

檔案命名與數位資源管理的關係

Using Naming Systems with the Concept of Uniform Resources Names for Digital Resources Management in Libraries

張文熙

Wen-Hsi Chang

國家檔案局籌備處資訊組科長

Section Chief of Archives Information Division

The Preparatory Office of National Archives

【摘要】

數位資源的應用，基本上是以檔案名稱作為內容識別的基礎。而當今數位資源最主要的應用方式是透過全球資訊網 WWW 是以 URL (Uniform Resource Locators) 來對網際網路上的數位檔案資源進行定址，檔案名稱也是定址範圍的一部分。而除了主機名稱的部分可以透過網域名稱服務 DNS (Domain Name Service) 以相對位址的方式來解決，而檔案所在的路徑及名稱，仍然是以絕對位址的方式來命名。以絕對位址來命名的方式問題在於當絕對位址變更時，直接會影響到網頁或程式的存取位址。若改採相對位址的定名方式則可以解決，大量檔案在異動頻繁變更絕對位址時造成的資源檢索的錯誤。為了讓資源名稱可以更簡便的維護，相對位址的命名方式是比較理想的。本文旨在說明以相對位址的命名的目的及相關原則提供需要管理大量數位資料檔案的圖書館作為管理上的參考。

【Abstract】

Most of digital resources were identified by different file names in applications. We locate a certain digital resource on Internet by URL. Basically, URL defines the physical location of a certain digital resource by an absolute name addressing method as referred a specific file name. However, such a way is not convenient for maintaining file names. Once a file was moved to a different server or storage, the users could not get it correctly any more by the original linking description on the web page. So far, digi-

tized resources indirect name addressing method can resolve this kind of problems as DNS system deals with host names to map their absolute IP addresses by an indirect addressing method on Internet. The gist of the article explains what the indirect naming system is and why we need in library systems.

關鍵詞：檔案資源命名、WWW、網際網路

Keywords: URN, Uniform Resource Name, WWW, Internet

壹、前言

隨著網際網路的普及數位化資料對圖書館的重要性也跟著越來越高，各類型的圖書館提供越來越多的數位化資源以滿足讀者的需要。同時傳統資料也越來越需要透過網頁來輔助，以加強它被應用的可能。例如圖書目錄及期刊目次等。因為網頁資料的超連結功能擴大了讀者檢索能力，加上瀏覽器使用簡便自然使得願意選擇數位資料的讀者越來越多。很明顯地，全球資訊網（WWW, World Wide Web）已然成為現今網際網路的重心，因為大部分的讀者使用網際網路的資源是透過 WWW 來檢索或取得資料為主，所以當網路上數位資料越來越多時，也使得超連結的層次越來越深，開始出現幾個問題：

1. 相同的數位資源重複貯存於不同媒體，而且比例太高。
2. 相同的數位資源以不同的檔名重複貯存而不自知。
3. 從數位資源貯存的名稱看不出資料來源，使得應用上變得困難。
4. 數位資源因貯存媒體或主機變更時，原有的網頁連結若不能及時更新，檔案資源就此消失。
5. 直接使用數位資源實體貯存的檔名，使得程式的可攜性(Portability)降低。

這些問題的解決方式可以從幾個方向來思考：

1. 定義全球統一的命名規則。
2. 連結資源使用間接命名方式，而不直接叫用實體貯存的檔名。
3. 建立一個可以控管命名或是取號的自動化系統。
4. 能和現有的編碼或命名系統（Legacy System）並存，減少變更現有資源命名系統困難度。

筆者在此介紹網路資源統一命名的需要及統一命名的機制的概念提供圖書館界的先進們對數位檔案資源管理的另一個思考的方向。

貳、命名系統的緣起

讀者之所以可以在網際網路輕易連結不同的數位檔案資源，是由瀏覽器透過 URL (Uniform Resource Locators) 指定連結的資源所在位置，因此換言之，URL 是 WWW 運作的基礎。由於 URL 指定資源的方式，是先指定主機，再指定路徑然後再對特定檔案進行存取，也就是針對資源所貯存的實體位置描述；這種方式稱為絕對位置命名 (Absolute Naming Address)。一旦主機變更或檔案更名時，從其他的網頁透過原來的連結就會無法找到原有的資源。我們經常在網頁瀏覽時可以看到 Http v1.0/1.1 error 404 這樣的訊息，就是表示所指定的連結位址找不到該資源。

