CNS 國家標準草案：數位物件識別符語法

(ANSI/NISO Z39.84—
2005 Syntax for the digital object Identifier)

中央研究院. 數位典藏國家型科技計畫後設資料工作組
與
國立台灣師範大學圖書資訊學研究所
研譯

經濟部標準檢驗局 **96** 年度 委辦計畫

規劃與建置數位內容與數位生活應用之技術標準環境
(案號：**1D15960125-20)**

Syntax for the Digital Object Identifier

目錄

1.適用範圍

本標準定義數位物件識別符（DOI）之字元串語法。

本標準僅限於定義DOI字元串之語法。至於管理DOI識別符的指定與使用政策則係由國際DOI基金會（International DOI Foundation , IDF）所訂定，不在本標準的討論範圍內。

This standard defines the syntax for a character string called the Digital Object Identifier (DOI). limited to defining the syntax of the DOI character string. Policies governing the assignment and use of DOIs are determined by the International DOI Foundation (IDF) and are outside the scope of this document.

2.用語釋義

2.1 置入(Deposit)

將 DOI 以及相關會使用到的資訊鍵入 DOI 目錄中之動作。

2.2 數位物件識別符(DOI)

由字元串所組成的識別符，其符合 IDF 所制定的規則，同時也會置入 DOI 目錄中進行管理。

2.3 目錄(Directory)

存放 DOI 識別符以及其伴隨置入地點的儲存庫。

2.4 目錄管理者(Directory Manager)

代表 IDF 管理目錄的組織。

2.5 DOI 前綴(DOI prefix)

由登錄代理者所核發給登錄者之目錄及登錄者碼，係用於由登錄者配置的 DOI 中。

2.6 DOI 後綴(DOI suffix)

由登錄者所指定的字元串。此後綴必須要在特定的 DOI 前綴下確保其唯一性。

2.7 國際 DOI 基金會(International DOI Foundation, IDF)

該組織藉由建立並管理數位物件識別符(DOI)系統，訂定使用策略，指定系統服務提供者，並監督確保系統的正常運作，以支援智慧財產團體在數位環境中之需求。

2.8 登錄者(Registrant)

登錄者可以是組織或個體，以請求登錄代理者配置一或多個 DOI 前綴。

2.9 登錄(Registration)

由登錄代理者針對登錄者進行 DOI 前綴配置之動作。

2.10DOI 登錄代理者(DOI Registration Agency)

此係由 IDF 指定授權派發以及登錄 DOI 前綴的組織，同時也接受登錄者進行 DOI 識別符的置入動作。

2.1 Deposit

The act of entering into the Directory a DOI and associated information necessary

for the DOI to be used.

2.2 DOI

A character string used in a System conforming to the rules of, and deposited in the Directory administered by, the IDF.

2.3 Directory

A repository in which DOIs are deposited and attendant locations are maintained.

2.4 Directory Manager

The organization that manages the Directory on behalf of the IDF.

2.5 DOI prefix

The Directory and the Registrant codes issued by a Registration Agency to a Registrant for use in the DOIs allocated by that Registrant.

2.6 DOI suffix

The character string assigned by a Registrant. The suffix shall be unique within the set of DOIs specified by the DOI prefix held by the Registrant.

2.7 International DOI Foundation, IDF

The body set up to support the needs of the intellectual property community in the digital environment by establishing and governing the DOI System, setting policies for the System, appointing service providers for the System, and overseeing the successful operation of the System.

2.8 Registrant

An organization or entity that has requested and been allocated one or more DOI prefixes by a Registration Agency.

2.9 Registration

The act of allocating the DOI prefix to a Registrant by the Registration Agency.

2.10 DOI Registration Agency

Foundation to register and allocate DOI prefixes to Registrants, and which subsequently accepts DOIs being deposited by Registrants.

3.引用標準

參考標準是指用來輔助建構DOI識別符的相關標準。其他第二級的標準與參考包含了可以與DOI銜接使用的文件（參照附錄C）。

Referenced standards are those that need to be used to construct a DOI. Secondary standards and references include citations to documents that can be of use in conjunction with the DOI. See Appendix 3 for related standards and references.

