

## **Digital Object Identifier: implementing a standard digital identifier as the key to effective digital rights management.**

**Norman Paskin**  
**Director, The International DOI Foundation**

April 2000    © 2000 The International DOI Foundation

### **Introduction: the Digital Object Identifier (DOI) initiative**

The future of the content industry is now clearly tied to the future of the Internet and its successors, requiring a fundamental shift from physical to electronic dissemination. All current varieties of intellectual property (e.g. books, music, journals, video, software), and those types yet to be invented, will be involved in this migration, and the boundaries that exist between current content types, especially at the level of supporting infrastructure, will be eliminated. Instead of (or in addition to) 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 problems and opportunities in delivering their content to their consumers.

One of the key issues in the move from physical to electronic distribution of content is the evolution of a common technical and procedural infrastructure to identify, or name, pieces of content in a digital environment. A common and well understood approach to referencing objects is essential to the evolution of services required for content providers to grow and prosper in the era of increasingly sophisticated computer networking. The referencing system must also work with existing systems for naming tangible and intangible intellectual property entities.

In 1998 the International DOI Foundation (IDF) was established to address this challenge, assuming a leadership role in the development of a multi-tiered framework of infrastructure, policies and procedures to support multinational, multi-community content provider requirements. Membership in the IDF is open to all organisations (for-profit or not-for-profit) with an interest in electronic publishing and its related enabling technologies. Major components of the IDF mission involve stimulating interest in and understanding of this framework, encouraging alliances and collaborative activities designed to explore in depth the complex issues to be addressed, and influencing the development of standards which would ensure the appropriate level of value-added and quality control across the spectrum of participation. The main activity of the IDF is to encourage the implementation and use of a standard digital identifier: the *Digital Object Identifier* (DOI).

The IDF was an early supporter of the INDECS (Interoperability of Data in E-Commerce) initiative, providing financial support and promoting the concept of well-formed interoperable metadata. It has used indecs as its metadata model.

### **The fundamentals of the DOI system**

The DOI system can be defined as "A unique resolvable identifier and multiple pieces of associated typed state data in an information management substrate". To explain the pieces of that definition further:

- A "unique identifier": DOI assignment is unique to a piece of intellectual property. The definition of that piece is specified by some key pieces of information about it (metadata) which depend - in detail though not in concept - on the particular genre: whether the entity is an article or a video clip, for example. The identifier is an opaque string; it does not contain any syntactic intelligence about the entity.

- "resolvable"; with "associated state data" : Routing via a proven internet system from that identifier to one or more pieces of associated data. These represent the current state (value) of some specific type of data (e.g. a URL). Those pieces of data may either represent, or direct to, services using DOI as an entry point.
- "an information management substrate": once a piece of data has been obtained by resolution, metadata about the identified entity is able to interoperate with metadata from other sources (e.g. about context) to construct automated services and transactions. This interoperability is achieved through managing the metadata in a controlled way, conforming to a generic interoperable architecture which enables DOI to offer applications beyond a simple persistent identifier.

The DOI system achieves this by a mix of technical implementation (using existing technologies) and specific policies as to that implementation (e.g. what a DOI can be assigned to, and who can assign it).

- The technical tools used by the DOI system include the Handle system<sup>®</sup> for resolution, and the indecs framework for describing intellectual property entities and ensuring metadata interoperability.
- The policies and overall governance are determined by the International DOI Foundation (IDF). The IDF is a not-for-profit organisation supported by members from a wide range of content and technology organisations

### The business case for DOI

The Digital Object Identifier (DOI) has a number of practical advantages:

1. It identifies the entity, not the location. There are important consequences:
  - If the location changes, the DOI stays the same (persistence);
  - The same entity can be identified and accessed at multiple locations;
  - In the World Wide Web digital environment, this overcomes the inability of browsers to deal with Names. URLs alone cannot deal with the necessary expression of *relationships* and *multiple instances*.
2. It is an open system, anyone can use it, and it is free at the point of use. Anyone encountering a DOI can click on it and be linked to services. The cost of assigning a DOI (which is very small) is paid when the content owner assigns it.
  - Anyone can build systems which integrate DOI usage into local environments - e.g. to access locally licensed copies of an entity
3. It is a fully managed system
  - DOI assignment via registration agencies, conforming to overall DOI policies, assures uniqueness
  - Routing is via a proven internet system to any associated data or services designated by the content owner/ DOI assigner, or to any services provided by others using DOI as an entry point
  - Provides some key information about the item identified (metadata).

These three points are the definition of what is often called an "actionable identifier"; such identifiers are not just a location tool but serve to open a transaction of some form.

4. DOI can be applied to any form of intellectual property, at any (arbitrary) level of granularity
  - The key information about the item identified, and its place within the managed information substrate, makes this possible, while still allowing the system to be used predictably
  - Because it has such a wide applicability, it does not replace existing standard identifier systems (such as ISBN, ISSN, ISRC, etc) but it can work with them where these are applicable and make them actionable. It can therefore become but a unifying identifier system to enable interoperability.

- It is applicable in any digital environment; the conceptual framework is at a level of abstraction not specific to one environment. Initial implementations are WWW based because of the current dominance of this environment in intellectual property concerns, but DOI could be applied to any future Internet protocols or environments (which are likely to arise just as the Web arose only a few years ago on the basis of the underlying Internet architecture<sup>1</sup>)
5. The associated information (metadata) is managed in a controlled way; therefore it can express *relationships*; and it can work with other metadata from other sources to construct *services and transactions* (by conforming to a standard way of using metadata - e.g. by expressing the context in which a piece of intellectual property is being used). This enables the construction of predictable, automated transactions.
6. The DOI system is *extensible*. Further developments are possible in an open standards environment. For example, the Handle resolution system has the capability for trusted transactions; the indics framework provides the basis for a full rights management architecture. Each of these functions is essential to fully automated and efficient intellectual property commerce.
7. DOI is not limited to current environments. It is neither restricted to digital entities, nor to the World Wide Web applications which are likely to form the obvious early implementations. The concepts are designed to be usable in any future information environments.

### **The business background: rights management**

Whilst it is recognised that internet commerce has been markedly successful in facilitating physical trade, it is clear that if digital commerce of digital property is to grow dramatically, a number of issues need to be dealt with by establishing underlying consistent standards for transactions, especially in recognising the treatment of intangibles<sup>2</sup>. Whether or not the present economic growth of E-commerce continues, there will be a need to put such standards in place. If all the current E-commerce is not sustained in a move to intellectual property transactions, there will be a re-evaluation followed by a more realistic assessment of the tools needed to put an open automated system in place. Copyright still matters; technology can adapt to enable it to serve its function.<sup>3</sup> The architecture evolved for the Internet hardens the boundaries between "use" and "prevention"; whilst it is easier to copy illegally, it also becomes easier to implement access prevention measures. Lessig<sup>4</sup> has argued that in addition to copyright, a content owner should think of "copyduty" - the need to make access possible. Shapiro & Varian point out<sup>5</sup> that a content owners aim is to *maximise value* of their information objects (not to minimise copy infringement): that is, to achieve the optimal strategy of reducing copy infringement but increasing accessibility. There are a number of ways that this can be achieved (e.g differential pricing); but fundamental to the transaction is identifying what it is you wish to control (allow access to).

