

\*\*\*ARCHIVED DOCUMENT\*\*\*

This report has been superseded by DOI 10.1000/6382-1, October 99

## DOIs used for reference linking

Version 1.0  
Date: 16 March, 1999  
Author: Norman Paskin  
Status: For comment/additions  
Circulation: Open  
File: C:\WPWIN\WPDOCS\DOI\reflinksNP2.doc

### Analysis of citation reference linking problem

See appendix one for detailed analysis of the problem and definitions of terminology used throughout this paper.<sup>1</sup> The analysis shows the need for two steps:

- Step 1: From a citation to a work
- Step 2: From a work to a manifestation

The current form of citation usually **declares** attributes of a physical manifestation (volume, issue, page, etc.), though **these are** increasingly likely to be attributes of a digital manifestation (file format, location, etc). The linkage to a work requires a number which identifies a Work irrespective of manifestation. Therefore the system requires:

- Identification and metadata of the manifestation(s)
- Identification and metadata of the underlying work
- Recognition that work and manifestation need not have 1:1 correspondence

The following analysis deals only with the issue of simple linkage. It proposes (*Appendix 2*) a syntactic convention which would further simplify the application (but is not itself essential to implementing DOI-based linkage).

*Appendix 3* outlines the further possible compatible developments which would be possible using DOI technologies, providing seamless solutions for resolution from multiple manifestation possibilities, local resolution, and linkage to other media forms.

### Principles of proposed solution

Information is collected on DOI registration sufficient to construct DOI metadata for *at minimum* two related entities:

- DOI of the Work
- DOI of the manifestation which displays the citation (usually, but not necessarily, the print manifestation)
- A link is possible between these (shared data elements or explicit declaration)

---

<sup>1</sup> Analysis presented to the DLF/NFAIS/NISO/SSP reference linking mail list as a result of the discussion at the workshop of the same name held in Washington, Feb 11, 1999.

A database is constructed from this metadata. This would be an Application database for this particular set of DOIs (Genre). With these in place:

- reference linkage between any two DOI registered works can be made
- citations can be converted to the corresponding work identifier

The work identifier may then resolve to a service, location, or DOI of a manifestation.

From the point of view of registrants:

- The registrant need not submit two separate DOI inputs, or even know that two DOIs are being constructed: a consolidated set of metadata for any article is collected and "de-constructed" by the registration agency. (This may assist early inexperienced registrants)

From the point of view of users:

- A look up from citation to Work identifier (step 1) could occur within the DOI system, using the Application database, or in external A&I systems carrying equivalent data;
- A resolution to one or more manifestations (step 2) is possible from the DOI of the Work, if the registrant so chooses;
- The DOI of the Work would be recommended as the *declared identifier in a citation*

### Metadata components of proposed solution

Table 5 summarizes how all the elements needed to complete Tables 2-4 can be provided in a single registration.

**Note that registrants and users of the proposed reference linking application need ONLY USE TABLE 5.**

The other tables (1-4) are shown here for clarity and explanation but users need not be concerned with these.

There is a *DOI kernel* metadata set which applies to anything which is given a DOI. That proposed kernel is shown in Table 1, and is explained in more detail than in the earlier DOI Metadata discussion paper..

The proposed extended kernels for the three types of Journal Article (Work, Print and Digital) are shown in Tables 2, 3 and 4 respectively.

### Mechanism

Metadata as specified in table 5 to be provided to a central collection point in XML. IDF will create a simple XML DTD in a standard format to accomplish this.

### Other metadata

This requires provision solely of the small set of metadata described here, and nothing more.

Any other metadata which the registrant may have and normally make available (e.g. to A&I services) is not collected by the IDF linkage database and the registrant (publisher) is free to continue existing arrangements.

**DOI Genre: Journal Article**

A proposed definition of the scope of this Genre is : *a textual article , with or without accompanying illustrations, tables or photographs, published in a journal in print or digital form.* If necessary, this genre definition can later be amended to take into account multimedia, component identification, etc. However, the main aim of the present proposal is to accommodate the current universe of reference-linked articles.

**The "Journal Article" Genre kernel**

In order to deal with the real complexities of component creations and multiple contributors, the approach set out in the metadata paper relies on grouping together certain attributes as *Events*. However, because of the fixed and limited design of the DOI kernel, it is NOT necessary to recognise events in this sense to complete the kernel. All items of data can be unambiguously shown simply as direct attributes of the journal article itself, and this makes for a straightforward element set. This set will also be compatible with any future development for this Genre, or for other DOI Genres based on the proposed DOI/INDECS approach.