相同的數位資源可能存在不同的實體位置，使用者通常不會在意它實際存在的位置，只要找到這個內容就好了，因此需要一個可以說明這兩個不同的命名具有相同內容的機制。用電話隨身碼的概念來解釋這個議題很容易了解；當使用者同時有兩隻電話時，當你要找這個人時，您必定要確知對方這兩支的電話號碼，當其中一支不通時，您可以打另一支。如果您不知道另一隻號碼時，你就找不到對方。隨身碼的方式是不論您申請了多少支不同門號的電話（現有電話門號就是電話的絕對位址命名的方式）都可選用一個好記的號碼或代表名稱代表您個人的隨身碼，方便友人找到您。友人只要記您的隨身碼就可以不必記憶您申請所有電話的絕對號碼。這個隨身碼或是代表號，對擁有者而言就是一個可以間接定址號碼，可藉這個號碼找到擁有者這個實體。而如何讓隨身碼好記，最好的方法就是讓取碼有固定的規則，同時可以從號碼直接解讀一些內容上的意義。這個方法在商業電話上的應用經常可見，如 28825252（餓爸爸餓我餓我餓），就是運用諧音達成助憶的典型範例。

再以行動電話的電話簿例子來說明，只要記得您朋友的小名就可以直接撥出預存的號碼，小名可以固定，他的電話如果便更了只要更改號碼與小名間的對應關係，所以即使他換了電話更新一次電話簿之後，您還是可以直接以小名撥出正確的電話。另外也可以存在多個小名對應多支電話號碼的情形。也就是說，實體存檔的檔名具唯一性，但是應用的方式是可以容許一對多的情形存在，這些就是間接定址命名最大的優點。

如果網路資源的命名規則上能夠做到這一點，除了可以任易更換實體的貯存位置而不影響應用程式或網頁的連結之外。此外，檔案的名稱經常變更，使用絕對名稱的直接定址方式，在網際網路的資源不斷地增加時，出現了一些應用上的困擾。因為資源太多了，檔名太過複雜沒有助記的規則，使用者很容易找不到資源所在，如何讓檔案名稱能夠更具永久性，如同前述的電話隨身碼概念，使其不會受到外界貯存體的變更時就必須跟著改變，在大量的網路資源應用時會方便許多。並且，在當檔案所在的實體貯存路徑描述錯誤時，還可以從檔案名稱來反推其所在位址，這樣一來對大量的數位檔案資源管理帶來許多方便性。

基本上，電腦檔案資源管理在作業系統的概念是以一個唯一的識別碼 (Identifier)

來區別不同的檔案內容，藉由這個識別碼來取出該資源，通常就是以檔名（Filename）做為貯存資料內容的識別碼。但是檔名只有在同一主機內（同一作業系統）內或是指定的路徑範圍內有唯一性。例如，在同一個硬碟內同一路徑下不能有相同的檔名存在，遇到使用者命相同的名字時，通常會自動加上其他字元去區別兩者是不同的檔案。一旦這個檔案資源對網際網路環境釋出時，檔名往往無法作為唯一的識別碼，因為在不同的主機中可能使用相同的檔名。此時用以區別資源的方式就是加上其所在的實體位址，它包含主機名、路徑名再加上檔名來判斷所要的網路資源所在。URL（Uniform Resource Locator）的解決方式就是以前述三項名稱（hostname、path、filename）再加上取用這個檔案的通訊協定（http、ftp、gopher 等）。但是 URL 除了主機名稱是以間接位址定命，主機名以下的路徑及檔案都依照原有的直接定址的方式來決定位置，所以 URL 還是屬於直接定址的使用絕對命名的方式。所以一旦主機變更貯存路徑或檔案名稱，使用原來網頁中 URL 的連結敘述就無法正確定位。因此使用 URL 作為唯一的檔案識別方式有實質上的問題。