There is a movement from *mass production* (one single product sold to many customers) to *mass customisation* (building individual products to order), which brings a need to identify components (and also the terms under which they are available). Component availability has business benefits in allowing multiple re-use of entities in different products, sometimes called "recombinant publishing".

### **The technical background: the Internet**

A recent paper reviewing the present status and prospects for the Internet<sup>6</sup>, written by two of its pioneers, notes: "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." Brad Cox makes a similar point in *Superdistribution*<sup>7</sup>, using an analogy with the settlement of the American West: initial penetration by communication (Pony Express), followed by later penetration by actual transport (freight railroads).

One of the current problems of WWW usage is that the majority of applications are via one single interface, the web browser, which has limited functionality. Given the likely future proliferation of possible devices, services, and protocols, Web browsers will no longer be the sole platform for Internet-enabled applications, except when they are targeted purely at casual users. Frequent users will need an optimized interface that takes full advantage of the device they are using (and what they are attempting to achieve)<sup>8</sup>. Efficiency is not likely to be maximised with a single design across all problem sets, so there may be multiple sector specific software tools in which general Handle/DOI capability will be useful. Even if we accept that browsers are the unavoidable tool of today, it is clear that there is a problem: current browsers deal only with locations (URLs); they currently do not deal transparently with identifiers of names. Specifications and procedures for URNs (Uniform Resource Names), having some of the desirable characteristics noted in the "Business Case" section above, have been defined but are not implemented in common browsers, despite being available for several years and implemented in several early experimental systems<sup>9</sup>. There is active debate within the W3C community as to whether the URN concept should be supported, or superseded by Uniform Resource Identifiers (URIs) which are essentially stable URLs. The DOI uses a technology (Handle) which can be used with browsers or with any other appropriate future interfaces, and can work with any defined URN or URI syntax.

## Identifiers

The point of including a unique resolvable identifier and multiple pieces of associated typed state data in an information management substrate is that such an identifier is needed *to track and use objects in a dynamic network*. There is no other way to accommodate the differing uses and relationships of entities but to put in a complex level of indirection. The Handle system<sup>®</sup> (which is used by the DOI system) aims to be the optimum way to carry that indirection. It is a tool for persistence, multiple resolution, scalability, security and authentication, distributed administration, and internationalisation, with a light weight resolution method and an open protocol.

Because we as individuals have one name we may be deluded into thinking that every entity has "one unique name". In practice of course this is not true. We name entities according to functional purposes in a variety of contexts: sometimes I refer to my laptop as a Toshiba 3110; sometimes I refer to it by its serial number; the two have markedly different purposes, each valid.

## Persistence

Identifiers can be used to link an identified entity to its location. However, a problem with using identifiers digitally is ensuring that they are persistent. Two general approaches are being suggested to making identifiers persistent on the Internet:

- Don't change the URL. Design the URL to be an unchanging persistent label for the resource at a maintained location. Accept that persistence is a social issue.
- Assign a name to the entity, and use redirection: accept that URLs may change, so assign a name ("URN") (which does not change) and a mechanism of resolving this to the URL (purl.org and redirect.net are examples of such services).

There are ongoing debates about which approach is best. A reader doesn't care which of these methods underlies any system he uses, provided it works. What is inhibiting this from happening is that there are currently problems with both methods, so *neither* works. People do change URLs, and there may be legitimate reasons for such changes (e.g., the sale of a journal from one publisher to another; archiving;

etc.). There may legitimately be multiple URLs corresponding to one entity. Book-marking in browsers uses not the reference "mid point" (the name) but the end point - the URL one ends at; when this changes, the book-marked link is broken. (Neither of these exhausts the lists of difficulties with each approach, but they are sufficient to make either unusable for consistent reliable services). The Handle system used by DOI adopts the approach of using names, and multiple redirection. Even with Handle there are some implementation issues, though these are not insurmountable. For example, bookmarking of DOIs is supported in Internet Explorer 5 (right click on a DOI, select 'add to favourites' from the pop up menu, and the DOI will get recorded as e.g. doi:10.1000/79), bookmarking the entity and not the page that contained it, which is otherwise the default bookmark behaviour. But none of this is standards based; it is an artefact of specific design of specific browser peripheral tools. Having standards-compliant supporting tools is a great help for widespread deployment, but the incentive to create the tools is low without widespread deployment, which may imply that we (IDF and CNRI) have to make those tools available i.e. do some of the basic seeding development or encourage others to do so.

### *Instances, classes and other relationships*

We recognise that it can be useful to group items as "the same" for some purposes, at some times. For example, all copies of the printed US edition of a book (many thousands) are treated for some purposes as one "Book ". When I refer to that book, page 43, I refer to the set of all copies. But when I actually want to read it, I need to get a particular instance (a copy); for example the copy bearing my own name and signed by the author is a specific instance of the edition; no one identifier can serve all purposes. Since sometimes we need to group things in one way, and at other times in another way (sometimes we want to identify an instance; sometimes a class), we cannot use just one simple number. Instead we must use a network of identifiers; and the relationships between them (this identified entity is an instance of that identified class, etc) are expressed by links - that is, metadata about the entity (well-formed metadata is no more than a network of identifiers, of metadata entities). In the digital world there are already many cases where the same article is legitimately available from more than one content provider, or at mirror sites: i.e. at more than one location. This is analogous to the case of multiple copies of a book (multiple instances of a class). I may find a citation (to an intangible work, or set of all manifestations) and need to find one specific instance among many possible ones. There may be many copies for legitimate reasons; yet all are the "same". The functionality of specifying an instance from the identifier of the entity - the "appropriate copy problem" - is currently being developed further by, amongst others, the Digital Library Federation in conjunction with the IDF<sup>10</sup>.

With intellectual property, relationships can be much more complex than simply class/instance. One piece of intellectual property may have many relationships to others (excerpt, modification, compilation, performance, etc). A single identified "manifestation" can embody an intangible work: when that same work has more than one manifestation, or a manifestation is related to several works, it becomes vital to clearly understand and express what is being identified. Take as an example a scholarly research article; in print-only publishing, the publication process resulted in a single published entity (a printed manifestation). With digital publishing, the production line branches at the end of the editorial process to produce multiple entities (e.g. a printed article, and an on-line PDF or HTML file). Those entities are related: they are "the same article"; the number that is on the production line item and carries forward into both the published entities (telling that they have something in common) is an identifier of the intangible "abstraction". Identifiers of abstractions are typically used for *citations*; identifiers of manifestations are typically used for *transactions* (there are exceptions to each: e.g. analysis of manuscript variants, and sales of rights, respectively). For a detailed exposition of this issue, see the DOI-X prototype usage of DOIs<sup>11</sup> and especially the discussion of the analytic framework<sup>12</sup> of "E-Citations").