**Table 1: DOI kernel elements**

These are the kernel elements which apply to *all* DOI Genres, shown here for completeness:

<i>Element</i>	<i>Definition</i>	<i>Status</i>	<i>Number</i>	<i>Value and Qualifications</i>	<i>Qualifications</i>
<b>DOI</b>	A DOI	Mandatory	1 only	DOI	This is the key for this record
<b>DOI Genre</b>	A class of resources with common characteristics defined by the IDF community	Mandatory	1 minimum	From DOI Genre tables	
<b>Identifier</b>	A unique identifier (e.g. from a legacy scheme) applied to the resources	Qualified	1 minimum	String	Mandatory <i>unless</i> the DOI Genre extension rules specifically exclude it
<b>Title</b>	A name by which the resource is known	Qualified	1 minimum	String	Mandatory <i>unless</i> the DOI Genre extension rules specifically exclude it
<b>Type</b>	The primary structural type of the resource	Mandatory	1 only	From: <i>Work</i> <i>Physical Manifestation</i> <i>Digital Manifestation</i> <i>Performance</i>	
<b>Origination</b>	The process by which the resource is made	Mandatory	1 minimum	From: <i>Original</i> <i>Derivation</i> <i>Excerpt</i> <i>Compilation</i> <i>Replica</i>	
<b>Primary Agent</b>	The name or identifier of the primary agent(s)	Mandatory	All	Identifier or Name from an agreed Genre namespace	The specification of Primary Agent (normally but not necessarily the creator) is determined in the extension rules for each DOI Genre.
<b>Agent Role</b>	The role(s) played by the primary agent(s)	Mandatory	1 minimum	Role from an agreed Genre namespace	

These are supplemented by a group of “Registration” elements, which covers the administrative metadata needed for the kernel declaration for “Level 1” DOIs.

<b>Metadata Registrant</b>	The identity of the person or organisation registering the DOI	Mandatory	1 only		
<b>URL resolution</b>	The current web location of the resource	Mandatory	1 only	URL	
<b>Metadata record version</b>	The version number of the metadata	Mandatory	1 only	Number	
<b>Date of metadata declaration</b>	The date on which this declaration was made by the registrant	Mandatory	1 only	YYYYMMDD	

**Table 2: DOI kernel plus extensions for DOI Genre=Journal Article, Type=Work**

These are the kernel elements required for a *work* in this genre:

<i>Kernel</i>	<i>Extension</i>	<i>Value</i>	<i>Summary</i>
<b>DOI</b>	<b>DOI of Work</b>	<i>DOI</i>	Key
<b>DOI Genre</b>		<i>“Journal Article”</i>	Fixed value for this Genre
<b>Identifier</b>	<b>Work Identifier</b>	<i>PII or Proprietary ID</i>	PII must be provided if one has been allocated by Publisher, otherwise an identifier is not mandatory for this Genre.
	<b>DOI of first digital manifestation</b>	<i>DOI</i>	Mandatory if work is declared with a digital manifestation
	<b>DOI of first print manifestation</b>	<i>DOI</i>	Mandatory if work is declared with a print manifestation
<b>Title</b>	<b>Article Title</b>	<i>String</i>	Mandatory
	<b>Title of Journal of first manifestation</b>	<i>String</i>	Optional. The title of the journal in which the Article was first manifested
<b>Type</b>		<i>“Work”</i>	Fixed value for this Genre
<b>Origination</b>	<b>Work Origination</b>	<i>From: Original Derivation Excerpt Compilation</i>	One value is mandatory
<b>Agent Identifier*</b>	<b>First named author</b>	<i>Author Name</i>	Mandatory first author name
	<b>Additional authors</b>	<i>"et al"</i>	"et al" flag mandatory if >1 author
	<b>Publisher Name(s)</b>	<i>Publisher Name(s)</i>	All Publisher names
<b>Metadata registrant</b>		<i>Publisher Name(s)</i>	
<b>URL resolution</b>		<i>URL</i>	
<b>Metadata version</b>		<i>Number</i>	
<b>Date of submission</b>		<i>YYYYMMDD</i>	

The kernel elements of *Primary Agent* and *Agent Roles* can be combined here as the possible roles are defined in the extension.

**Table 3: DOI kernel plus extensions for DOI Genre=Journal Article, Type=Physical Manifestation**

These are the kernel elements required for a *physical manifestation* of a work in this Genre:

<i>Kernel</i>	<i>Extension</i>	<i>Value</i>	<i>Summary</i>
<b>DOI</b>	<b>DOI of print manifestation</b>	<i>DOI</i>	
<b>DOI Genre</b>		<i>"Journal Article"</i>	Fixed value for this Genre
<b>Identifier</b>	<b>Article Identifier</b>	<i>SICI</i>	SICI must be provided if one has been allocated by Publisher, otherwise an identifier is not mandatory for this Genre.
	<b>Journal identifier</b>	<i>ISSN, CODEN</i>	Minimum 1
	<b>Publication Date</b>	<i>YYYY</i>	Mandatory
	<b>Journal Volume No</b>	<i>Number</i>	Mandatory
	<b>Journal Issue No</b>	<i>String</i>	Mandatory
	<b>Article Start Page No</b>	<i>String</i>	Mandatory
	<b>Article Start Page Seq No</b>	<i>Number</i>	Mandatory (default=1)
<b>Title</b>	<b>DOI of Work</b>	<i>DOI</i>	Mandatory
	<b>Article Title</b>	<i>String</i>	Mandatory
<b>Type</b>	<b>Journal Title</b>	<i>String</i>	Mandatory
		<i>"Print Manifestation"</i>	Fixed value
<b>Origination</b>	<b>Manifestation Origination</b>	<i>From: Original Derivation Excerpt Compilation Replica</i>	One value is mandatory
<b>Agent Identifier*</b>	<b>First named author</b>	<i>Author Name(s)</i>	Mandatory first author name
	<b>Additional authors</b>	<i>"et al"</i>	"et al" flag mandatory if >1 author
	<b>Publisher Name(s)</b>	<i>Publisher Name(s)</i>	All Publisher names
<b>Metadata registrant</b>		<i>Registrant (Publisher) Name(s)</i>	
<b>URL resolution</b>		<i>URL</i>	
<b>Metadata version</b>		<i>Number</i>	
<b>Date of submission</b>		<i>YYYYMMDD</i>	