參、統一數位資源命名的基本概念

由於統一資源命名及間接定址的概念、理論基礎及命名控制的理念很容易讓人接受，並且有實際上的應用需要，所以在不同的應用領域中，有人提出各種不同領域的統一命名系統；如出版界的 DOI（Digital Object Identifier），期刊控制的 SICI 等都是不同應用領域的統一命名系統。在網際網路中最廣為人知而全球奉行的統一命名系統就是網域名稱 DNS（Domain Name Server）系統。網域的定名是以階層式的領域名稱對應實體絕對的 IP 位址。IP 位址在網際網路上是區別電腦設備唯一識別碼，理論上在網際網路的範圍中必須各自獨立，不同的設備使用重複的 IP 位址；也就是 IP 位址是網路的絕對定址命名方式，而網域名（Domain Name）就是以間接的方式完成的命名，應用上使用者只要記住網域名即可，不需直接使用 IP 位址，因為選用有意義具助記性的網域名取代難以記憶的 IP 位址比較方便。而 DNS 是透過階層式管控的機制，讓選用的網域名具有全域性（Global）的唯一特性，這就是數位資源選用間接命名的最佳範例。

傳統上電腦是使用檔名（Filename）區別資源內容，也就是以檔名作資源的識別碼。如果可以從檔案的命名中直接了解資源的所在及意涵，就可以使用者及管理者在應用資源的速度加快，如何檢索過程及索引的製作會更有效率。基本上，網路資源命名（Naming）的目的有三：一為識別（identification），二為說明所在（location），三為可以具有助記性（memoric）及語意性（semantics）（註 1·參考 RFC2276）。檔案依據這樣的原則命名去命名，目的就是希望當使用者看到檔名時就可以根據這個命名原則對該案的內容或屬性有初步的了解，如此可以減少系統及人工作業的時程。如果能將全球的網路資源選用具一致性資源命名方式，建立出一組命名組合的原則同時，可以將其

其資源實體貯存的位址或識別碼相互對照，這樣可以解決大量資料應用的問題。

所以 1993 年由網際網路協會 (IETF) 統一網路資源名稱的計劃 (URN, uniform resource names) 就是希望採行一種使用間接定名的方式去統一數位檔案資源名稱機制，徹底解決指定連接網路資源的問題。其不僅是指定其所在位址而已，而在使用間接的定址方式去指定網路資源，將網路資源的名稱和資源實體存在位址相互獨立，中間透過中介的命定址伺服機制去轉換實體的名稱及位址，這樣的定址觀念形成了 RFC1737 (註 2)。從 RFC 1737 希望能訂出一套可以永遠持續運用，不會輕易受到貯存媒體變更影響的命名機制。RFC1737 所定義的統一命名機制應該符合下列八項需求：

1. 全域性：是以一個完整的網域環境為應用範圍，而非局限在某個區域，因此所使用的名稱，不因所在地點不同而有不同意義，在任何地點均需有相同意義。
2. 唯一性：二件不同的資源不可能使用相同的名稱，同一名稱不得重複使用。
3. 永久性：選用的統一名稱是永久有效的，名稱存在的效期甚至比所指向之資源實體更持久。
4. 包容性：可以為目前所有網路上出現的不同資源來命名。
5. 相容性：必須支援現有的命名系統並滿足舊有命名系統的需求。
6. 延展性：必須具有延展性，以提供未來發展。
7. 獨立性：命名方式與解譯系統之間應該相互獨立，命名方式不會被特定的解譯系統所限制。
8. 解譯性：能將統合的資源名稱轉換為實體網路資源位址 URL。