### *Identifiers, recognition and metadata*

As the example above shows, it is not sufficient to assign an identifier to an entity without clearly expressing what that entity is. An identifier assigned to "a book" or "a paper" is not useful, since it is not clear what is meant by that term (an instance? a manifestation? an abstraction?). It is necessary to declare

that "I am using this identifier to refer to the paper in the form of its abstraction defined as follows:". In the digital world, that expression must be done by assigning metadata. In an interoperable digital world, that metadata must be well-formed (unambiguous), and follow a data model to enable relationships to be expressed<sup>13</sup>.

When an identifier is used, the associated metadata will be brought into play. One key use for the metadata is to confirm that the entity is the one requested, by providing some readily recognisable piece of data: it is not possible for the identifier itself to carry all the appropriate descriptive information<sup>14</sup>, and entities may not be directly inspected either for practical reasons or because they are by their nature not directly perceivable (abstractions). For example, when an ISBN is used it is helpful if associated data such as "Title" of the book can be returned as a confirmation step. Many standard identifiers have such metadata associated with them in practice, though this may not necessarily be in an interoperable form. There is now discussion (e.g. within ISO TC46) as to whether mandatory metadata standards for identifiers should be specified, to ensure this identification and recognition.

This important point leads to the principle of identification and recognition: "an identifier must declare sufficient metadata to unambiguously define the entity which it identifies, and to enable the entity to be recognised from the metadata".

### *Not just digital*

An early decision taken in the DOI's development was that a DOI may be applied to any intellectual property entity, not just to digital objects<sup>15</sup>. This is essential if we are to deal with the full range of intellectual property relationships. E-Commerce of intellectual property involves transacting intellectual property entities, or the rights to intellectual property entities. The commonly used phrase "Digital Rights Management" in fact implies both *Management of Digital Rights* and *Digital Management of Rights*<sup>16</sup>. If DOI is to be used for rights management, it must therefore imply both *Identification of Digital Objects* and *Digital Identification of Objects*.

The entities identified by DOIs will be those which are useful for some practical purpose and of significant persistence (e.g. digital manifestations, books, abstractions). However there are many related entities generated unavoidably in the digital world (some of which are transitory, though that term is relative: the distinction may be important in for example legal considerations such as whether a cached entity is a copy). "The same abstract forms may be represented or embodied in many different physical entities. Therefore, physical objects of widely divergent natures could be characterised by the name of the same abstract form. The name of the book *War and Peace*, for example, could refer to an abstract form conceived by Tolstoy or to an embodiment of that form in a physical object made of paper and ink. When computers are used to represent such things, the number of entities, both physical and abstract, is multiplied:

- When *War and Peace* is encoded for computer processing, it becomes a pattern of bits, which is another abstract form that hardly resembles the one that Tolstoy conceived.
- When a computer program formats the bits to recreate a humanly readable copy, the abstract pattern passes through a rapid succession of physical embodiments: reflecting spots on one disk, magnetic spots on another disk, currents flowing in transistors, pulses of light in a laser beam, electrically charged spots on a drum, and dust particles that are attracted to the drum and baked on the paper.
- Despite the profound differences between the physical embodiments, they could all be called by the name of the same abstract form, *War and Peace*. For different purposes, the same physical entity could also be described by different forms: the bound volume could be called *War and Peace* to emphasise its content, or it might be called 'a book' to emphasise its physical structure.

William of Ockham admonished philosophers to avoid multiplying entities, but computers multiply them faster than his razor can shave"<sup>17</sup>.

### Resolution: the Handle system

The resolution mechanism used by the DOI is the Handle System<sup>®</sup>. It is logically possible to conceive of using other resolution systems for DOI resolution. Handle was chosen for a number of reasons, of both fundamental design and practical implementation; its advantages include:

- Scalability
- Reliability
- Resolution speed
- Flexibility in configuration
- Ad hoc data typing
- 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

In addition, the originators of the Handle System, the Corporation for National Research Initiatives<sup>18</sup>, have considerable expertise and practical experience in many digital library and related activities which benefits us in considering future development strategies. Current technical challenges or improvements being responded to by CNRI for the future development of Handle include:

- Locality:
  - Optimize selection among parallel choices - a performance issue
  - Provide a local answer to a global question - an organisational issue
- Administrative flexibility
  - Configuration for rapid administration (many writes to few reads)
  - Use of the Handle System as a Public Key Infrastructure
  - Multi-master – allow administrative actions at any point in a distributed system
- Decentralization
  - Heterogeneous name spaces
  - Ad hoc interoperability among differing resolution systems

Of particular interest for commercial uses, the latest release of the Handle technology includes two features suitable for trusted transactions:

- While there is some overhead involved if you want to utilise this feature, the new system has been built to allow ad hoc *trusted resolution* across the entire distributed system..
- Distributed trusted administration. The permissions are at the handle (DOI) level, not at the server level. Again, this is especially useful for a distributed system. If you have administrative privileges on a given DOI, you don't have to worry about where or how many or who even owns or runs the servers.

Whilst Handle is not yet as ubiquitous as DNS, the cost structure is identical to better. The root service (the Global Handle Server) is free for use, naming authorities are free, and servers and client library are free. Only services and policy layers need support, at least at the moment.

The Handle System allows handles to be both created and resolved in a distributed fashion. Both creation and resolution can be accomplished using dedicated clients, common clients such as web browsers using special extensions or plug-ins, or unextended clients going through various proxies. In all cases, communication with the Handle System is carried out using Handle System protocols and in all cases those protocols have both a formal specification and some specific implementations, all freely available from CNRI. The resolution protocol has a corresponding API available as a library of C functions. This library, also known as the client library, has been used by CNRI in the creation of an http-to-handle proxy server, handle-aware extensions to the Netscape and Microsoft web browsers, a handle caching server, and in various prototype projects.