ISBN 0-321-18578-1 國際統一碼協會（The Unicode Consortium）的統一碼標準（*The Unicode Standard*）第 4.0.1 版。是由統一碼標準第 4.0 版所指定（Reading, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1），並經 4.0.1 版修正（http://www.unicode.org/versions/Unicode4.0.1/）。

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

4.DOI 的格式及特徵

DOI字串之語法以<DIR>.<REG>/<DSS>表示。

DOI是由前綴及後綴所組成。前綴是由目錄碼（Directory Code） <DIR>以及登錄碼（Registrant Code）<REG>組成。後綴則是由DOI後綴串<DSS>為主。

DOI字串的長度基本上沒有實際限制，其組件亦同。處置系統（Handle System）允許字串最大長度為4GB，在UTF-8編碼下每一個ASCII字元使用一個位元組，因此在使用ASCII編碼下DOI可以允許大約40億個字元）。

DOI 字串中所使用的字元 ˋaˊ ˋzˊ 以及 ˋAˊ ˋZˊ 沒有區分大小寫（例如10.123/ABC和10.123/AbC完全相同）。這些DOI字串中的字元都會在登錄及解析階段先一律轉為大寫字元，若一個已登錄的DOI識別符是10.123/ABC，則10.123/abc可以成功的解析，而隨後如果有10.123/AbC想要註冊成為DOI字串，此登錄動作將會被拒絕，同時傳回一個錯誤訊息告知此DOI識別符已經存在。針對兩個DOI識別符的比較，兩個識別符都會先轉成大寫字，然後對整個DOI字串進行一對一的八位元比較動作。

The syntax of the DOI string is: <DIR>.<REG>/<DSS>
The DOI is composed of the prefix and the suffix. Within the prefix are the Directory Code <DIR> and the Registrant Code <REG>. The suffix is made up of the DOI Suffix String <DSS>.

There is no practical limit on the length of a DOI string, or any of its components (the Handle system allows strings of up to 4 GB; under UTF-8 encoding each ASCII character takes one byte, hence in ASCII encoding a DOI may be approx 4 billion characters).

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

4.1 DOI 字元集

合法字元係統一碼（Unicode）字元集中所表達的圖像字元。此範圍意味著會將控制字元

（0x00-0x1F和0x80-0x9F）排除在合法字元之外，也就是不會成為DOI識別符可以使用的正確字元，也不會存放在DOI識別符相容系統中。至於特定的保留字元，則留到前綴與後綴段落進行說明。

Legal characters are the legal graphic characters of Unicode. This specifically excludes the control character ranges 0x00-0x1F and 0x80-0x9F, which are therefore not valid characters for DOI strings, and will never be present in DOI conformant systems. Reserved characters, if any, are listed in the following descriptions of the prefix and suffix.

4.2 前 綴

<DIR>目錄碼（必要）

目錄碼可以使用的有效字元（參照附錄A），而更新這些有效字元值有特定的維護組織。目錄碼是數值化的資料值，目前唯一合法的資料值是<DIR>=10。

<REG>登錄碼（必要）

此部份與<DIR>資料以 ‘.’ 進行區隔。此係由登錄者指定。

DOI前綴元集

在DOI字元集中的任何字元則是根據以上定義。

<DIR>和<REG>係由IDF指定。

<DIR> Directory Code (required)

See Appendix 1 for all valid values for the Directory Code. The Maintenance Agency is responsible for updating the list of valid values. The Directory Code is numeric; currently the only valid value is <DIR>=10.

<REG> Registrant’s Code (required)

Separated from <DIR> by “.”. This is assigned to the Registrant by the International DOI Foundation.

DOI Prefix Character Set.

Any character within the DOI Character Set as defined above.

<DIR> and <REG> are assigned by the International DOI Foundation.

4.3 後 綴

<DSS>DOI後綴串（必要）

由登錄者指定。

.

DOI後綴元集

在DOI字元集中的字元則根據以上定義。但是後綴不能包含 ‵*′ 字元，此係因 ‵*′ 代表任何單一字元。此部份係保留給未來使用。DSS部份沒有區分大小寫。

<DSS> DOI Suffix String (required)
This is assigned by the Registrant.

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

附錄 A

DOI 規格

1.範圍

本附錄並非數位物件識別符語法規範（ANSI/NISO Z39.84 2005）的一部份，本附錄提供有關DOI系統語法實作層面的相關訊息，此部份由國際DOI基金會所訂定，同時也不會影響DOI語法在本標準中的定義。

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only. This appendix provides information on aspects of the DOI system syntax implementation which are determined by the International DOI Foundation and which will not change the DOI syntax defined in this standard.