肆、識別統一網路資源的方式

對網際網路資源的檢索而言 URL 的出現其實已經是相當進步的方式，不過經過這幾年的應用，逐一發現了它的不足之處。最主要是因為在 URL 組成的結構中主機名以下的路徑及檔案都是針對資源實體的絕對存檔位址（命名），使得 URL 的可以因應環境改變應用範圍的活動力及可攜性（Portability）變低。所以 URI 工作小組（Uniform Resource Identifier working group, URI-WG）提出了 URN（Uniform Resource Names）去進行永久性（persistent）、不受限實體資源所在地（location）的網路資源識別方式。但是如何去架構這樣的命名解譯機制卻很令人傷腦筋，如能配合現有的機制理論上是最簡便的方式。URL 的前半部主機名是被 DNS 所決定的，所以利用 DNS 系統是最容易被想到的一種方式，如果使用 DNS 作為命名解譯的架構因為 DNS 已經廣泛佈署在全球各地也廣為眾人了解，所以選用 DNS 作為命名的解決管道可想而知在推展上困難較低。但是若直接用 DNS 來進行統一的名稱的識別機制（Uniform Resources Identification）有下列幾個問題：

1. DNS 並無意去控制太多的記錄。

2. DNS 的資料有長度上限制，並不適合作為編目用的資訊。
3. 網域名稱在性質上也不適合作為唯一的資源名稱。

如果可以將 DNS 和 URL 結合的方式加以改善，應該可以解決統一資源名稱在應用上的問題。所以 RFC2168（註 3）提出了 DNS（Domain Name System）結合 URLs 的解決方案，這兩個系統結合起來成為一個網路資源名稱的識別系統，稱之為 Uniform Resource Identifiers（URI）。URN 建立出第一個資源命名解譯服務程式（RDS resolver discovery service）其是建置在一種新的 DNS 之上的應用程式，稱為 NAPTR（Naming Authority PointeR）它提供了 URI 對應成 URL 的規則。藉由變更對應規則的方式，我們就可以變更解譯的主機。NAPTR 提供了其他的命名系統得以納入 URN 的系統之下；它提供了名稱分派系統（Name Assignment System）及解譯協定系統（Resolution Protocol System），它允許多重服務如（Name to Location, Name to Description, Name to Resource 等）。

因此要建立一個名稱對應的解譯資料庫才能讓間接定名的方式落實，而這個資料庫可以用來提供 URN 識別資源所用的相關資訊，這些資訊包含資源的所在地（Location）、書目描述資料及資源本身。因為是以 URL 作為存取描述的方式，故便於瀏覽器的程式使用。這個特殊功能的資料庫便是解譯器（Resolver）。

伍、命名管控系統的設計

前面述及 URL 允許網際網路的資源以 DNS 及資源所在位置名來組合資源命名（local name），而所在位置名稱包含所在位置檔案路徑（local file path）或指向（reference）該所在地其的服務程式（如 CGI 程式），所以可以用 HTTP 指向透過 NFS（Network File System）等其他網路服務方式取得資源。但是這裡並沒有解決實體貯存檔名本身的唯一性問題。URL 規則命名中只有 DNS 具備唯一性，其後的路徑及檔名描述並不具備唯一性，必須使路徑及檔名具備唯一性才能使完整的 URL 也具有唯一性。

假設全球的檔案資源以流水編碼方式進行，流水號的優點就是具有唯一性；當使用者開啓新檔時先向命名管控中心軟體取號作為實體存檔檔名，如此就不會發生檔名相同的衝突。但流水號的缺點是不易記，從檔名也看不出代表內容的任何意義，如果可以再給一個易記的識別碼作為對外應用的名稱，這樣的命名機制可以解決大部分的命名衝突。困難在這個易記的識別碼的編碼該如何訂定，就有許多的考量。首先考量的就是現存編碼系統的問題。如果現存的編碼系統已經具有全域（Global）唯一性的考量且廣為人知時，如能納入統一資源編號的取號規則，會使得統一命名的方式更讓人接受，減少對新系統的適應性。現存命名系統（Legacy Naming or Numbering System）或是編碼系統，如果它具有存在的唯一性時，而這個編號系統對資源識別能力（Identification）又