**Table 4: DOI kernel plus extensions for DOI Genre=Journal Article, Type=Digital Manifestation**

These are the kernel elements required for a *digital manifestation* of a work in this Genre:

<i>Kernel</i>	<i>Extension</i>	<i>Value</i>	<i>Summary</i>
<b>DOI</b>	<b>DOI of digital manifestation</b>	<i>DOI</i>	
<b>DOI Genre</b>		<i>"Journal Article"</i>	Fixed value for this Genre
<b>Identifier</b>	<b>Article Identifier</b>	<i>SICI</i>	SICI must be provided if one has been allocated by Publisher, otherwise an identifier is not mandatory for this Genre.
	<b>Journal identifier</b>	<i>ISSN, CODEN</i>	Minimum one of these
	<b>Publication Date</b>	<i>YYYY</i>	Mandatory
	<b>Journal Volume No</b>	<i>Number</i>	Mandatory
	<b>Journal Issue No</b>	<i>String</i>	Mandatory
	<b>Article Start Page No</b>	<i>String</i>	Mandatory
	<b>Article Start Page Seq No</b>	<i>Number</i>	Mandatory (default=1)
<b>Title</b>	<b>DOI of Work</b>	<i>DOI</i>	Mandatory
	<b>Article Title</b>	<i>String</i>	Mandatory
<b>Type</b>	<b>Journal Title</b>	<i>String</i>	Mandatory
		<i>"Digital Manifestation"</i>	Fixed value
<b>Origination</b>	<b>Digital format</b>	<i>MIME type etc</i>	Mandatory
	<b>Manifestation Origination</b>	<i>From: Original Derivation Excerpt Compilation Replica</i>	One value is mandatory
<b>Agent Identifier*</b>	<b>First named author</b>	<i>Author Name</i>	Mandatory first author name
	<b>Additional authors</b>	<i>"et al"</i>	"et al" flag mandatory if >1 author
	<b>Publisher Name(s)</b>	<i>Publisher Name(s)</i>	All Publisher names
<b>Metadata registrant</b>		<i>Publisher Name(s)</i>	
<b>URL resolution</b>		<i>URL</i>	
<b>Metadata version</b>		<i>Number</i>	
<b>Date of submission</b>		<i>YYYYMMDD</i>	

**Table 5: input metadata collection form for article registrant**

This table shows in a single form all the kernel metadata required for registration of a work and the first digital and physical manifestations in this Genre.

<b>DOI</b>	<b>DOI of work</b>	<i>DOI</i>	
	<b>DOI of first digital manifestation</b>	<i>DOI</i>	Mandatory if work is declared with a digital manifestation
	<b>DOI of first print manifestation</b>	<i>DOI</i>	Mandatory if work is declared with a print manifestation
<b>DOI Genre</b>		<i>"Journal Article"</i>	Fixed value for this Genre
<b>Identifier</b>	<b>Work Identifier</b>	<i>PII or Proprietary ID</i>	Provide PII if one has been allocated by Publisher, otherwise an identifier is not mandatory for this Genre.
	<b>Article Identifier</b>	<i>SICI</i>	Provide SICI if one has been allocated by Publisher, otherwise an identifier is not mandatory for this Genre.
	<b>Journal identifier</b>	<i>ISSN, CODEN</i>	Minimum 1
	<b>Publication Date</b>	<i>YYYY</i>	Mandatory
	<b>Journal Volume No</b>	<i>Number</i>	Mandatory
	<b>Journal Issue No</b>	<i>String</i>	Mandatory
	<b>Article Start Page No</b>	<i>String</i>	Mandatory
	<b>Article Start Page Seq No</b>	<i>Number</i>	Mandatory (default=1)
<b>Title</b>	<b>Article Title</b>	<i>String</i>	Mandatory
	<b>Journal Title</b>	<i>String</i>	The title of the journal in which the Article was first manifested
<b>Type</b>			Automatically derived from DOIs. Fixed values: see tables 2-4
	<b>Digital format</b>	<i>MIME type etc</i>	Mandatory if work is declared with a digital manifestation
<b>Origination</b>	<b>Manifestation (physical) Origination</b>	<i>From: Original Derivation Excerpt Compilation Replica</i>	One value is mandatory for Work and one for any manifestations declared.
	<b>Manifestation (digital) Origination</b>		
	<b>Work Origination</b>		
<b>Agent Identifier*</b>	<b>First named author</b>	<i>Author Name</i>	Mandatory first author name
	<b>Additional authors</b>	<i>"et al"</i>	Mandatory if >1 author
	<b>Publisher Name(s)</b>	<i>Publisher Name(s)</i>	All Publisher names
<b>Metadata registrant</b>		<i>Publisher Name(s)</i>	
<b>URL resolution</b>		<i>URL</i>	
<b>Metadata version</b>		<i>Number</i>	Sequentially updated, latest overwrites previous
<b>Date of submission</b>		<i>YYYYMMDD</i>	



## Notes on the input elements

These notes are provided for clarification and illustration of tables 1-5.