2.目錄碼中可使用的資料值

<DIR> <REG>是由國際DOI基金會所指派，前綴是由數字所組成，合法的<DIR>資料值為10。

DOI識別符是永久儲存的，定義在IETF RFC 1737中。在一致性資源名稱（Uniform Resource Names）的功能需求（http://www.ietf.org/rfc/rfc1737.txt）中說：〝URN的生命期應該是永久的，也就是說，URN具有全域唯一的特性，而且可以適當地用在資源參考上，或是任何擁有URN名稱的命名授權者上。〞

編碼目前是由處置系統（Handle System）所指定使用。因此，所有的統一碼字元都將會以UTF-8規格進行編碼。

處置系統（Handle System）目前用來當作DOI系統的基礎，同時也允許無限長度的DOI識別符字串。但無論如何，後綴（<DSS>）在實用上建議使用盡量短的字串，以利人的閱讀以及便利各式對長度有需求的系統使用上（例如浮水印的使用上）。

<DIR> <REG> is assigned by the International DOI Foundation. The prefix is numeric.
Valid value for <DIR> = 10
DOIs are persistent, as defined in IETF RFC 1737. Functional Requirements for Uniform Resource Names. (http://www.ietf.org/rfc/rfc1737.txt): “It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.”
encoding is mandated by the Handle System. Therefore, all Unicode characters must be encoded using UTF-8.
The Handle System used as the basis for the DOI system allows an unlimited length for the DOI string. However it is recommended that the suffix (<DSS>) be kept as short as possible to allow for human readability and ease of use in systems where size may be a consideration (e.g., watermarking).

附錄 B

數位物件識別符範例

1.範圍

本附錄並非數位物件識別符語法規範（ANSI/NISO Z39.84 2005）的一部份。

DOI識別符登錄者針DSS可以使用多種形式的字串，包括私有的識別符以及既存的標準，如SICI（Serial Item and Contribution Identifier）。識別符數值編碼方式就是可以運用任意既存的識別符語法來表達成可以適用在DOI系統中的字串形式。

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only.

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

2.範例

例1.Authors' Licensing and Collecting Society's Byline service所使用之DOI：

10.054/1418EC1N2LE

例2.John Wiley & Sons出版的American Society for Information Science期刊中的一篇文章所使用之DOI（與SICI語法結合之DOI）：

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

例3.期刊Journal of the American Medical Association內的一篇文章所使用之DOI：

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

例4.一篇出自由Academic Press所出版之Encyclopedia of Immunology Online第二版的文章〝ABO Blood Group System〞所使用之DOI：10.1006/rwei.1999.0001

The following are examples of Digital Object Identifiers:

e.g.1: DOI for the Authors' Licensing and Collecting Society's Byline service: 10.054/1418EC1N2LE

e.g.2: DOI (incorporating a SICI) from an article in the Journal of the American Society for Information Science, published by John Wiley & Sons:

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

e.g.3: DOI for an article from JAMA, the Journal of the American Medical Association:

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

e.g.4: DOI for the article "ABO Blood Group System" from Encyclopedia of Immunology Online, 2nd edition, published by Academic Press: 10.1006/rwei.1999.0001

附錄 C

相關標準及參考

本附錄中包含了對其他標準的參考與引用，這些對DOI識別符是有用且針對DOI識別符提供了更多有用的資訊。以下引用到美國國家標準的部份將會以將來改版之後的版本為主。

This appendix includes references to other standards and citations that may be useful with DOIs or which provide additional information on the DOI. When American National Standards cited below are superseded by a revision, the revision shall apply.

ANSI X3.4: 1986　1986年資訊系統所用之美國標準中的編碼字元集（7-bit American National Standard）

DOI 10.1000/182 DOI識別符手冊（DOI Handbook）

DOI factsheets　DOI情況說明書（DOI與處置Handle、DOI與數值編碼方式、DOI與資料辭典、DOI與網際網路識別符規格、DOI應用、DOI系統加值情形）

Handle System　處置系統

RFC 3650　處置系統概述（Handle System Overview）

RFC 3651　處置系統命名空間與服務定義（Handle System Namespace and Service Definition）

RFC 3652　處置系統協定規範（Handle System Protocol (Ver 2.1) Specification）

RFC2044　UTF-8，Unicode及ISO10646轉換格式（UTF-8, A Transform Format for Unicode and ISO10646）

ANSI X3.4:1986 American National Standard for Information Systems – Coded Character Sets – 7-bit American National Standard Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986. Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986。

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

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

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

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

Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle System Namespace and Service Definition". RFC 3651, November 2003. http://www.handle.net/rfc/rfc3651html

Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". RFC 3652, November 2003. http://www.handle.net/rfc/rfc3652html

