposing a level of managed redirection. The lack of persistence in identification of entities on the Internet is a commonplace. Across the Internet, the rate at which once-valid links start pointing at non-existent addresses – "link rot" -- is as high as 16 percent in six months (Dowling, 2001). When writing this paper, I consulted two articles of interest that I had printed for perusal from current web sites in recent weeks; on checking the URLs, one had changed to an archival URL and the other, initially free, has reverted to one accessible only to subscribers. This demonstrates that not only *location*, but also other relevant properties like *access rights*, may change and need to be considered in managing persistence.

4.2 Metadata

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

identifiers tell you something about the thing that they identify – for example, since "ISBN" is the acronym of "International Standard Book Number", the identifier "ISBN 1-900512-44-0" can reasonably safely be assumed to identify a book (always assuming that ISBN rules have been correctly followed, which is not universally the case). However, to find out *which* book it identifies, it is necessary to consult metadata – the identifier links the metadata with the entity it identifies and with other metadata about the same entity. Metadata is an integral part of making the identifier useful.

A metadata system designed for stability needs to be multimedia, multi-functional, multi-level, multilingual, multinational and multi-platform. Such an approach is said to be well formed. <indecs> (Interoperability of Data for Electronic Commerce Systems) (Rust, 2000a and 2000b) was a project that with the backing of the European Commission brought together a global grouping of organizations with an interest in the management of content in the digital environment. The <indecs> project was created to address the need, in the digital environment, to put different creation identifiers and their supporting metadata into a framework where they could operate side by side, especially to support the management of intellectual property rights.

<indecs> was a time-limited project, which finished its work early in 2000. Its output is highly regarded and its analysis has been adopted in a number of different implementations. The resulting indecs framework was a *reference model*. In order to implement the reference model, a practical reference implementation was required. DOI-Namespace was an implementation of indecs initiated by the International DOI Foundation for use with DOI; later this was widened to indecs2RDD (IDF Announcements, 2001), which is a generic Rights Data Dictionary (RDD), a common dictionary or vocabulary for intellectual property rights based on the <indecs> framework, and hence a *reference implementation* of indecs, which provides benefits of easier and widespread interoperability.

Because rights metadata is inseparable from other metadata, and because the indecs framework specifies a general metadata framework, the work done in developing indecs2RDD also deepened and expanded the original indecs framework, building on it whilst providing a practical reference implementation. The dictionary resulting from this activity was adopted (in Dec 2001) as baseline technology for the ISO-MPEG-21 Rights Data Dictionary standard and is now being actively developed further.

What does it mean for metadata to be "well formed"? There are only two types of metadata that can be regarded as well formed (Rust, 1998).

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

- All other metadata (if it is well formed) needs to be drawn from a controlled vocabulary of values, which are supported by a data dictionary in which those values are concisely defined. This means that the values in one metadata scheme can be mapped to those in another scheme; this mapping may not be exact – where two definitions in one scheme both overlap with (but are not wholly contained within) a single definition in another, for example. However, the use of a data dictionary avoids ambiguity: where precision of meaning is essential, human beings can clarify definition through a process of dialogue. This is not generally the case with computers.

The need for something like indecs has arisen from the growth of the digital world but in theory could have been created in a non-digital, non-network world, since indecs is essentially a general ontology, independent of any digital network – it is not in other words in any way tied to the Web in preference to other mechanisms. <indecs> will be implemented on the Internet and other networks through implementing things such as DOI services using it, linking resolution and metadata.

The mapping between different metadata schemes may be more or less exact. It may also involve considerable loss of information or no loss of information at all. It is obviously advantageous to achieve as close a mapping as is possible; this is most easily achieved between schemes that share a common high-level data model. The <indecs> data model underlies all DOI metadata. The same analysis underlies ONIX International, rapidly becoming widely accepted as the metadata dictionary for the publishing industry internationally. Similar developments are now occurring in other media sectors (through e.g. the adoption of indecs by MPEG-21).

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

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

The <indecs> data model was devised to cover the same field of endeavour as the DOI – all types of intellectual property ("creations" in <indecs> terminology). The fundamental principles it defines are however applicable to any data representation. It is an open model, which is designed to be extensible to fit the precise needs of specific communities of interest.