In the absence of ubiquitous support for the Handle protocols in standard browsers, there are two routes to achieve the functionality:

- The CNRI **Handle Resolver** is a free, downloadable web browser extension that is installed once and then run behind-the-scenes, like a Plug-In. Similar to the way that a plug-in enables a web browser to encounter the string "<http://www.somesite.org/paper.pdf>" and recognise the file at that location as PDF and properly display it, the Handle Resolver enables a web browser to encounter the string "doi:10.1000/85", and know to use the Resolver to recognise the string as a DOI, retrieve the URL associated with that DOI in the DOI System, and return the resulting location or file to the browser. The CNRI Handle System Resolver works with Netscape (version 3.0 and higher) and Internet Explorer (version 3.0 and higher) running under Windows 95, NT 3.5 and 4.0, and Mac OS 68040 and PowerPC.
- **Proxy Servers.** Without the Resolver installed, a web browser that encounters "doi:10.1000/85" will not know how to respond, and can only proceed if the DOI is combined with the address of a DOI "proxy server". A proxy server is a special server that understands how to communicate with both a web browser and the DOI System. For example, the string "<http://dx.doi.org/10.1000/85>", constructed by combining the address of a DOI proxy server with the DOI 10.1000/85, will direct the web browser to the proxy server, which in turn knows how to contact the DOI System, resolve the DOI in the string, and return the associated location or file to the browser. CNRI maintains two public proxy servers addressed by "hdl.handle.net" and "dx.doi.org". The extension enables browsers to respond to the URNs "hdl:" and "doi:" in the same way it responds to "<http://>". The syntax in use today for proxy server resolution of a DOI is "<http://dx.doi.org/10.1000/1>". The syntax for resolving that DOI using the web browser extension would be "doi:10.1000/1" or "hdl:10.1000/1".

There are advantages to downloading and using the plug-in, for native Handle protocols, to resolve DOIs, instead of relying on a proxy server. These include *performance* and, more fundamental in the long term, *functionality*. Proxy server resolution requires that the web browser connect to the proxy server, and is affected by network response time and the availability of the server itself.; the proxy introduces another layer of redirection (communication back and forth across the net) which can slow resolution. It is important to start to use the resolver because more *functionality* can (and in the future will) be included in the Resolver that won't be in the proxy; for example, a prototype has been created to use various data types in the format of associated services. Such services are an important potential use of DOIs<sup>19,20</sup>. Such extra functionality is more easily built into a client software (incorporating the Handle resolver) than by using a proxy. The next version of the resolver plug-in (to be released shortly) will improve resolution of DOIs which have *multiple URLs* (e.g., there are five URLs associated with the DOI for each D-Lib Magazine article, one for each of the five mirror locations plus the primary site) by dynamically determining which URL will provide the fastest response, before sending that URL to the web browser. The proxy server will only randomly choose one URL to send to the web browser. It is anticipated that more functionality, perhaps in the area of services such as rights management, could be built into future versions of the resolver plug-in. The resolver plug-in is frequently upgraded, and registered users are automatically notified of the availability of new versions.

### **Intellectual property management: the metadata framework**

Why do we need to associate metadata with a DOI? *If a DOI can apply to anything, we must be able to say what it identifies.* Consider the wider implications of having such a persistent identifier. If the DOI is seen as "just" a single-point resolution routing system (from a DOI occurrence to a single URL at a publisher-maintained web-site which is the sole source of the entity), then there is little need for interoperable metadata, as any related information about the object so identified can be held within that "proprietary" web site. The need for interoperability arises in the wider vision of DOIs as persistent names (identifiers) for content entities, and which can be used openly; that is, independent of the source of that content (as noted earlier, restricting all uses of material to routing via the publishers web site is neither possible nor productive). Metadata is essential to commerce as it must be possible to process transactions via unique identifiers without recourse to physical inspection of the items being traded, which may be either inconvenient or impossible. Multiple-resolution schemes for names (such as URNs) necessitate the existence of some basic metadata associated with the resolution process, in order to provide parameters for an intelligent selection process. Once the DOI is seen as a potential unifying identifier which can be used to access seamlessly a variety of entities from unrelated sources, and to access different instances of the same

resource at different sources (using multiple-resolution, local resolution, and building services dependent on the persistent identifier), then the need for a common vocabulary and data model for DOI metadata is clear.

Metadata associated (predictably and logically) with each identified entity is essential if a DOI is to be used as anything other than a simple routing to a single defined location. In considering the logical framework necessary design the associated metadata scheme to guarantee interoperability of DOIs, a number of requirements became clear. The metadata must be capable of describing any Creation, or intellectual property entity; rights descriptions (for transactions and services); use existing identifier schemes where these are established; and be capable of full automation (which requires no ambiguity; the metadata must be well-formed. Fortunately, the development of the DOI coincided with the publication of just such a set of principles, from the indecs project. We used this logical framework; the alternative was an ad-hoc or "informed guesswork" approach driven by the need of specific applications with the hope that other, yet undefined, applications will find the same set of metadata terms equally useful. The top level principles that indecs identified are:

- *Unique identification*: every entity needs to be uniquely identified within an identified namespace;
- *Functional granularity*: an entity needs to be identified only when there is a reason to distinguish it;
- *Designated authority*: the author of metadata must be securely identified;
- *Application independence*: metadata structures should be independent of any technology platform;
- *Appropriate access*: everyone will need access to the metadata on which they depend (appropriate privacy is a necessary counterpoint).

The indecs framework is:

- Generic for all types of content: convergence renders differentiation meaningless at a transaction level;
- Focussed on intellectual property management, but does not enforce one specific model of management;
- Enabling, not replacing, other schemes, thereby creating an interface;
- Broad in scope, covering identification, description, and transaction;
- Based on tested real world models. The concepts of the fundamental data model are built on earlier work in the library sector (IFLA's Functional Requirements of Bibliographic Records) and the music sector (the Common Information System).

We found indecs to be a powerful analysis tool in early DOI prototype work, in analysing the distinction between one work and several formats for citation reference linking (the "DOI-X" prototype). It has produced practical results which are being used in the commercial sector in both the EPICS standard for the book and serial sector (Editeur Product Information Communication Standard) and the audiovisual sector (Muze's MerchEnt system), which will use a common XML expression and dictionary. DOI will endorse this common set of expressed terms as the basis for definition of DOI metadata; the coverage of the books, serials and audiovisual sector provides a sound basis for initial use and future extension in the DOI system.

This framework provides a means to define a set of metadata about each identified entity, available to users and conforming to an overall defined architecture which enables DOI applications offering functionality beyond a simple persistent identifier. To ensure that metadata about the identified entity can interoperate with metadata from other sources (e.g. about context) to construct services and transactions, this information is managed in a controlled way to provide a basis for interoperable applications by conforming to a generic interoperable metadata framework.

In the course of 1999 a DOI information management layer was specified, derived from the indecs generic framework. Because of IDF's participation as a sponsor of indecs we were able to use the generic data model, mapping, and schema activities immediately to define kernel and specific genre metadata sets (e.g. in the DOI-X prototype). The overall architecture is now clear, though some implementation details still

need to be documented. Indecs documentation is now being completed: a common set of terms and a multimedia data interchange schema (<indecs>XML expression) covering a wide range of content types will be available shortly. This schema will be supported by a common multi-media dictionary derived from the indecs model and its mapping to existing schemas, and an indecs registry (a metadata registry for mapping other schemas and terms to the indecs model).