“UTF-8, A Transform Format for Unicode and ISO10646”, RFC2044, October 1996, Yergeau, Francois - http://www.normos.org/ietf/rfc/rfc2044.txt

附錄 D

應用議題

1.範圍

本附錄並非數位物件識別符語法規範（ANSI/NISO Z39.84 2005）的一部份。

除了本標準中所提的特殊需求（例如使用統一碼標準以及保留字元）之外，其餘DOI識別符可以使用的字元都不再會有任何的限制或是假設。附錄D中討論編碼方面的議題，是當DOI識別符使用在

特定的應用情境下才需要討論，如：URL之於HTTP協定的情境下。其他的應用情境也會有類似型態的需求與限制。但無論如何，這樣的編碼需求與限制僅適用於將DOI識別符使用在特定的應用情境下，且這些都不會成為本文件中所述之DOI語法標準的一部分。

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84-2005. It is included for information only.)
Except for the specific requirements imposed by this standard (such as use of Unicode and reserved characters), no restrictions are imposed or assumptions made about the characters used in DOIs. Appendix D discusses some encoding issues that arise when using DOIs in specific application contexts like URLs and with the HTTP protocol. Other application contexts in which DOIs are used may have similar types of requirements or restrictions. However, such requirements for encoding or restrictions on the use of particular characters only apply when DOIs are used within those particular application contexts. They are not part of the DOI syntax itself as defined by this document.

2.UTF-8編碼

處置系統（Handle System）規定UTF-8為DOI字串的編碼方式。ASCII字元在UTF-8編碼方式下保有其原來的呈現形式。因此ASCII字元在UTF-8編碼下不需要有任何的改變。統一碼的預設編碼使用16位元（2個八位元組）的方式，UTF-8編碼則是統一碼編碼方式的變形，允許字元被編成一到六個八位元組。UTF-8編碼目前在非ASCII字元的顯示上有著一定的地位，舉例來說，日本字〝日本語〞是寫成：

The Handle System specifies UTF-8 as the encoding for DOI strings. ASCII characters are preserved under UTF-8 encoding. No changes need to be made to ASCII characters to comply with UTF-8 encoding. The default encoding of Unicode is that each character consists of 16 bits (2 octets). UTF-8 is a variation of the Unicode encoding that allows characters to be encoded in terms of one to six octets. UTF-8 encoding plays a role when non-ASCII characters are used. For example, the Japanese word “nihongo” is written as:

漢字〝日本語〞的統一碼資料表達為：65E5 672C 8A9E
而這些資料如果使用UTF-8進行編碼則會表達為：E6 97 A5 E6 9C AC E8 AA 9E
如需參考更多UTF-8資訊（參照〝UTF-8, A Transform Format for Unicode and ISO10646〞, RFC2044, October 1996）。

The Unicode sequence representing the Han characters for “nihongo” is: 65E5 672C 8A9E
These may be encoded in UTF-8 as follows: E6 97 A5 E6 9C AC E8 AA 9E
further information on UTF-8 see “UTF-8, A Transform Format for Unicode and ISO10646”, RFC2044, October 1996.

3.將識別符用於URL時的編碼建議

目前的Web瀏覽器技術還需要加入一些額外特性，才能讓瀏覽器具有運用DOI識別符的能力：也就是說，瀏覽器需要新增特色，期待將來瀏覽器會將支援識別符的解析特色加入。

目前有免費的〝解析器外掛〞可供下載（http://www.handle.net/resolver/），解析器外掛用於Netscape以及Microsoft IE瀏覽器，係用來讓瀏覽器處理代碼協定（Handle protocol）。

另一種支援代碼解析的方式，是不需要針對瀏覽器的功能進行擴充。DOI識別符可以透過使用一個預設公開的DOI代理伺服器（http://dx.doi.org）來進行運作。DOI識別符的解析在這種情形下透過使用URL語法來完成，例：〝doi:10.123/456〞可以寫成〝http://dx.doi.org/10.123/456〞。

DOI亦主要使用於HTML網頁上。例：要DOI識別符〝10.1006/rwei.1999".0001〞當作是一個HTML網頁，可以寫成〝<A HREF=”http://dx.doi.org/10.1006/rwei.1999%22.0001">10.1006/rwei.1999%22.0001</A>〞

請注意引號〝“〞已經被編碼過了（參照下節），因為DOI在URL的表達中必須和週邊的字元進行區隔。在此DOI識別符的呈現已是編過碼的狀況，此形式可讓使用者在瀏覽器中直接將這種形式的DOI識別符鍵入。

Current Web browser technology requires additional functionality to allow the browser to make full use of DOIs: additional browser features are necessary. It is anticipated that features supporting resolution will commonly be built into browsers in the future.