可以符合部分我們對統合性網路資源的需求，將其整併進入統一網路資源命名系統的一部分是最理想不過的方式。一則可以讓使用者繼承原有編碼或命名的習慣，同時也可以簡化新命名系統編碼上細節的設計。例如國際標準書號 ISBN 的納入 URN 就是一個很好應用的方式。

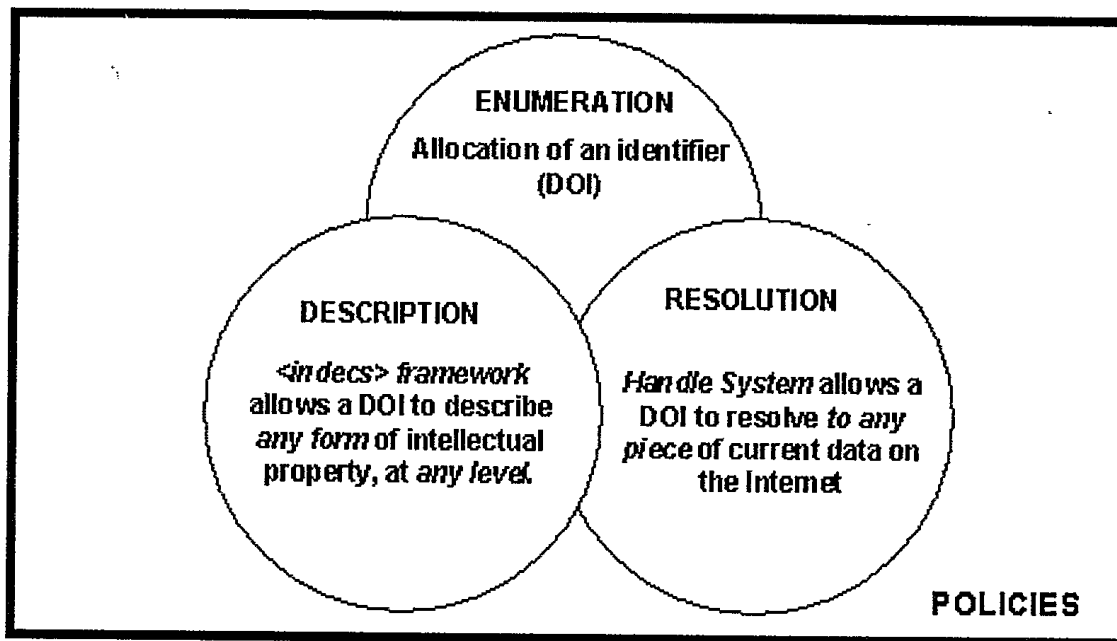
陸、如何維持資源命名的統一

訂定一套統一資源系統，除了要制定一個統一的命名原則，並且這個命名原則必須具備前述命名的八大項原則（RFC1737）（註 4）；此外，必須有一個有效的控管的機制，才能使統一命名得以落實。控管機制的目的主要在確保取號的過程不會重覆，同時這個機制可以解譯所選用的資源名稱所代表的意義。所謂解譯功能（Resolution）的意義是對任一個網路服務程式送出的資源識別碼給命名的控管機制，然後接受相關於這個識別碼的訊息回傳到這發出的識別碼之上，這樣的動作就稱為解譯。由於網際網路上的各項資源普遍不具有永久性，使用者經常會看 Http Error 404 這樣的訊息，這表示所指定的網頁位址無法找到，意即對所指定的網頁之 URL 不能被解譯。

柒、命名系統與應用環境的關係

命名系統的運作考量現存編碼系統就是為了基於現在各界現存實際應用的模式，所以會因為不同的應用領域會有不同的需要。美國出版者協會為了處理出版品資源的統一命名問題制定了一套系統稱之為 DOI（Digital Object Identifier）系統。DOI 系統將自己定位成一個整合性的系統結構，不僅是單純的命名系統也就是把控管的機制和實際運作的考量納入。DOI 系統分成四大項的組成項目：