Related projects which now need further support include:

- work on a Directory of Parties;
- A project to extend <indecs>xml and <indecs> terms to encompass descriptions of offers, agreements and statements of rights;
- A legal namespace metadata dictionary;

All of these will be useful additions for wider DOI use in E-commerce, but they are not required for initial implementations and developments. Because of the close relationship between practical DOI implementations and practical indecs implementations, the DOI and indecs initiatives share common concerns and opportunities for future development and promotion. We are currently considering how we might bring some aspects of governance and development together to make a more effective common alliance.

### **Essential simplicity, and power, of the DOI**

DOIs are built on a foundation of two technical tools: a resolution system and a metadata framework. Each of these tools provides tremendous flexibility, since they provide very simple conceptual foundations which enable wide applicability:

- The Handle resolution system allows a DOI to be linked to any piece of related data (more precisely, a current state value).
- The Metadata framework allows a DOI to identify any form of intellectual property, at any arbitrary level of granularity.

This flexibility produces a system of great potential power. The inevitable price is that this power also represents a danger; unless carefully managed, misuse of these fundamental abilities could lead to chaos - if anything is possible, nothing is predictable. The policies which IDF is building on top of these technical foundations are designed to set management constraints and ensure that the system is predictable and useful. For example:

- Handle resolution is controlled by a registration process which will conform to policies laid down by the IDF;
- A DOI must declare sufficient metadata to unambiguously define the entity which it identifies, and to enable the entity to be recognised from the metadata

Some of the more attractive options for potential application of DOI, such as for example resolution of a DOI to a local (or "appropriate") copy, will need careful analysis and clear implementation rules to avoid confusion.

Existing identifiers are more limited in scope, which makes them easier to deal with: an ISBN has a context which largely prevents its inappropriate use. However the necessity of well thought out and implemented policies and implementation is apparent even in such well established identifier schemes as advances in technology push the boundaries of existing policies: for example a number of E-book creator/distributors are assigning their own ISBNs to publishers' books that they have converted for distribution. That is, the book still belongs to the publisher, and the E-book creator is being treated as a distributor, but the ISBN comes from the E-book creator. This is like a bookseller assigning its own ISBNs on every title publishers ship to them; this will create some interesting havoc in the short term.

## Metadata implementation issues

In practical implementations of the full functionality, entities identified by a DOI will also need declaration of a minimal set of metadata, to enable systems to understand what is being identified.

1. Metadata has value. Declaring some common elements of metadata, and making these "freely available" (at least on the basis of a known-item search, i.e. given the identifier, what is the associated metadata) is still seen as a business hurdle in some areas. Some companies generate income from sales of metadata and there is concern that such a step will cannibalise this business. In response it should be noted that:

- DOI resolution to declared metadata provides known-item look up, but not the more widely required look up from a search to identifier (resource discovery); that functionality is the role of applications using DOIs, not the DOI system itself. An example of such an application is the recently announced CrossRef initiative<sup>21</sup>, which provides a service to link reference citations of articles to source items, using DOIs as the linking identifier of the cited work.
- The metadata is not required to be held in the DOI system; the DOI must simply provide a pointer to its declaration
- The provision of some freely available metadata is a means of maximising value of information by promoting access while not granting full access without some further step. It is analogous to a publishers catalogue, and is simply one possibility at the beginning of a spectrum which can range through making small portions of a piece available (e.g. an abstract) to making a number of sample available in complete form. There is no single model which fits all forms of property, but the minimal step of making simple metadata available is a lowest common denominator.

2. Making metadata available is also a technical hurdle; how is it declared in practise? Some options (which are not mutually exclusive) could include:

- <meta> tags in html. Easy to do but relatively unstructured and providing no interoperable framework;
- XML (Extensible Markup Language) declarations: offering a logical syntactic framework, and increasingly widely supported, but interoperability is not guaranteed. If the right connections "run through" an XML expression to indecs it will enable a move towards interoperability later without losing short-term investment;
- RDF (Resource Description Framework): offering a further level of semantic expression, but currently not developed to the point where widespread easy implementation is available. RDF appears to be in the long run a good solution for the business of making metadata interoperable, but those whose immediate concern is with achieving short-term practical improvement within existing business operations need to be convinced that this is the way to go now;
- a separate database (the initial DOI-X and possibly CrossRef model): providing interoperability only within a specific environment
- a pointer record in the DOI Handle to a location: this is an attractive option and is being considered as a compulsory requirement for DOIs, since it offers a platform-independent means of providing the information.

3. What metadata elements must be defined for declaration? We have adopted the indecs framework as a top level means of definition, but it is clear that this is appropriate for constructing mappings and deriving sub sets, not immediately as a list of required elements. We recognise that no single "grand unified data model" will be adopted by all communities, since "all classifications of the world are arbitrary"<sup>22</sup>; but some are better than others, and useful ones must be mappable and become interoperable. Appropriate specific data elements will differ between intellectual property types; a measure of extent such as "number of pages" is appropriate for one type, but another measure such as "duration" may be more appropriate for another. What is needed is defined elements (building blocks) one can mix and match in a controlled, managed way giving interoperability. We have chosen to go as close as possible to the indecs model by defining an overall framework for a DOI "Genre" (type of intellectual property) to be expressed

in specific terms appropriate for that type and declared with each DOI in future, full implementations of the system:

*DOI*: key record

*DOI Genre*: a class of entities with common characteristics

*Identifier*: a unique identifier e.g. from a legacy scheme applied to the entity

*Title*

*Type*: primary structural type: e.g. work, manifestation

*Origination*: e.g. original, derivative, replica

*Primary Agent*: name or identifier of primary agent

*Agent role*: role of primary agent

An extended description of this genre set has been published in an earlier paper<sup>23</sup>; we now need, and intend, to publish a more detailed description and set of guidelines for this set, to enable users to create DOI Genres within the indecs framework.

We must also recognise that whilst flexibility and variety must be allowable, an infinite number of genres built from our component data element "blocks" is not useful. This also raises a practical business issue: who pays to validate a series of "guidelines" or community standards (genres) and how many should there be?

4. If there are rules, there will be mis-use of these rules. How should this be minimised? In our early initial implementations of DOI (using the simple single redirection to one URL, without any associated metadata (what we have begun to call the "zero genre") we have already seen examples of mis-use:

- one content owner using a co-publishers prefix, leading to attempted double use of one DOI;
- a content owner declaring DOIs without entering them into the Handle directory (the DOI is therefore "dead" and not actionable!)
- Authors using DOIs as citations without clearly understanding what they identify (because in a zero genre this is not stated)

Who will police these incorrect usages and errors? We are building in appropriate checks into our Doi registration processes, but cannot control misuse outside the system, which requires a more fundamental process of education.