The adoption of the <indecs> metadata model gave DOI metadata a firm basis in an intellectual analysis of the requirements for metadata in a network environment that has been tested in real world applications. It will provide easy interoperability with other metadata schemes constructed using the same analysis and a basis for interoperability with metadata schemes based on alternative analyses.

However, it does not greatly matter to the DOI whether the <indecs> analysis and developments based on its framework come to be widely used for the management of intellectual property on the Internet (although we believe it will be very helpful if they do). Good data models, based on rigorous analysis, provide a good basis for the management of intellectual property entities in the network environment, and will be essential. Data dictionaries and transfer protocols based on the <indecs> analysis are already being implemented in commercial contexts.

All a DOI needs is a few kernel elements and a map to a consistent data model. We use an underlying model as a way of guaranteeing that those few elements are useful when people want to extend on them. The reason for using the <indecs> model is that it is alone in having demonstrated its extensibility to real-world transactions, through rights management. DOI describes what is identified in a structured way and allows services about digital content objects to be built for any purpose. IDF guidelines provide a standard way of doing this, and hence a means of mapping to existing standards such as ONIX, Dublin Core and so on, allowing each community to bring its own identifiers and descriptions into play. Wrapping these tools into a social and policy framework, through the Registration Agency federation, allows the development of DOIs in a consistent

quality-assured way across many sectors, opening the possibility of managing multimedia objects seamlessly.

4.3 DOI and 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 is free at the point of use, there is a small fee to an assigner for creating a DOI (a few cents), because we have deliberately chosen to make the DOI a self-funding (though not for profit) system. Our task now is to show that DOI offers value for money as a tool which producers such as image databases can use.

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 (Brin, 1999). URLs for example have a clear technical infrastructure (standards for how they are made), but a very loose social infrastructure (anyone can create them, with the result that they are unreliable alone for long term preservation use as they have no guarantee of stability let alone associated structured metadata). Product bar codes, Visa numbers, and DOIs have a tighter social (business) infrastructure, with rules and regulations, costs of maintaining and policing data – and corresponding benefits of quality and reliability (When a credit card is presented, we can be reasonably certain that the number is valid, and has been issued only after careful correlation with

associated metadata by the registrant). It does not necessarily imply a centralised system – it may be a distributed system (like domain names), but it must have some form of regulation.

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

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

DOIs be made available at "no charge", if the costs of doing so can be met from elsewhere (there is no such thing as "free", only "alternatively funded"). IDF itself is willing to allocate a DOI prefix free of charge to organizations or limited experimental non-commercial uses. For the longer term, the business model includes two separate steps: a business relationship between IDF and an RA (the "franchise fee"); and a business relationship between an RA and a DOI registrant (the "registration fee"). The two are not directly connected; this enables the RA to offer to registrants any business model whatever, which suits its needs. This could include assigning DOIs without charge. Hence DOIs can be used in both commercial and non-commercial settings, interoperably. Like any

other piece of infrastructure, an identifier system (especially one which adds much value like metadata and resolution) must be paid for eventually by someone. So an organization could, if it wished, assign DOIs freely (registration fee zero to registrants) and subsidize this added-value service by paying a franchise fee to IDF from a central fund, as an acceptable cost for supporting the service.

5. Current Applications of DOIs

Over 200 different registrant organizations have so far allocated approximately 5 million DOIs. Because the origins of the DOI were in the text sector, an initial large implementation covering half of these registrants was from traditional print-publishing companies that have already established major online publishing programs. However the fundamental design of the system is applicable to any media or content. The IDF is working closely with many businesses in other sectors of the "content industries" to extend the application of the DOI to many other types of intellectual property

DOIs are now widely used for the identification of scientific articles (and their citation electronically) (Paskin, 2000a), which form the backbone of the peer-reviewed record of science. Through the CrossRef consortium, in which 101 publishers are collaborating at the time of this seminar, over 4.3 million DOIs have been allocated so far to scientific articles, including extensive back files (CrossRef, 2002 and Brand, 2001). In addition to the benefits of persistent resolution, and defined metadata and services, there are some instant benefits in interoperability where none existed previously: CrossRef is now considering application of DOIs to other scientific publication types, including conference proceedings, encyclopaedia entries, and book chapters; and is exploring identification of ancillary material such as images.

Learning Objects Network Inc has been appointed as a DOI Registration Agency and intends to register DOIs for use in the management of learning objects, 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. There are other related projects such as Curriculum Online in the UK, which are also considering the use of DOI. In the software sector, the Software & Information Industry Association has recommended DOI as "The Keystone for Digital Rights Management" (SIIA, 2002).