There is a freely available “resolver plug in” that can be downloaded from http://www.handle.net/resolver/. For both Netscape and Microsoft IE browsers, the plug-in extends the browser's functionality so that it understands the Handle protocol.

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

DOIs are also primarily used in HTML pages. The DOI 10.1006/rwei.1999".0001as a link in an HTML page would be: <A HREF=”http://dx.doi.org/10.1006/rwei.1999%22.0001">10.1006/rwei.1999%22.0001</A>

Note that “ has been encoded (see next section) to distinguish the DOI in the URL from the surrounding text. The DOI is displayed in its encoded form since users may type the DOI directly into their browsers.

4.編碼議題

當DOI識別符的使用場合是HTML、URL或HTTP環境時有特別的編碼需求。一致性資源識別碼（Uniform Resource Identifier，URI）的語法相較於DOI識別符來說，是更為嚴格的。所謂一個一致性資源識別碼URI，可以是一個一致性資源定位器（Uniform Resource Locator，URL）或是一個一致性資源名稱（Uniform Resource Name，URN）。

當一個在DOI識別符中的字元使用在URL或是URN的場合時，如果URL或URN無法允許顯示該字元，或該字元另有其他意義時，就必須使用十六進位編碼來進行該字元的表達。所謂的十六進位編碼表達方式，就是將特定字元使用其十六進位資料值，並再前方冠上一個百分符號。因此，〝#〞字元就會表示成為〝%23〞，而〝http://dx.doi.org/10.1000/456#789〞則會表示成為〝http://dx.doi.org/10.1000/456%23789〞。目前瀏覽器並不會遇上單獨的〝#〞字元，此字元通常都是用在一個URL的結尾部分並表示其他參數段落的開始。請注意：DOI識別符本身並不會因編碼方式而改變，以上說明僅僅是指在URL下的表達。一個編碼過後的DOI識別符會先進行解碼，然後再送到DOI登錄端（Registry）進行處理。目前解碼的動作是由代理伺服器（http://dx.doi.org/）來進行處理。只有未經編碼的DOI識別符會被存放在DOI註冊端資料庫中。例：一個DOI註冊端中的識別符應該叫做〝10.1000/456#789〞，而不是〝10.1000/456%23789〞。此外，在任一個URL中，百分符號〝%〞都會編碼成為〝%25〞。

DOI識別符數字字串本身並沒有太多的字元限制，但是當DOI識別符使用在URL上時，就必須要遵守URL語法的命名習慣。但是同一個DOI識別符在別的應用情境下並不一定要遵守相同的命名習慣。

There are special encoding requirements when a DOI is used with HTML, URLs, and HTTP. The syntax for Uniform Resource Identifiers (URIs) is much more restrictive than the syntax for the DOI. A URI can be a Uniform Resource Locator (URL) or a Uniform Resource Name (URN).

Hexadecimal encoding must be used for characters in a DOI that are not allowed, or have other meanings, in URLs or URNs. Hex encoding consists of substituting for the given character its hexadecimal value preceded by percent. Thus, # becomes %23 and http://dx.doi.org/10.1000/456#789 is encoded as http://dx.doi.org/10.1000/456%23789. The browser does not now encounter the bare #, which it would normally treat as the end of the URL and the start of a fragment, and so sends the entire string off to the DOI network of servers for resolution, instead of stopping at the #. Note: The DOI itself does not change with encoding, merely its representation in a URL. A DOI that has been encoded is decoded before being sent to the DOI Registry. At the moment the decoding is handled by the proxy server http://dx.doi.org/. Only unencoded DOIs are stored in the DOI Registry database. For example, the number above is in the DOI Registry as “10.1000/456#789” and not“10.1000/456%23789”. The percent character (%) must always be hex encoded (%25) in any URLs.

There are few character restrictions for DOI number strings per se. When DOIs are embedded in URLs, they must follow the URL syntax conventions. The same DOI need not follow those conventions in other

contexts.

5.DOI識別符存放與URL的強制與建議編碼方式

表1及表2整理了DOI識別符的編碼指導綱要。URL使用最多限制之字元集。表1列出的字元〝必須〞以十六進位編碼，但是表2列出了一些附加的字元，這些字元是〝建議〞使用十六進位編碼。兩個表之間的差異是基於現行Web瀏覽器環境的實際經驗以及URL語法的正式規格所形成。但是在DOI目錄中，所有的字元都只需要使用自身的表達方式。

Tables 1 and 2 summarize the encoding guidelines for DOI. URLs have the most restricted set of characters. Table 1 lists the characters that should always be hex encoded. Table 2 lists additional characters where it is recommended that characters be replaced by hex-encoding. The distinction between the lists is between practical experience with current Web browsers and the more formal specification of URL syntax. In the DOI Directory all characters represent themselves.