5. Therefore there is a need for education or "missionary work" in the identifier/metadata/rights management arena:

- A need to market the concepts and get the general concepts known to a wider audience;
- A need to ensure that all metadata intended for use in automated systems is well-formed. There is still a need to get this message across. Even though there is now little problem in convincing people of the cost of ambiguity in other areas (e.g. the Y2K problem), and there is recognition that implementation of a standard office technology project would be impossible without a logical underlying model (entity/relationship schemes, etc), this thinking is not always transferred to thinking about intellectual property rights - perhaps because this has traditionally not been an area where automation has been strong.
- A need for understanding some of the key principles of identification, and especially the key concepts of names and levels of indirection, which are at the heart of the issue of identifiers used on the Web: URLs and URIs don't provide the necessary levels of indirection to provide management capability at the level of intellectual property entities.

There are encouraging signs that over the last year, thanks in large part to the indecs initiative, awareness of these issues is growing. But much more needs to be done: who will pay for this? We can note but not emulate some other standards development promotions: "Having some large visible customers aboard can get the bandwagon rolling by directly building up critical mass. In November 1997 Sun took out full-page ads in the New York Times and other major newspapers reciting the long list of the members of the "Java coalition" to convey the impression that Java was the "next big thing".<sup>24</sup> We hope that some major implementations of DOI and indecs work will lead to positive feedback and encourage wider implementations, plus the development of complementary tools and uses.

6. Implementing DOI and indecs principles also means implementing several underlying standards or developing standards, some of which themselves require further funding. Some of these standards activities have a champion or home such as XML, RDF etc (though even the WWW Consortium has resource limitations); other aspects more specific to the intellectual property community have no funding or support structure; there is no "Intellectual Property Engineering Task Force" equivalent to the IETF for technical infrastructure. Activities like pre-competitive further development of metadata standards (indecs), and infrastructure development (e.g. Handle) are in need of more resource and wider recognition. The question of how we raise the funding for this is one which concerns the IDF and the indecs partners.

### **Business implementation issues**

In addition to these issues related to metadata implementation, there are some specific issues relating to the wider business aspects:

7. Who does the work to make the DOI metadata a good fit to the DOI Genre definitions, and ensure it is of good quality? How do they get paid? Who is the authority for each data element?

8. In developing an IDF business model, we have reviewed analogous systems such as the physical bar code system and defined the concept of Registration Agencies. If there are several possible Registration Agencies, how do they relate to each other?

9. Some entities identified by DOIs will also be identified by legacy systems (e.g. ISBNs, ISSNs, etc), which raises a number of issues. Those legacy systems may also be collecting metadata. Should the metadata be collected only once and agreement made to share, and if so on what basis? (this may of course not be possible, for commercial reasons, in some cases). If the metadata collected by the legacy system is not mapped to the DOI indecs model, or (more problematic) cannot be easily mapped, what steps should be taken to bring the two into line?

An IDF paper on "DOI Deployment" (see below) outlines the "vision" and desired development of the DOI initiative; it provides further details of how we get from where we are now (an active development activity) to where we want to be (a fully implemented, widespread, mature system). This process is characterised as *deployment* of the DOI. This paper represents a summary of work in progress. (The paper is an evolution of an earlier version 1.0 (October 1999) which was circulated only to International DOI Foundation members, and a few interested parties. Version 2.0 incorporates some suggested minor changes made by readers of the first version, and updates progress in several areas). The following is a condensed summary of the business issues arising in deployment.

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

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

The IDF intends that:

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

In order to recognise the fact that some charging for DOIs will be inevitable, the IDF initiated a simple economic model (currently in use) of charging \$1000 for allocation of a prefix (a one-off charge) allocated to one publisher, allowing an unlimited number of DOIs to be constructed using that prefix. The IDF now wishes to encourage a migration to a *wide variety of potential business models*, in recognition of the fact that such a simple model is not a "one size fits all" solution. Business models could include, but not be limited to, cost recovery via direct charging based on prefix allocation, numbers of DOIs allocated, numbers of DOIs resolved, volume discounts, usage discounts, stepped charges, or any mix of these; indirect charging via cross subsidy from other value added services, agreed links, etc. The IDF wishes to place minimal constraints on the business models offered by RA's, and to enter into discussion on practical implementation of these. In this vision, a Registration Agency would be free to establish a completely different pricing model: the business model of each RA is determined by the RA itself. The RA enters into agreements with its customers (DOI assigners); these may be radically different between RAs, depending on the needs of its customer communities, its own other business, the value-added services it offers, competitive pressures, etc. The agreements will however require conformance to the overall federal standards.

We also want to encourage a similar migration to a variety of *naming authority models*. Initially we implemented a model of prefixes as hierarchical naming authorities, which allowed local creation of individual DOIs by adding suffixes. Prefixes are now at the level of "content owner" (e.g. publishing house); but we should encourage migration to whatever level of granularity is the best fit for a particular community. For example, in reference linking in the scholarly community, it may be that prefixes at the level of "journal" are a better architecture. This is linked to the business model, by the prices set for individual prefixes (e.g. \$1000 per prefix is inappropriate at the journal level).

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

The RA offers its own business model to customers; it also enters into a contractual agreement with the IDF (potentially in future migrating to an Operating Federation consisting of all the RAs), guaranteeing the conformance to minimal criteria covering conformance to technical, information management, and economic criteria. The "economic criteria" for being a member of the Federation is a payment to support central Federation governance: a "participation fee", or "franchise fee".

There is however another set of relationships which needs to be considered, namely the relationships 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 will go to the local one. 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 specialised 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 favoured; whereas in smaller language markets, a geographical (or at least linguistic) basis may be more appropriate.

The IDF will be considering the best way to appoint and work with the initial likely registration agencies. The aim of the IDF will be to stimulate the widest use of the DOI system, by fostering a climate in which working as a registration agency is attractive, yet any long term monopoly is discouraged. Initial incentives could include subsidising the Participation Fee from membership funds, etc. The desired end result, however, is to have an Operating Federation which is self-sustaining.

## Standards issues

"As Internet access and services are provided by existing media such as telephone, cable and broadcast, interactions with standards bodies and legal structures formed to deal with these media will become an increasingly complex matter. The intertwining of interests is simultaneously fascinating and complicated, and has increased the need for thoughtful co-operation among many interested parties"<sup>25</sup>.