**Work Identifier:** if a publisher has issued a PII or another (proprietary) identifier for this work in his own system, it must be declared. This is a cross-check to guarantee that the publisher's own database retains an integral link with the DOI metadata, and also to support publisher systems where the DOI itself may not be a database field.

**Article Identifier:** if a publisher has issued a SICI or another (proprietary) identifier for this manifestation in his own system, it must be declared. This is a cross-check to guarantee that the publisher's own database retains an integral link with the DOI metadata, and also to support publisher systems where the DOI itself may not be a database field.

**Publication Date:** treated in this Genre as an aid to identification rather than a historic date of publication, in keeping with its normal use in citations (for example, the "December 1999" issue may actually be published in January 2000, or vice versa).

**Journal Issue Number:** this element may alphanumeric as it may include special issue names e.g. S1 (supplements etc.), therefore shown as *string* rather than *number* is chosen to allow for special numbering .

**Article Start Page Number:** this element may be alphanumeric, as above.

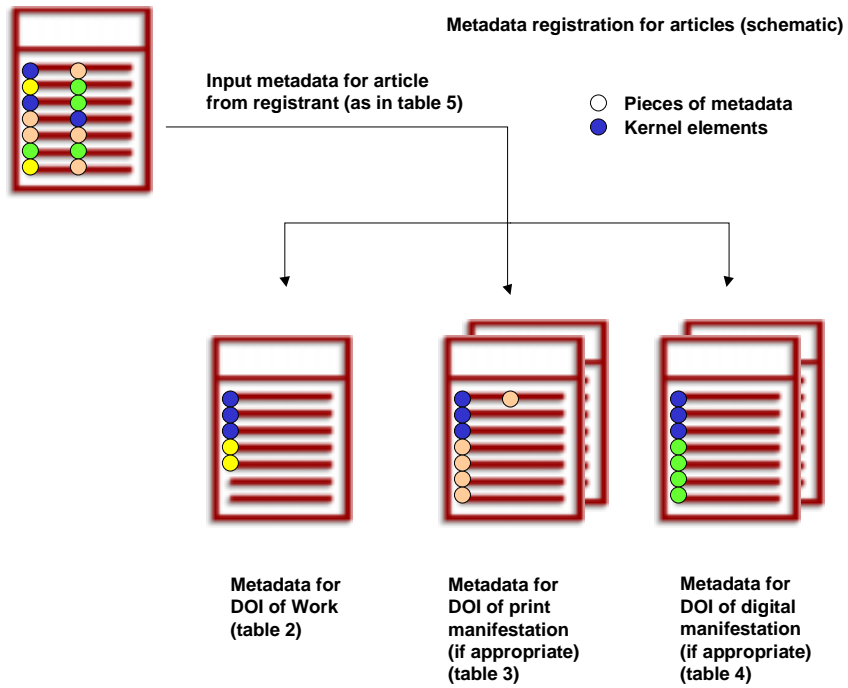
**Article Start Page Sequence Number:** This applies where more than one article appears on the same page of the journal. The sequence indicates where, in the normal sequence of reading, this article appears on the page. The default value is 1.

**Type:** In this Genre the primary types are denoted by *Work*, *Print (=Physical) Manifestation* and *Digital Manifestation*. The value *Performance* does not apply to this Genre.

**Origination:** Typically, the work and its related manifestations will have the same derivation. However, one manifestation may be a new *version* of another manifestation, while the work (which applies to both manifestations) is declared as an *original*, so separate values are required. The value *replica* cannot apply to a work. In this Genre the value *replica* applies to an exact copy of an existing manifestation, typically a local copy of a digital manifestation. See also appendix 3.

**Author Name/Additional Authors:** As this Genre is developed initially for the purposes of reference linking, only the first author name (for identification) is required. However, to support future extension, where other authors have contributed this must be declared by using the "*et al*" convention in the Additional Author field, indicating that additional information may be required to support future applications.

Figure 1: flow of information from registration to database



## Advantages

- Immediately implementable to achieve simple linkage between digital manifestations of a citing and cited article.
- Automatable by use of existing DOI system.
- Extensible as to resource type. The underlying mechanism and syntax are not restricted to articles. It allows linkage to "invocations" (Cronin et al, JASIS Dec 1998) of any type, and it allows for the future of what really happens in science: linkage to databases, web sites, software, list servers, resources, etc. The only requirement would be registration of a DOI for each.
- Does not require waiting for standardisation of work identifiers or digital file identifiers: "Names the nameless":
  - for those manifestations which commonly do not have standard assigned identifiers, such as electronic-only articles, the DOI provides a unique identifier
  - in the absence of an agreed standard for Work identification, the DOI of a Work provides a unique work identifier: DOI takes the place of an ISWC for text works.
- Resulting identifiers work with Handle DOI system and can be used instead of URLs, with increased persistence
- DOI registration of metadata allows extension for future uses and applications using the DOI as the unique identifier of a Creation
- Allows future extensions in a seamless fashion with planned DOI developments: see appendix 3