表1、強制編碼的字元

Character	Encoding
%	(%25)
"	(%22)
#	(%23)
SPACE	(%20)

表2、建議編碼的字元

Character	Encoding
<	(%3c)
>	(%3e)
{	(%7b)

英中名詞對照表

-A-	
Allocated	配置
-B-	
-C-	
Character string	字元串
-D-	
Deposit	置入
DOI	數位物件識別符
Directory	目錄
Directory Manager	目錄管理者
DOI prefix	DOI 前綴
DOI suffix	DOI 後綴
DOI Registration Agency	數位物件識別符登錄代理者
-E-	
-F-	
-G-	
-H-	
Handle System	處置系統
-I-	
International DOI Foundation, IDF	國際 DOI 基金會
Issue	核發
-J-	
-K-	
-L-	
-M-	
-N-	
-O-	
-P-	
-Q-	
-R-	
Registrant	登錄者
Registration	登錄
-S-	
-T-	
-U-	
-V-	

-W-

-X-

-Y-

-Z-

Syntax for the Digital Object Identifier

Contents

1.Scope

This standard defines the syntax for a character string called the Digital Object Identifier (DOI). This standard is limited to defining the syntax of the DOI character string. Policies governing the assignment and use of DOIs are determined by the International DOI Foundation (IDF) and are outside the scope of this document.

2.Terms and definitions

2.1 Deposit

The act of entering into the Directory a DOI and associated information necessary for the DOI to be used.

2.2 DOI

A character string used in a System conforming to the rules of, and deposited in the Directory administered by, the IDF.

2.3 Directory

A repository in which DOIs are deposited and attendant locations are maintained.

2.4 Directory Manager

The organization that manages the Directory on behalf of the IDF.

2.5 DOI prefix

The Directory and the Registrant codes issued by a Registration Agency to a Registrant for use in the DOIs allocated by that Registrant.

2.6 DOI suffix

The character string assigned by a Registrant. The suffix shall be unique within the set of DOIs specified by the DOI prefix held by the Registrant.

2.7 International DOI Foundation, IDF

The body set up to support the needs of the intellectual property community in the digital environment by establishing and governing the DOI System, setting policies for the System, appointing service providers for the System, and overseeing the successful operation of the System.

2.8 Registrant

An organization or entity that has requested and been allocated one or more DOI prefixes by a Registration Agency.

2.9 Registration

The act of allocating the DOI prefix to a Registrant by the Registration Agency.

2.10 DOI Registration Agency

Foundation to register and allocate DOI prefixes to Registrants, and which subsequently accepts DOIs being deposited by Registrants.

3.Normative reference

Referenced standards are those that need to be used to construct a DOI. Secondary standards and references include citations to documents that can be of use in conjunction with the DOI. See Appendix C for related standards and references.

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

4.Format and Characteristics of the DOI

The syntax of the DOI string is: <DIR>.<REG>/<DSS>

The DOI is composed of the prefix and the suffix. Within the prefix are the Directory Code <DIR> and the Registrant Code <REG>. The suffix is made up of the DOI Suffix String <DSS>.

There is no practical limit on the length of a DOI string, or any of its components (the Handle system allows strings of up to 4 GB; under UTF-8 encoding each ASCII character takes one byte, hence in ASCII encoding a DOI may be approx 4 billion characters).

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

4.1 DOI Character Set

Legal characters are the legal graphic characters of Unicode. This specifically excludes the control character ranges 0x00-0x1F and 0x80-0x9F, which are therefore not valid characters for DOI strings, and will never be present in DOI conformant systems. Reserved characters, if any, are listed in the following descriptions of the prefix and suffix.

4.2 Prefix

<DIR> Directory Code (required)

See Appendix A for all valid values for the Directory Code. The Maintenance Agency is

responsible for updating the list of valid values. The Directory Code is numeric; currently the only valid value is <DIR>=10.

<REG> Registrant's Code (required)

Separated from <DIR> by ".". This is assigned to the Registrant by the International DOI Foundation.

DOI Prefix Character Set.

Any character within the DOI Character Set as defined above.

<DIR> and <REG> are assigned by the International DOI Foundation.