1. 給號及取號（Enumeration）：付予每一個網路資源一個識別碼。
2. 由識別碼解譯出描述資源的敘述（Description）：使用 DOI 識別碼產生適當的詮釋資料(Metadata)描述。
3. 由識別碼解譯出適當的對應動作（Resolution）：能夠透過 DOI 系統解譯出的資訊而使得所用的識別碼具有作某些動作的能力。
4. 政策設定（Policies）：用來控制這個系統運作的規則，也就是軟體參數的定義及設定的方式。



(資料來源 http://www.doi.org/handbook_2000/components.html) (註 5)

DOI 和其他的命名系統最大的不同處是在於識別碼本身所含的動作功能。DOI 識別碼是可以被解譯的，它可指引使用者到任何網際網路可以取得的任何資料。但這個不意味著 DOI 必須解譯其所指向的資源本身，雖然可能在某些情況之下 DOI 是可以對資源本身作若干的解譯，反之則不然。另外詮釋資料 (Metadata) 是 DOI 系統相當必要的一項元件，對所有的已登錄的 DOI 識別碼都是不可缺的必備項目。其所謂核心的詮釋資料 (Kernel Metadata) 是儘可能縮小它的範圍，同樣地對任何可以使用 DOI 來識別的資源實體都適用。這些說明解譯系統的功能除了前項 RFC 律定的規則之外，還有相當多功能上的發展空間。

因為 DOI 它主要是在解決出版商在線上引用的問題，因此控管系統在設計勢必會遷就線上引用的方便性為最主要的考量。反過來說當出版商需要使用統一命名系統時，DOI 會比較先考慮。因此它在電子商務及智慧財產權 (Intellectual Property) 有多方面的應用考量。DOI 最常使用的方式是用來作更新期刊論文的引用工具，出版商可以指定 DOI 到適當的期刊論文，所以它們就可以輕易地在書目資料中引用。這樣的服務就發展成了連結參考引用 (link reference citation)，這些參考書目就可以線上處理引用的作業。也因為這樣的機制可以把相關的文獻放在不同的伺服器，也可以由不同的出版商來進行出版。Link.Openly 就是一個以 S-Link-S (Scholarly Link Specification Framework) 為基礎的系統，它是以 XML 來發展用來簡化出版商間的相互參照連結。

IDF 則是另外一組發展的模式，主要是以目次內容 (Content industries) 業者發展出來的 DOI 應用，其主要是針對許多不同型態的智慧財產表示方式。以 Metadata 的描述而言可以清楚的了解，描述期刊論文的 DTD 和描述樂譜或描述照片的方式必然是不

相同的。也因為在 Metadata 詮釋資料的項目在目的上不一樣，所以也有可能需要用不同的應用程式來作對的處理，所以 DOI 有所謂的 DOI-AP Profile 去記錄這個部分。每一個資源在向 DOI 記錄之後同時也會由 DOI 系統產出部分的詮釋資料核心（Kernel Metadata）。這種 DOI-AP profile 的設計方式提供應用程式使用彈性，也使得統一命名方式應用的可行性提高。

捌、階層性的控制機制

控制機制的階層結構設計，爲了就是分層負責，讓控管的範圍分散到不同主機，減輕集中單一主機的工作負荷。所以 URN 和 DOI 這兩大系統都是採用階層控管的方式，不論命名對照及解譯都是採用階層式的處理，就如同 DNS 的管理網域名稱的方式一樣。就名稱解譯的觀念和來看 DOI 系統使用的控管機制 Handle system 和 DNS 的名稱解譯很類似，兩者最大的不同在於 DNS 並不是定位成全功能（General Purpose）的資源命名系統，它只對主機及網域的名稱來作控管而已。但是 Handle 系統則是希望的是一個全方面的資源命名解譯系統。又從控管作業的角度來看 DNS 一般都是定位在網路管理者這一個等級的作業，除了網路系統管理者之外，沒有其他的管道針對這個名稱來進行管理的動作。而 Handle 系統在這幾點上剛好和 DNS 設計理念相反，它是以 DNS 系統所不能爲的部分開始思考進行設計，一開始就是設計成可以針對大量實體命名管理來發展，而這一項作業卻視爲在一般應用程式層級的管理作業動作，而不需要像 DNS 需由網路管理者這一層來管理。因此命名控管機制階級的分級方式和成員多寡有關，故圖書館在選擇命名系統時也應該考慮所選的命名系統可能影響系統管理工作的分配，這是需要預先設計的部分。