## Issues of note

- Rights considerations. This mechanism provides a linkage between references. It does not deal with the issue of authorisation or payment for such linkages i.e. when a publisher wishes to exercise rights in the cited article. However the existence of a mechanism with associated identifiers and metadata should encourage and facilitate the development of such authorisation mechanisms if required.
- DOI metadata database. This particular implementation of a DOI-based system illustrates the construction of a metadata database from the information registered with each DOI entity. This may not be the case for all DOI applications: the DOI requirement is probably that metadata for a DOI entity be accessible in a standard form, not necessarily at a central point. A commonly agreed means of declaring metadata for each entity in an accessible manner (e.g. as an RDF declaration/registry) would serve the same purpose (this is the subject of ongoing DOI work). For reference linking, there is no existing central comprehensive source of such metadata and so the construction of a database appears necessary and desirable to promote usage of DOIs. It is not obvious yet if this will be true for other DOI genres: e.g. DOIs allocated to musical works, manifestations, recordings etc. could make use of systems such as the CIS databases and may not require a separate DOI database.
- Multiple manifestations. Where multiple manifestations are published, the registrant should be encouraged to register all manifestations via DOIs, for completeness and to stimulate usage of the DOI service.
- Print and digital manifestations differ. Example: a digital "version" of the print paper which contains supplementary digital-only material. Analysis of this problem shows that these are in fact two separate works, one a component of the other, with consequences for manifestation identification. This is readily solvable in the DOI/INDECS schema, but the proposal won't solve this immediately. There is obvious work needed on this problem but it's outside the immediate scope of this paper. (The IDF has invited submission of "**Conundrums**" which will be used to test the existing draft schema and provide solutions

## Appendix 1: analysis of citation reference linking problem

### Introduction

The Feb 11, 1999 workshop on reference linking (NISO/NFAIS/SSP/DLF) noted the following problems about reference linking:

- “Multi-dimension” articles (multiple manifestations; no longer just print)
- Proliferation of citation types (a reference may be more than an article citation)
- Actionable links = URL are unreliable
- Bilateral agreements are too complex
- Automation of the process is needed

As a result of the workshop, it has been established that we have the following basic requirements:

- Identifiers for works
- Identifiers for digital manifestations
- Multiple-resolution mechanisms<sup>2</sup> for work identifiers
- Multiple-resolution mechanisms for digital manifestation identifiers
- A system which is extensible to other genres etc in the future, to allow other forms of links such as databases, software, etc. These need not be defined now, but the system should be designed to be extensible and not restricted to thus the discussion should not be solely guided by one implementation, e.g. the print article model with electronic text equivalents.
- Only the first two requirements are necessary for simple linkage between one digital manifestation of a citing and cited article.
- The other three requirements are necessary for a full solution incorporating the likely future requirements for journal publishing in the digital world (manifestation resolution, local resolution).

The rest of this appendix is an attempt to demonstrate the logic behind these requirements; it can be ignored if you already agree and support these propositions. It might be considered that this is "over-analyzing" the problem; however, in the DOI discussions we have found it necessary to have a rigorous analysis of the problem since otherwise it can be easy to fall into verbal traps and hidden assumptions, and the following analysis of the problem by the IDF is offered to others for comment and as a possible vocabulary to assist the group. Criticisms and corrections are welcome.

The IDF is of course also working on potential solutions to the problem, but these are not discussed here. However I note that:

- Only the first two requirements are necessary for simple linkage between one digital manifestation of a citing and cited article.
- The final two requirements are necessary for a full solution incorporating the likely future requirements for journal publishing in the digital world.

### Work and manifestation.

Not everyone agrees on definitions here (or even sees that the distinction is needed) so let's use an operational, functional definition. Traditionally, a publisher has a production line which is editing articles for publication, and the publisher will use a number to identify the entity he is processing. In print-only days, that process resulted in a single published entity (a printed manifestation). In digital days, the production line will bifurcate

---

<sup>2</sup> "multiple-resolution mechanism" means a mechanism which resolves one input to several outputs, or selects one of these outputs intelligently. It does not mean "several different resolution mechanisms".

at the end and produce two (or more) entities (e.g. a printed article, and an HTML file). Those two entities are related (they are "the same article"). The number which is on the production line item and carries forward into both the published entities (telling that they have something in common) is the identifier of the Work. (This was the origin/intent of the PII).

A more formal analysis would say the Work is an abstract Creation, made of concepts, whilst the published entities are examples of physical Creations (manifestations) and are made of atoms and/or bits. This is another way of getting to the same distinction, and it does not matter which you prefer from the point of view of this analysis.

### Conventions used

In order to achieve clarity it's first necessary to have a few conventions for the discussion. Consider one simple link and use the following definitions and symbols:

Entity 1 the citing article

Entity 2 the cited article (i.e., 2 appears in the reference list of 1)

Article: a generic term embracing "work and manifestations of the work" (Note that this alone is therefore insufficiently precise for a detailed analysis)

Use: **M** to indicate a manifestation entity (printed article, HTML file, etc)  
**W** to indicate a work entity (abstract underlying work)

Subscript: <sub>1</sub> to indicate the entity 1, the citing article  
<sub>2</sub> to indicate the entity 2, the cited article  
Example: M<sub>1</sub> is a manifestation of entity 1

Superscript <sup>P</sup> to indicate a printed manifestation  
<sup>D</sup> to indicate a digital manifestation  
Example: <sup>D</sup>M<sub>1</sub> indicates a digital manifestation of entity 1.