4.3 Suffix

<DSS> DOI Suffix String (required)

This is assigned by the Registrant.

DOI Suffix Character Set

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

Appendix A: DOI Specifications

1. Scope

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only. This appendix provides information on aspects of the DOI system syntax implementation which are determined by the International DOI Foundation and which will not change the DOI syntax defined in this standard.

2. Valid values for Directory Code

<DIR> <REG> is assigned by the International DOI Foundation. The prefix is numeric.

Valid value for <DIR> = 10

DOIs are persistent, as defined in IETF RFC 1737. Functional Requirements for Uniform Resource Names. (http://www.ietf.org/rfc/rfc1737.txt): “It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.” encoding is mandated by the Handle System. Therefore, all Unicode characters must be encoded using UTF-8.

The Handle System used as the basis for the DOI system allows an unlimited length for the DOI string. However it is recommended that the suffix (<DSS>) be kept as short as possible to allow for human readability and ease of use in systems where size may be a consideration (e.g., watermarking).

Appendix B: Examples of Digital Object Identifiers

1. Scope

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84 2005. It is included for information only.

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

2. Examples

The following are examples of Digital Object Identifiers:

e.g.1: DOI for the Authors' Licensing and Collecting Society's Byline service: 10.054/1418EC1N2LE

e.g.2: DOI (incorporating a SICI) from an article in the Journal of the American Society for Information Science, published by John Wiley & Sons: 10.1002/(SICI)1097-4571(199806)49:8<693::AID-ASI4>3.0.CO;2-O

e.g.3: DOI for an article from JAMA, the Journal of the American Medical Association: 10.1001/PUBS.JAMA(278)3,JOC7055-ABST:

e.g.4: DOI for the article "ABO Blood Group System" from Encyclopedia of Immunology Online, 2nd edition, published by Academic Press: 10.1006/rwei.1999.0001

Appendix C: Related Standards and References

This appendix includes references to other standards and citations that may be useful with DOIs or which provide additional information on the DOI. When American National Standards cited below are superseded by a revision, the revision shall apply.

ANSI X3.4:1986 American National Standard for Information Systems – Coded Character Sets – 7-bit American National Standard Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986.

Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986。

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

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

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

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

Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle System Namespace and Service Definition". RFC 3651, November 2003. http://www.handle.net/rfc/rfc3651html

Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". RFC 3652, November 2003. http://www.handle.net/rfc/rfc3652html

“UTF-8, A Transform Format for Unicode and ISO10646”, RFC2044, October 1996, Yergeau, Francois - http://www.normos.org/ietf/rfc/rfc2044.txt

Appendix D: Application Issues

1. Scope

This appendix is not part of *Syntax for the Digital Object Identifier*, ANSI/NISO Z39.84-2005. It is included for information only.)

Except for the specific requirements imposed by this standard (such as use of Unicode and reserved characters), no restrictions are imposed or assumptions made about the characters used in DOIs. Appendix D discusses some encoding issues that arise when using DOIs in specific application contexts like URLs and with the HTTP protocol. Other application contexts in which DOIs are used may have similar types of requirements or restrictions. However, such requirements for encoding or restrictions on the use of particular characters only apply when DOIs are used within those particular application contexts. They are not part of the DOI syntax itself as defined by this document.

2. UTF-8 Encoding

The Handle System specifies UTF-8 as the encoding for DOI strings. ASCII characters are preserved under UTF-8 encoding. No changes need to be made to ASCII characters to comply with UTF-8 encoding. The default encoding of Unicode is that each character consists of 16 bits (2 octets). UTF-8 is a variation of the Unicode encoding that allows characters to be encoded in terms of one to six octets. UTF-8 encoding plays a role when non-ASCII characters are used. For example, the Japanese word “nihongo” is written as:

The Unicode sequence representing the Han characters for “nihongo” is: 65E5 672C 8A9E

These may be encoded in UTF-8 as follows: E6 97 A5 E6 9C AC E8 AA 9E

further information on UTF-8 see “UTF-8, A Transform Format for Unicode and ISO10646”, RFC2044, October 1996.

3. Encoding Recommendations When Used in URLs

Current Web browser technology requires additional functionality to allow the browser to make full use of DOIs: additional browser features are necessary. It is anticipated that features supporting resolution will commonly be built into browsers in the future.

There is a freely available “resolver plug in” that can be downloaded from http://www.handle.net/resolver/. For both Netscape and Microsoft IE browsers, the plug-in extends the browser's functionality so that it understands the Handle protocol.