The DOI development process can be described as having three parallel tracks: the initial implementation (a single redirection tool for persistent naming); the full implementation (adding structured metadata and using multiple resolution); and standards tracking. There is a strong practical reason for this work with standards. DOI and indecs have each opted for the creation of open standards. Network economics and positive feedback make cooperation in this way more important than ever. Standards promote a larger number of users of technologies conforming to the standard, by enhancing compatibility, reducing uncertainty and fear of being "locked in", and shifting efforts from competing for the market to competing within the market. It is true that standards development can be frustratingly slow, but the benefits are worthwhile: when the value of a product to one user depends on how many other users there are, economists say that this product exhibits network externalities, or network effects. Technologies subject to strong network effects tend to exhibit long lead times followed by explosive growth.<sup>26</sup>

We can add to the nine issues already identified for implementation a tenth relating to standards:

10. Who manages an open standard once it is accepted and successful? "Managing successful open standards can be especially tricky. Truly open standards face two fundamental threats. First, if there is no clear sponsor, who will be in charge of setting the direction in which the standard will evolve? Will the standard stagnate, or will crippling compatibilities arise, since no-one can exert control? Second, without a sponsor, who will invest the resources to make improvements and thus keep the standard from stagnating?"<sup>27</sup>

One of the aims of the IDF is to act as a unifying open standard forum, which can achieve widespread support and prevent a futile splintering into competing standards which in the extreme can prevent the very step forward they are aiming to achieve (see the story of AM Stereo standards cited by Shapiro & Varian: "Never heard of AM stereo? Our point exactly")<sup>28</sup>. We are working actively as a member or liaison body with many standards groups, including:

- W3C (World Wide Web Consortium), especially URI and RDF activities
- NISO (National Information Standards Organisation)
- MPEG (ISO/IEC JTC1/SC29/WG11) - both MPEG7 "Multimedia content description interface" and MPEG 21 "Multimedia Framework"
- ISO TC46 (international Standards Organisation/ Information and Documentation)
- IETF (Internet Engineering Task Force)

As DOI activities reach a level appropriate for formal standardisation we are putting these into the formal process: e.g. the DOI syntax will become a NISO standard, Z39.84. We are also liaising closely with any related activities such as SDMI (Secure Digital Music Initiative), CIDF (Content ID Forum), National Bibliography Numbers (NBN) projects of libraries, etc; although many of these are focussing on one specific area of concern, it becomes clear that there is convergence and much to be gained by working together as they share many of the same problems and opportunities. We also have strong relationships with bodies which influence these activities, whilst not being formal standards bodies themselves, such as WIPO (World Intellectual Property Organisation), GBD (Global Business Dialogue), ICC (International Chambers of Commerce), etc. Our participation in the indecs project also brings liaison with many bodies too numerous to list here; (see [www.indecs.org](http://www.indecs.org) list of "connections").

## Intellectual property rights issues of the DOI system

The policy of the IDF is that DOIs must be freely resolvable. The IDF is reviewing the current situation of its rights protection in the interests of ensuring system integrity: DOI and DOI.ORG are already registered service marks. The DOI system has some inherent costs (registration, resolution, governance) which need

to be covered; we have outlined a possible way of doing this, by means of Registration agencies - exactly analogous to other schemes such as the physical bar code system.

Why are we reviewing IP issues? A helpful analogy might be to consider why I can't say "I accept VISA payments", create a number that follows the VISA syntax (which is a public specification; individual VISA bank identifiers are readily available on the Net), and process credit card payments through my own bank, charging retailers a fee and not paying VISA International? The answer is that the VISA system is trademarked and/or patent protected and/or enjoys other protection measures, as a means of protecting its interoperability, reputation, and the quality of the overall system, and to enable VISA International (a not-for-profit) to recover its operating costs. Exactly the same would apply in the case of someone setting out to sell physical bar codes outside the existing UCC/EAN system, and in many other areas. The aim of establishing protection is "defensive". The process of review we are now undertaking is exactly what happened with the Imprimatur Music Trial, a good example of an open development in the area of content, and we are looking at that as an example. DOI is already a trademark but we need to know if that is enough to protect the integrity of the system. Clearly another group "setting its own rules" and calling them DOIs would be disastrous for consistent operation of the system. Additional measures of securing integrity are likely to be covered in contractual arrangements made with registration agencies and, through them, with content owners assigning DOIs.

For those who think this precautionary work is overly cautious, reality tells us otherwise. We have already seen evidence of possible "cybersquatters" trying to set up and/or sell us DOI-like domain names; "squatting" the business model is a next step. The aim of the IDF is to create a system which can be used by everyone with credibility, reliability and integrity. We need to take appropriate steps to ensure this by reviewing and if necessary improving our existing protection measures.

### **Issues and plans for the future**

In putting a standard digital identifier in place, the IDF is necessarily dealing with many issues. These include practical business issues of implementation (Registration Agencies etc), but also the wider aim of promoting the advantages of interoperable standards for identifiers and metadata, to aid the development of an improved infrastructure: to encourage the tide to enter the harbour and raise all the boats to everyone's advantage.

The Foundation plans for 2000 include several areas of work which we feel will be of benefit to the initiative and also to the development of the area of Intellectual Property commerce in general. The following list is not exhaustive and we may modify the list as developments unfold:

#### *Marketing and external relations*

- Recruitment and retention of Foundation members
- Development of new standard presentation tools, a CD-ROM/video stand-alone demo, and additional meetings and speaking engagements. A "simplified" guide and articles.
- Running workshops in US and Europe for DOI registrants
- Running information/training seminars in US and Europe

#### *Technical*

- Review of DOI Genre metadata proposals and conformance testing against indecs-compliant schema
  - Development of XML and possibly RDF schema for each approved DOI Genre
  - Mappings of DOI genres to other important schema e.g. text, images, music areas
  - Development of "proof of concept" software tools for services using multiple resolution
  - Development of Java or similar tools for multiple resolution capability .
- (Note that in the long term, activities such as the first three of these items will probably be carried out by DOI Registration Agencies, but in 2000 we assume IDF will need to carry these out .)

*Legal and related*

- Patents, trademark and other IP issues to ensure integrity of the system is guaranteed.

*Standards activities*

- Participation in all relevant activities; of particular interest at present are MPEG 21 (new multimedia framework activity); W3C interest groups etc (e.g. URIs, RDF); NISO and ISO meetings; monitoring related activities.

*Technical writing*

- preparing a "DOI Handbook" version 1 (Web and print format), the complete guide to what DOI is, accompanied by DOI syntax, Metadata and genre guidelines, scope and usage guidelines, etc.
- preparing materials for DOI user workshops, training materials, and materials for further adaptation by Registration Agencies

*Outreach*

A positive effort to attract companies and develop applications (either prototype or full) in market segments of significant size where we currently have little or no activity.

**Funding and governance**

The existing IDF organisation follows the W3C model: the funding is from Members, and there is no direct relation to the operational running costs of the DOI system. The member-based IDF currently subsidises all the operational running of the system specific to the DOI implementation of Handle technology. The members of IDF, as with members of W3C, pay a membership fee to support development of the system as a pre-competitive standards activity, which if implemented would enable costs savings or new business opportunities in the community (analogy: allowing the tide to enter the yacht harbour, enabling each yacht to float free but with no advantage to any one). The IDF is run by its Members, via an elected Board and appointed Director. The aim however in the long term is to reduce reliance of the DOI system on this membership fee, enabling the fee to be reduced, abolished, or redirected (as the Membership decrees), and seek income directly related to DOI system usage. As a necessary step in promoting wider take up of DOI we are also considering how a "DOI Development Forum" or similar grouping of practical implementers could be created, financed, and managed.