- If it is necessary to distinguish more than one digital or physical manifestation then the superscript is extended as <sup>D1</sup>, <sup>D2</sup>, <sup>D3</sup> etc  
Example: <sup>D1</sup>M<sub>1</sub> and <sup>D2</sup>M<sub>1</sub> are two different digital manifestations of the same article (e.g. a HTML file and a pdf file)

≡ The equivalent sign means "represent the same underlying work" and hence share the same work identifier, like a PII, in their metadata. In this terminology, the set of entities with such equivalence are the "article".

[ ] The identifier of the term bracketed, e.g.  
<sup>D</sup>M<sub>1</sub> indicates a digital manifestation of entity 1.  
[<sup>D</sup>M<sub>1</sub>] indicates the identifier of digital manifestation of entity 1.

Instance: a single copy of a manifestation

Reader: any user of the information.

### Case 1: the print-only world

There has been no need historically to separately identify works, because there is only one manifestation of each article. The concept of a separate work identifier has not therefore been necessary or introduced.

$$\begin{aligned} & {}^P M_1 \equiv W_1 \\ \text{and} \quad & {}^P M_2 \equiv W_2 \end{aligned}$$

Reader encounters  ${}^P M_1$  and sees the reference to 2 as a bibliographic string which can be "resolved" by consulting a library system either:

- directly ("which shelf is item 2 on?"), to then retrieve  ${}^P M_2$ ; or
- by constructing a SICI or similar, or using an identifier if given, as  $[{}^P M_2]$  which then goes into an on line system or similar to order an instance.

The process of "looking up the citation" is therefore a two step process:

- From the instance  $M_1$ , determine the identifier  $[{}^P M_2]$ , which is  $[{}^P M_2 \equiv W_2]$
- Resolve from  $[{}^P M_2]$  to an instance of  ${}^P M_2$

### Case 2: print plus one digital manifestation

Assume initially these exist in the case both of citing and cited article. However we assume here that we are starting from any  $M_1$ . In automated linking we are interested in the digital manifestations.

$$\begin{aligned} & {}^D M_1 \equiv {}^P M_1 \equiv W_1 \\ & {}^D M_2 \equiv {}^P M_2 \equiv W_2 \end{aligned}$$

Reader is accessing  ${}^D M_1$  and encounters the citation to  ${}^D M_2 \equiv W_2$

The citation is a bibliographic reference (possibly a URL which is an instance of  ${}^D M_2$ ).

- From the instance  ${}^D M_1$ , determine the identifier  $[W_2]$ 
  - Conclusion: we need identifiers for works<sup>3</sup>
- Resolve from  $[W_2]$  to  $[{}^D M_2]$  to a location instance of  ${}^D M_2$  i.e. a URL
  - In current practice commonly the only  $[{}^D M_2]$  is the URL (there is no separate URN for the digital object), which is not sufficient (does not enable local resolution, i.e. multiple instances)
  - If there is no  ${}^D M_2$  (e.g. a historical reference), default to  $[{}^P M_2]$  and deal with as in Case 1
- Conclusion: URNs for  ${}^D M$  are needed, i.e we need identifiers for digital manifestations (not just locations).

### Case 3: one print plus multiple digital manifestations

This is the model where an article appears in different digital forms such as pdf, html, etc (or future formats to be determined):

$$\begin{aligned} & {}^{D1} M_1 \equiv {}^{Dn} M_1 \equiv {}^P M_1 \equiv W_1 \\ & {}^{D1} M_2 \equiv {}^{Dn} M_2 \equiv {}^P M_1 \equiv W_2 \end{aligned}$$

As for case 2, but step 2 (Resolve from  $[W_2]$  to  $[{}^D M_2]$ ) is no longer a single point resolution. A multiple resolution system is needed (i.e., a system which can resolve from one input,  $[W_2]$ , to a set of outputs,  $[{}^{D1} M_2] \dots [{}^{Dn} M_2]$ ). The resolution would ideally be "intelligent" i.e. return the correct choice of digital

<sup>3</sup> PII has recently been introduced as a work identifier and is carried in some databases. Citation of PII as a work identifier  $[W_2]$  is possible in theory now with some publishers and some databases. Resolution of this requires a database which carries  $[W_2]$  and associates this with  $[{}^P M_2]$ :

- From  $M_1$ , determine the identifier  $[W_2]$ , which is  $[{}^P M_2 \equiv W_2]$
- Resolve from  $[W_2]$  to  $[{}^P M_2]$  to an instance of  ${}^P M_2$

manifestation depending on some criteria.

- Conclusion: we need multiple resolution mechanisms for work identifiers

(the same logical analysis applies to multiple print manifestations, e.g. an article in a journal later reprinted in a book, but the problem is usually left to cataloguers to sort out!)

### **Local resolution**

Irrespective of linking and digital manifestation identification, we need mechanisms to solve the "Harvard problem" or local resolution problem: I may not wish to resolve to a location that is a publishers site, but instead to a local copy (site license, cache, archive, intermediary, library, etc.).