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

DOIs are also primarily used in HTML pages. The DOI 10.1006/rwei.1999".0001as a link in an HTML page would be: <A HREF=”http://dx.doi.org/10.1006/rwei.1999%22.0001">10.1006/rwei.1999%22.0001</A>

Note that “ has been encoded (see next section) to distinguish the DOI in the URL from the surrounding text. The DOI is displayed in its encoded form since users may type the DOI directly into their browsers.

4. Encoding Issues

There are special encoding requirements when a DOI is used with HTML, URLs, and HTTP. The

syntax for Uniform Resource Identifiers (URIs) is much more restrictive than the syntax for the DOI. A URI can be a Uniform Resource Locator (URL) or a Uniform Resource Name (URN).

Hexadecimal encoding must be used for characters in a DOI that are not allowed, or have other meanings, in URLs or URNs. Hex encoding consists of substituting for the given character its hexadecimal value preceded by percent. Thus, # becomes %23 and http://dx.doi.org/10.1000/456#789 is encoded as http://dx.doi.org/10.1000/456%23789. The browser does not now encounter the bare #, which it would normally treat as the end of the URL and the start of a fragment, and so sends the entire string off to the DOI network of servers for resolution, instead of stopping at the #. Note: The DOI itself does not change with encoding, merely its representation in a URL. A DOI that has been encoded is decoded before being sent to the DOI Registry. At the moment the decoding is handled by the proxy server http://dx.doi.org/. Only unencoded DOIs are stored in the DOI Registry database. For example, the number above is in the DOI Registry as "10.1000/456#789" and not"10.1000/456%23789". The percent character (%) must always be hex encoded (%25) in any URLs.

There are few character restrictions for DOI number strings per se. When DOIs are embedded in URLs, they must follow the URL syntax conventions. The same DOI need not follow those conventions in other contexts.

5. Mandatory and Recommended Encoding for DOI Deposit and URLs

Tables 1 and 2 summarize the encoding guidelines for DOI. URLs have the most restricted set of characters. Table 1 lists the characters that should always be hex encoded. Table 2 lists additional characters where it is recommended that characters be replaced by hex-encoding. The distinction between the lists is between practical experience with current Web browsers and the more formal specification of URL syntax. In the DOI Directory all characters represent themselves.

Table 1: Mandatory Encoding

Character	Encoding
%	**(%25)**
"	**(%22)**
#	**(%23)**
SPACE	**(%20)**

Table 2: Recommended Encoding

Character	Encoding
<	**(%3c)**
>	**(%3e)**
{	**(%7b)**

參考資料

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

ANSI X3.4:1986 American National Standard for Information Systems – Coded Character Sets – 7-bit American National Standard Code for Information Interchange (7-bit ASCII) New York: ANSI, 1986.

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

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

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

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

Sun, Sam; Reilly, Sean; Lannom, Larry. "Handle System Namespace and Service Definition". RFC 3651, November 2003. http://www.handle.net/rfc/rfc3651html

Sun, Sam; Reilly, Sean; Lannom, Larry; Petrone, Jason. "Handle System Protocol (Ver 2.1) Specification". RFC 3652, November 2003. http://www.handle.net/rfc/rfc3652html

“ UTF-8, A Transform Format for Unicode and ISO10646”, RFC2044, October 1996, Yergeau, Francois - http://www.normos.org/ietf/rfc/rfc2044.txt

爭議事項

無

英中名詞對照表

-A-

Allocated	配置

-B-

-C-

Character string	字元串

-D-

Deposit	置入
DOI	數位物件識別符
Directory	目錄
Directory Manager	目錄管理者
DOI prefix	DOI 前綴
DOI suffix	DOI 後綴
DOI Registration Agency	數位物件識別符登錄代理者

-E-

-F-

-G-

-H-

Handle System	處置系統

-I-

International DOI Foundation, IDF	國際 DOI 基金會
Issue	核發

-J-

-K-

-L-

-M-

-N-

-O-

-P-

-Q-

-R-

Registrant	登錄者
Registration	登錄

-S-

-T-

-U-

-V-

-W-

-X-

-Y-

-Z-

中英名詞對照表

DOI 前綴	DOI prefix
DOI 後綴	DOI suffix
目錄	Directory
目錄管理者	Directory Manager
字元串	Character string
核發	Issue
配置	Allocated
國際 DOI 基金會	International DOI Foundation，IDF
處置系統	Handle System
登錄	Registration
登錄者	Registrant
置入	Deposit
數位物件識別符	DOI
數位物件識別符登錄代理者	DOI Registration Agency