IDF has taken an active role in introducing the concept of the digital object identifier to the MPEG-21 multimedia framework activity [http://mpeg.telecomitalialab.com/]. The 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

audio-visual works. What MPEG calls a "digital item" can be considered a sub set of what DOI calls a "Digital Object"; hence DOIs can be used to identify MPEG-21 Digital Items (MPEG-21, 2001).

A number of applications in the business sector, including tender documents, financial prospectuses, and mortgage documents are under discussion; DOIs are being considered for scientific data sets, including the allocation of DOIs to biological images within the E-BioSci project (EMBO, 2002). These are just a few of the prospective applications which are being discussed.

6. Some Issues in the Identification and management of images

6.1 Existing standards activities relating to identification and metadata

A brief survey shows that initiatives for standardisation of image identification and metadata have largely ignored interoperability issues. Major commercial image houses use internal numbering schemes (e.g. Corbis (2002); within Getty (2002) each brand uses its own numbering system), perhaps viewing interoperable identifiers as a threat; in fact, they are not: the text industry recognised many years ago that having a standard identifier for books (ISBN) was a significant aid to supply chain interoperability (Ehlers, 1994), and the development of internet identification schemes such as URL have clearly led to an opening of opportunities. DOI has been presented to trade associations of picture associations, agencies and libraries in Europe (CEPIC, 2002) and the US (PACA, 2002). An early analysis for DOI implementations in technical images was developed with the BioImage consortium (BioImage, 2000); this is now being developed further in the E-BioSci project, which is a member of the IDF.

The DIG35 Initiative Group, operating within the sphere of the consortial 80-member Digital Imaging Group, focuses upon digital image metadata standards to support efficient archiving, indexing, cataloguing, reviewing, and retrieving individual images (XML, 2001) NISO is developing technical metadata for still images (NISO, 2001). AIIM International has established a new standards project to produce Guidelines for the creation of Metadata in the Document Imaging Environment (AIIM, 2002). Each of these efforts could integrate well idf account is taken of interoperability tools such as indecs and DOI. Three years ago the Visual Creators Index (1999) outlined proposals for image identification at source, a project, which does not seem to have progressed significantly.

6.2 Rights

Managing the "bundle" of rights for a creative work becomes extremely complicated in a digital environment (Rosenblatt, 2002). Due to the quantity of rights that must be administered: web sites for example can contain text, artwork, design elements, audio, still and moving images that appear

inextricably bound together on the site, but each may arise from separate sources with separate copyright interests. For re-use, copyright in each element must be identified and cleared. Multimedia works are created from a series of separate products protected by copyright, as well as materials that may be public domain, all assembled in a way that itself forms a new and separately copyrightable creative work (Zorich, 1998).

Images can have extremely intricate layers of rights that make their digital distribution legally complicated. Multiple copies or "generations" of an image can be developed, with different rights emerging at each step of the process (Sowa, 2000 and Trant, 1996). Use of a digital image may involve rights clearance with a publisher, the original photographer, one or more copy photographers, and the creator of the work portrayed in the image. Documentary photographs of works in the public domain may have layers of rights associated with them.

<indecs> and DOI were designed to be readily extensible into the field of rights management metadata (Paskin, 2000b), the data that is essential for the management of all e-commerce in intellectual property. The <indecs> analysis asserts that it is essential for the dynamic data necessary for the management of rights to be built on a foundation of the rather more static data that identifies and describes the intellectual property, and that these two layers of metadata can easily interoperate with one another. There is no logical separation of rights metadata from many other metadata; indecs2RDD is in fact a deepening and extension of the fundamental indecs model which has been widely endorsed. The extension of indecs2RDD on the basis of digital rights management does not imply in any way a model which is only applicable to "commercial" data; indeed the metadata tools we are building are highly relevant to public data, and in the indecs model a transaction can be free of any charge but still follow the fundamental model of usage.

6.3 What precisely is being managed?

A major issue in the management of images is what [precisely] is going to be identified and hence managed. Electronic publications are not centrally "registered" anywhere. Nor is there a standard identifier that is widely used to identify electronic publications (as the ISBN and ISSN are used for print publications), and which could be used to compile such a listing. Further, images may be now referenced in several different versions.