Conclusion: we need multiple resolution mechanisms for digital manifestation identifiers

## Appendix 2: A syntactic convention which will simplify the use of DOIs.

The use of the DOI enforces a more rigorous (necessary) approach to the distinction of Work and Manifestations. Registering a text article in this way may result in three different DOIs: for Work, Physical manifestation, and Digital manifestation. As currently described, three dumb numbers would be produced. This may be confusing to registrants and users of the system. However introducing a syntactic convention as follows would immediately alleviate this problem:

**"The DOI suffix string will open with the string (\*), where (\*) is one of the following values of Type as defined in the DOI metadata schema<sup>4</sup>:**

- (W) Work: an abstract creation of the mind whose existence is revealed through expression as a performance or manifestation
- (P) Physical manifestation. A creation realised in physical form (atoms)
- (D) Digital Manifestation. A creation realised in digital form (bits)
- (R) **Performance**. A creation realised in space and time and recorded within a physical or digital manifestation that allows for its re-creation."

- On registration of an article, a DOI is created for the Work and simultaneously the same string generates for the DOI of the "first" digital and physical manifestations. Note that "first" is functionally defined: simply the one the publisher wants to call first.
  - Example:
 

10.1000/(W)131	the Work
10.1000/(D)131	the corresponding (first) digital manifestation
10.1000/(P)131	the corresponding (first) printed manifestation
- There may of course be multiple manifestations, either on registration or occurring later which will have DOI suffix strings not identical to the corresponding DOI Work identifier. But the main problem, of easy identification and registration and consequent linking article 1 to article 2, is solved. The connection of multiple digital manifestations is solved by automation (Work identifier resolution).
- The DOI of the Work would be recommended as the declared identifier in a citation
- The practical rules for dealing with versions, format versions etc, will be developed as best practice (the IDF will take an active role). The guiding principle is that of functional granularity: a creation should be distinguished separately (different DOI) only when there is a practical reason to do so.
- The DOI of the Work would be recommended as the declared identifier in a citation

The convention above has been proposed to the NISO committee working on the DOI syntax standard. The convention brings benefits to the immediate reference linking application, but also has more general benefits. Following are extracts from the justification of this proposal to the Syntax committee (full details available on request):

- **Metadata is the single step which is key to this proposal.**

There are two key issues about identification of an entity:

- identify it (assign name/identifier)
- identify what it is (assign key metadata)

At the inception of the syntax committee we had only the number to consider. We have now arrived at a

<sup>4</sup> exact wording to be reviewed; some wording in the syntax standard will tell where to go for more information on this: at present this is the DOI Metadata discussion paper, version 3, section 5.2.6; shortly it will be the DOI Handbook but at present that is only in preliminary draft; eventually we hope it will be reflected in a metadata schema standard derived from INDECS, Dublin Core 2.0, or similar.



fundamental structure which will underlie the future of the DOI (see DOI discussion paper on Metadata, version 3 February 1999). This is the single most important step since the DOI's conception.

- **But we agreed the number would be "dumb"**

No identifier is ever truly dumb. To be useful it must have or acquire some intelligence from its structure and/or context. Take for example a PII, explicitly stated to be a "meaningless" number: yet it is always encountered as PII:S1384107697000225, or in a field otherwise identified as "PII". That is, it has some minimal intelligence: we know it is a PII (a work identifier). Without that it could be anything or nothing. The question is not whether an identifier acquires intelligence in context, but whether that intelligence is wholly predicatable or not.

- **The DOI is fundamentally different in Type scope from other identifiers**

It will be true that any DOI will be equally intelligent: we will know that it is a DOI. But recall that the scope of a DOI is "any Creation": it may be a *Work, Manifestation, or Expression identifier*. No information identifier has ever before covered such a vast range: it is as though we have lumped together ISSN, ISBN, ISMN, ISRC, ISRN, ISAN and ISWC, calling them "ISXN" and deciding to have them all "equally dumb". One cannot tell anything about what the identifier relates to other than a "Creation". Even the UPC bar code at least says something about Type: everything it covers is a Physical Manifestation.

- **This is not a re-introduction of the previously rejected "bracketed type"**

Earlier, I and others argued for rejecting codes such as (SICI) etc in the DOI suffix string. This is *still correct*: such codes are arbitrary, not guaranteed to be consistent (e.g. for Work, one person might use PII, another ISWC), and may fall into disuse or be replaced. We do not want to DOI to be at the mercy of other standards timetables, success, etc.

By contrast the Type Code is derived from the *fundamental underlying logic* of the DOI metadata scheme (see DOI Discussion Paper 2 on metadata) ; will be *consistent* and *persistent*; and will be of tremendous *practical use* in implementation of DOI.

The code for Type will of course be carried in the kernel metadata. Why then should we include it in the number (and why not other elements of kernel metadata...isn't this the start of a slippery slope?). This is explained below.

- **A fundamental category**

Type is a fundamental category and the DOI (unlike any other identifier) covers several of these categories (the ISXN analogy). The Type is a *single value* and a *compulsory* element (unlike "other identifier").

- **A type code will be needed for local resolution uses**

- DOI is both an identifier, and a system to use that identifier to access the identified content;
- we have described in outline how in "local resolution", a DOI could be used just as an **identifier** (*without recourse to any Handle system or metadata system*); and how it can seamlessly become **identifier+system** when required.