Indecs, by contrast, has been run as a fixed term development project, with funding from a variety of partners but no continuing governance structure beyond the first proposed deliverables. As indecs comes to a close (at least of its first phase), the IDF has an interest in ensuring that there is continuing support for activities such as maintenance of the data model, schemas, registry of mappings, etc. It is also clear that indecs could benefit from a grouping of developers. A possible way forward for indecs is also the creation of a member-based organisation and structure similar to that of IDF. Both require the encouragement of practical application development, through the creation of a developers forum or similar bodies, which will initially require active stimulation. The likely overlap in both scope and market for such an organisational model has prompted both parties to consider whether some elements of their future management could be taken forward more advantageously through a joint undertaking. This is under active discussion at present.

**Conclusion**

The future of the content industries is now clearly tied to the future of the Internet, and its successors. One of the key issues in this move to electronic distribution of data is infrastructure to put names on pieces of content in a standard and consistent way. This is essential to ensure it's possible to know what a piece of content is and what can be done with it. Only by doing this will different vendor systems be able to communicate with and understand each other.

Two years ago the International DOI Foundation was established to address this challenge, and to put such a system in place. We've made a lot of progress, but I know a lot more needs to happen to provide a framework for automated content transactions.

Please join us in making this happen.

### **Related documentation**

#### *On DOI:*

Two documents provide additional material. The first Annual Review of the Foundation, published in July 1999, provides information on the aims and governance of the Foundation, and outlines issues of scope and metadata framework. It is available at the DOI web site (<http://www.doi.org>) as well as in printed form distributed at this conference. It includes an article on DOI Current Status and Outlook first published in D-Lib magazine<sup>29</sup>. Further discussion on aspects of DOI deployment (how we get from where we are now -an active development activity- to where we want to be -a fully implemented, widespread, mature system) is carried in a discussion paper available on the DOI web site as "DOI Deployment"<sup>30</sup>

#### *On the indecs metadata framework:*

The indecs web site at [www.indecs.org](http://www.indecs.org) carries latest work in progress. Documentation of many aspects of the project, including an XML expression and terms for an initial wide-ranging dictionary, is being prepared. A good starting point is the document on "The <indecs>metadata schema: Building Blocks" prepared for the indecs evaluation conference in November 1999 and available from the indecs web site (printed copies are also distributed at this conference)

#### *On the Handle resolution technology:*

The web site at [www.handle.net](http://www.handle.net) provides comprehensive documentation on the Handle resolution system, ranging from introductory guides to the full technical specification of the protocols.

---

### **References**

<sup>1</sup> What is the Internet (And What makes it Work). Robert E. Kahn and Vinton G. Cerf. Internet Policy Institute, December 1999. [http://www.internetpolicy.org/briefing/12\\_99\\_story.html](http://www.internetpolicy.org/briefing/12_99_story.html)

<sup>2</sup> Blown to Bits: How the New Economics of Information Transforms Strategy. Philip Evans & Thomas S. Wurster. Harvard Business School Press, 2000

<sup>3</sup> see e.g. The Soft Edge: A Natural History and Future of the Information Revolution. Paul Levinson. Routledge, 1997

<sup>4</sup> Code and Other Laws of Cyberspace. Lawrence Lessig; Basic Books, 1999

<sup>5</sup> Information Rules: A Strategic Guide to the Network Economy. Carl Shapiro and Hal R. Varian; Harvard Business School Press, 1999

<sup>6</sup> Kahn and Cerf, op. cit.

- 
- <sup>7</sup> Superdistribution: Objects as Property on the Electronic Frontier. Brad Cox. Addison-Wesley, 1996
- <sup>8</sup> Jakob Nielsen's Alertbox, 6 Feb 2000 <http://www.useit.com/alertbox/20000206.html>
- <sup>9</sup> see e.g. Uniform Resource Names: a progress report. The URN implementers. D-Lib Magazine, February 1996. <http://www.dlib.org/dlib/february96/02arms.html>
- <sup>10</sup> Choosing the Appropriate Copy: P. Caplan & D. Flecker.: Digital Library Federation Architecture Committee. <http://www.niso.org/DLFarch>
- <sup>11</sup> Reference Linking with DOIs: a Case Study. Atkins et al <http://www.dlib.org/dlib/february00/02risher.html>
- <sup>12</sup> E-Citations : actionable identifiers and scholarly referencing. Norman Paskin. Available on the DOI web site ([www.doi.org](http://www.doi.org)). To be published in "Learned Publishing", 2000
- <sup>13</sup> Metadata: The Right Approach. An Integrated Model for Descriptive and Rights Metadata in E-commerce Godfrey Rust. D-Lib Magazine July/August 1998 <http://www.dlib.org/dlib/july98/rust/07rust.html>
- <sup>14</sup> Toward Unique Identifiers. Norman Paskin. Proceedings of the IEEE, 87 (no.7) July 1999 pp. 1208-1227
- <sup>15</sup> DOI: Current Status and Outlook Norman Paskin D-Lib magazine, May 1999 <http://www.dlib.org/dlib/may99/05paskin.html>
- <sup>16</sup> Digital Rights Management Systems: Technology Update. Glen M. Secor. Paper presented at the conference "Digital Rights management and Digital Distribution for Publishing", New York, February 2000.
- <sup>17</sup> Knowledge Representation: Logical, Philosophical and Computational Foundations. John F. Sowa. Brooks/Cole, 2000
- <sup>18</sup> <http://www.cnri.reston.va.us>
- <sup>19</sup> The role of metadata supply chains in DOI-based, value-added services. John Erickson ICSTI Forum, No.30, April 1999
- <sup>20</sup> The DOI and Rights Management: Tying up Loose Ends. John Erickson. Trialogue, No 11, summer 1999. <http://www.ybp.com/yrm/trialogue/1199/1199doi.htm>
- <sup>21</sup> CrossRef: [www.crossref.org](http://www.crossref.org)
- <sup>22</sup> Sowa (op.cit.) ; and for an eloquent expression of this concept see the essay John Wilkins' The Analytical Language, Jorge Luis Borges (1942); translated in e.g. Borges: Selected Non-Fictions (ed. E. Weinberger), Viking, New York, 1999.
- <sup>23</sup> DOI: Current Status and Outlook Norman Paskin D-Lib magazine, May 1999 <http://www.dlib.org/dlib/may99/05paskin.html>
- <sup>24</sup> Shapiro & Varian, op.cit., chapter 1
- <sup>25</sup> Kahn & Cerf, op.cit.

<sup>26</sup> Shapiro & Varian, op. cit.

<sup>27</sup> Shapiro & Varian, op. cit., chapter 8

<sup>28</sup> Shapiro & Varian, op.cit., chapter 9

<sup>29</sup> DOI: Current Status and Outlook Norman Paskin op. cit

<sup>30</sup> DOI Deployment IDF Discussion paper:  
<http://www.doi.org/deployment2.pdf>