A deeper problem comes from considerations of precise structured metadata. Management has conventionally involved consideration only of "manifestations" of works. These physical manifestations – particularly books but also physical photographs– have been susceptible to relatively straightforward unique identification. For over a quarter of a century, almost all

the books deposited at the British Library will have had an ISBN, a way of identifying a specific manifestation. With printed publications, it is thus possible for relatively unambiguous communication to take place between library and publisher. As we move in the direction of electronic products, particularly online publications, the situation becomes more complex. While it may often make sense for different manifestations of the same intellectual content to have different identifiers (e.g. two different resolutions of a single underlying image), the underlying abstract work is the same. This makes decisions on what is and is not identical in intellectual content (and what to archive) even more difficult to discern; a fundamental requirement for well-formed metadata is the use of an ontology ("an explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them") which describes such relationships. In the digital environment, there can be a profusion and different, related, manifestations of an underlying work, as John Sowa puts it: "Computers multiply entities faster than Ockham's Razor can shave them."

Once inside an image database system, unique identification is a matter only for those who manage the system. Nevertheless, careful thought shall be given to the ways in which unique identifiers are used at the point of deposit and retrieval, particularly to the extent that they may be used in the future as finding and location aids. The long-term value of the image archive will depend, to some extent at least, on the approach taken to identification. In an analysis of digital preservation issues, Bide et al (1999) conclude: "In this connection, we find the approach being taken by the International DOI Foundation persuasive. This work, which is closely related to the work of the <indecs> project, proposes that a limited kernel of metadata should be deposited alongside every registered DOI. The kernel metadata will be supplemented and qualified for different genres of content. An approach similar to this, in which a minimal, but tightly defined set of metadata is expected to be deposited by the publisher, would appear to us to be a realistic approach".

6.4 Granularity

Digital publication also allows the publication and interchange of smaller (and hence many more) components – whereas in the physical world a book is transacted as a whole, digitally its component chapters may be manipulated as independent objects, perhaps with no information as to context. In the text world, the ISBN working group is setting up a sub-group to look at how ISBN can be extended to include fragments, especially of digital publications, which will include embedded graphics. This issue of granularity is one which indecs has analyzed closely.

6.5 Copy protection

A further issue which seems certain to cause much debate is the issue of copy protection and related licensing rights. The VisiCalc software, and the ability to study how an early PC program was designed, might have been lost to the public forever if Dan Bricklin's original company, Software Arts, had not turned out to have a solitary unprotected version (Lillington, 2001). That one copy became the download on what is now the most popular part of Bricklin's website. "Copy protection will break the chain of formal and informal archivists who are necessary to the long-term preservation of creative works," says Bricklin (2001). One need not go so far as Bricklin and other advocates like Laurence Lessig (1999) that "copyright doesn't work in the digital age" to recognise that this is a practical issue for preservation and access of images, and one which seems to get relatively little consideration in discussions of the problem of archiving. Copy protection per se is not a feature of an identifier mechanism like DOI (deliberately); rather, DOIs are "hooks" by which many different copy-protection mechanisms and other rights management features might connect for interoperation.

Literature Cited

AIIM Standards; Metadata Work
2002 http://stnds.aiim.wegov2.com/

Arms, W.; Hillman, D.; Lagoze, C.; Krafft, D.; Marisa, R.; Saylor, J.; Terrizzi, C.; Van de Sompel, H.
2002 "A Spectrum of Interoperability"; D-Lib Magazine. http://www.dlib.org/dlib/january 02/arms/01arms.html

Bide, M.; Potter, E. J.; Watkinson, A.
1999 "Digital Preservation: an introduction to the standards issues surrounding the deposit of non-print publications". http://www.bic.org.uk/digpres.doc

BioImage metadata framework, based on the <indecs> metadata framework
2000 http://www.doi.org/topics/metadata.rtf

Brand, A.
2001 CrossRef turns One. http://www.crossref.org

Bricklin, D.
2001 "Copy Protection Robs The Future"; www.bricklin.com http://www.bricklin.com/robfuture.htm

Brin, David
1999 "The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?"; Addison-Wesley; 384 pages.

CNRI (The Handle System)
2002 http://www.handle.net

Co-ordination of European Picture Agencies, Press and Stock (CEPIC)
2002 http://www.cepic.org/#

Corbis
2002 http://www.corbis.com/

CrossRef Web Site
2002 http://www.crossref.org

DOI Handbook
2002 http://www.doi.org/handbook_2002/index.html

Dowling, T.
2001 "One Step at a Time"; NetConnect; page 36.