When used as a (local) identifier, it will be just like the "ISXN" described above. It will be tremendously helpful for local systems to know what Type the DOI is and not have to consult or maintain metadata systems (analogy: like using ISBNs in EDI - at minimum, only the number is needed *because you know it is a Book*)

- **It will be of practical use in applications**

In the case of reference linking , DOIs must be assigned for both works and manifestations, and it will be very confusing if it is not known if an encountered DOI refers to an entity which is abstract, or something one can obtain. Note: in most cases, people will use Work identifiers and resolve them "invisibly" - it will not matter to authors: but when different manifestations come to be traded it will matter tremendously to publishers, agents, libraries and intermediaries. A type code enables Work and Manifestations to be clearly related (and also determine whether physical or digital).

### Appendix 3: possible future developments using DOI-based technology

The IDF has stated that it will work towards future implementation of the full Handle capability for multiple resolution. However initial implementations will be built using single-point resolution, and all future developments will be compatible with this implementation.

- **Using Handle multiple-resolution for multiple digital manifestations**

The Handle technology used with DOI would allow the DOI(Work) to be resolved to all the corresponding digital manifestations DOI(Manifestations). The first digital manifestation DOI(Manifestation) could be the default value (what we have sometimes called the level 1 DOI, the current implementation)

DOI is intending to adopt multiple resolution as part of its continuing development. This is not yet available for the DOI, but not required for the adoption of the initial reference linking implementation. This would however be a stimulus for practical developments of the technology by the STM community.

Note: other multiple-resolution mechanisms are logically possible to achieve this task. This paper focuses on Handle because its use would provide seamless continuity with the DOI.

- **Using Handle multiple-resolution for local resolution**

This has been called "The Harvard problem". Note that the issue described above is a separate problem from local resolution; but the Handle could be used for this too. Once the DOI of the required manifestation is known, that in turn can be resolved to a URL. At present a single URL, but in the future DOI implementation, multiple URLs. *A separate DOI paper on local resolution issues is in preparation.* (The point of the present paper is not to solve that problem, but the fundamental first problem: simple linkage between one digital manifestation of a citing and cited article). Other multiple-resolution mechanisms could be used to achieve this task too<sup>5</sup>. (Handle running in a local environment could be one such.)

Note that DOIs can be nested. So we can resolve DOI(Work) to DOI (Manifestation) to URL and at each step allow multiple resolution. DOI(W) to DOI(M) is step 2 of reference linking; DOI(M) to local URL is the "Harvard problem". Using DOI/Handle would allow, in the future, the same system to solve both problems. Seamless systems could be built to do both.

- **Possible allocation of DOIs to local copies**

The registration of a DOI requires that the registrant have rights in the entity (resource, creation) so identified. In its current initial implementation, this has typically been interpreted as publisher rights. However others in the information chain have defined or assigned rights, from author through to user. It may be that local usage of e.g. digital copies, having rights within a local environment (libraries) or for defined uses such as re-sale (intermediaries), may be helped by allocation of a DOI to local copies. This is possible using the DOI kernel (adopted from the INDECS schema), since the element "origination" allows identification of a Replica. For a defined genre or application it would be possible to allocate DOIs to replicas and record the original's DOI as a pointer within that genre kernel. Note that this application is outside the scope of the Journal Article genre and reference-linking application described here. Local resolution issues, as well as registration rights definitions, form a work item for the IDF during 1999; more work is needed before a definitive proposal can be made.

---

<sup>5</sup> For example, within Chemical Abstracts STNEasy service, a work is resolved locally to several different available options for providing a copy of a manifestation.

**Appendix 4: example**

Example of input form for a simple article with both a print and digital manifestation. The actual values used are fictitious.

<b>DOI</b>	<b>DOI of work</b>	10.2500/(W)123456789
	<b>DOI of first digital manifestation</b>	10.2500/(D)123456789
	<b>DOI of first print manifestation</b>	10.2500/(P)123456789
<b>DOI Genre</b>		Journal Article
<b>Identifier</b>	<b>Work Identifier</b>	<i>(none assigned)</i>
	<b>Article Identifier</b>	SICI 0264-1615(1999)<13:TDOISD>2.0.TX;2-A
	<b>Journal identifier</b>	ISSN 0264-1615
	<b>Publication Date</b>	1999
	<b>Journal Volume No</b>	27
	<b>Journal Issue No</b>	1
	<b>Article start page No</b>	13
<b>Title</b>	<b>Article Title</b>	The digital object identifier system: digital technology meets content management
	<b>Journal Title</b>	Interlending and Document Supply
<b>Type</b>		
	<b>Digital format</b>	pdf
<b>Origination</b>	<b>Manifestation (physical) Origination</b>	Original
	<b>Manifestation (digital) Origination</b>	Original
	<b>Work Origination</b>	Original
<b>Agent Identifier*</b>	<b>First named author</b>	Paskin, Norman
	<b>Additional authors</b>	
	<b>Publisher Name(s)</b>	MCB University Press
<b>Metadata registrant</b>		MCB University Press
<b>URL resolution</b>		<a href="http://www.mcbpubl.com/doiarticle">http://www.mcbpubl.com/doiarticle</a>
<b>Metadata version</b>		1
<b>Date of submission</b>		19990316