Dublin Core Metadata Initiative
2002 http://dublincore.org/

Ehlers, H.
1994 "Identification Numbering in the Book, Library and Information World". ISBN Review 15 pp. 89-214.

European Molecular Biology Organisation (EMBO), E-BioSci project
2002 http://www.e-biosci.org/

Getty Images
2002 http://creative.gettyimages.com/source/home/home.asp

IDF Announcements
2001 <indecs>2rdd adopted as MPEG-21 baseline technology. http://www.doi.org/news/020114-DRM.html#consortium.
2001 Major organizations to develop digital rights management (DRM) standard. http://www.doi.org/news/011101-DRM.html

Lessig, L.
1999 "Code and Other Laws of Cyberspace"; Basic Books.

Lillington, K.
2001 "Sentries at the gate"; Guardian. http://www.guardian.co.uk/Archive/Article/0,4273,4322913,00.html

MPEG-21 Digital Item Identification and Description (DII&D)
2001 Committee Draft; (approved), ISO/IEC JTC 1/SC 29/WG 11/N4532. http://mpeg.telecomitalialab.com/working_documents.htm

NISO Technical Metadata for Digital Still Images, Standards committee AU
2001 http://www.niso.org/committees/committee_au.html

ONIX Product Information Standards
2001 http://www.editeur.org/onix.html

Paskin, N.
1999 "Toward Unique Identifiers"; *Proceedings of the IEEE;* 87 (No.7); pp. 1208-1227. http://www.ieee.org/organizations/pubs/pub_preview/PROC/87proc07_toc.html

2000a "E-Citations: actionable identifiers and scholarly referencing"; Learned Publishing Volume 13, pp. 159-166. http://www.catchword.com/alpsp/09531513/v13n3/contp1-1.htm)

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

2002 "Proceedings of ICSTI Seminar:Symposium on Digital Preservation of the Record of Science - 14/15 February 2002", IOS Press, Amsterdam, the Netherlands. http://www.doi.org/topics/020210_CSTI.pdf

Picture Agency Council of America (PACA)
2002 http://www.stockindustry.org/

Uniform Code Council Website (UCC)
2002 http://www.uc-council.org

Rosenblatt, B.; Trippe, B.; Mooney, S.
2002 "Digital Rights Management: Business and Technology"; Hungry Minds; ISBN 0764548891.

Rust, G.
1998 "Metadata: The Right Approach, An Integrated Model for Descriptive and Rights Metadata in E-commerce"; D-Lib Magazine. http://www.dlib.org/dlib/july98/rust/07rust.html

Rust, G.; Bide, M.
2000a The <indecs> Metadata Framework: Principles, model and data dictionary. http://www.indecs.org/project.htm#finalDocs

2000b indecs Summary Final Report. http://www.indecs.org/project.htm#finalDocs

Sidman, D.; Davidson, T.
2001 "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).

Software & Information Industry Association (SIIA), Content Division Digital Rights Management Working Group
2001 "The Digital Object Identifier: The Keystone for Digital Rights Management". http://www.siia.net/divisions/content/doi.pdf

Sowa, J. F.
2000 "Knowledge Representation: Logical, Philosophical and Computational Foundations"; Brooks/Cole.

Sundt, C. L.
1997 "The CONFU Digital Image and Multimedia Guidelines: The Consequences for Libraries and Educators". http://oregon.uoregon.edu/~csundt/indy.htm

Svenonius, E.
2000 "The Intellectual Foundation of Information Organization"; MIT Press.

Technical Association for the Graphic Arts web site (TAGA)
2002 http://www.taga.org/

Trant, J.
1996 "Exploring New Models for Administering Intellectual Property"; The Museum Educational Site Licensing (MESL) Project. http://www.archimuse.com/papers/jt.illinois.html

Visual Creators Index: White Paper
1999 http://www.vci-uk.com/white/vci.PDF

XML Cover Pages: DIG35: Metadata Standard for Digital Images
2001 http://xml.coverpages.org/dig35.html

Zorich, Diane
1998 "Introduction to Managing Digital Access: Options for Cultural and Educational Organizations"; Getty Information Institute; ISBN 0-89236-546-3.