玖、結論

統一資源的命名對大量網路資料的管理是有其必要的。不論控管的機制是否爲加入現存的系統（如 DOI 及 URN 等）與否，自行訂立一套規則對實際管理作業而言都有相當的功效。在現在圖書館強調館際資源的整合運用之際，如果圖書館能夠確立一套統一的命名機制，應該對館際間的數位資源的相互應用有實質的效用，可以藉此建立分工貯存數位資料的基礎，可以增加未來數位資料分享的可行性，如此可以突破目前圖書館界以書目資料分享爲主的數位資料互動的方式，更進一步到廣泛數位資源的互通。

但是建立統一資源命名不是沒有困難的，首先在於建立控管的機制相當複雜，如果獨立控管機制，系統設計的花費可能相當龐大，圖書館界可能缺乏設計人力及經費。若加入國外的機制如 DOI 系統，其以單筆資料計價的方式對目前國內而言尚不可行。移植其控管軟體在台自行建置運作機制也是一種方式，不過該套裝軟體的價格目前也仍

然太高。選擇以 DNS 的運作機制來配合也有實質上的困難，一來增加現有 DNS 伺服器的工作量，可能會影響目前 DNS 的作業。另外 URN 的機制比較鬆散，將工作委由現有 DNS 管理人來處理，可能因為並非每一個單位都有大量數位資源管理的需求而認為沒有實質的必要。個人以為如何建立完整的作業機制是相當值得同道們進一步思考的問題，希望先進們可以想出一條比較可行的路，讓國內圖書館界在數位資料的管理方面可以向前跨一大步。

註釋

- 註 1：<http://www.rfc-editor.org/rfc.html>, <ftp://ftp.isi.edu/in-notes/rfc2276.txt>
 註 2：<http://www.rfc-editor.org/rfc.html>, <ftp://ftp.isi.edu/in-notes/rfc1737.txt>
 註 3：<http://www.rfc-editor.org/rfc.html>, <ftp://ftp.isi.edu/in-notes/rfc2168.txt>
 註 4：<http://www.rfc-editor.org/rfc.html>, <ftp://ftp.isi.edu/in-notes/rfc1737.txt>
 註 5：http://www.doi.org/handbook_2000/components.html

參考資料

1. Clifford Lynch "Identifier and Their Role in Networked Information Applications" <http://cause-www.colorado.edu/ir/library/html/cem9743.html> 1997-10
2. Lloyd A. Davidson and Kimberly Douglas, "Promise and Problems for Scholarly Publishing", <http://www.press.umich.edu/jep/04-02/davidson.html>
3. Juha Hakala and Truagott Koch. URN creation tool, Nordic Metadata Project. Help, "URN User Guide", 1998-02-23, <http://www.lub.lu.se/metadata/URN-help.html>
4. Sam X. Sun, Laurance Lannom, "Handle Ssystem Overview" <http://www.ietf.org/internet-drafts/draft-sun-handle-system-06.txt> 2001-4, <http://www.handle.net/overview-current.html>
5. Andy Powell, "Resolving DOI Based Urns Using Squid" 1998, 6: D-Lib Magazine <http://www.dlib.org/dlib/june98/06powell.html>
6. John. S. Erickson, "The Role of Metadata Supply Chains in DOI-Based, Value-added Services", <http://www.ybp.com./yps/papers/doiservicesapr99.htm>, 1999-4
7. DOI Handbook http://www.doi.org/handbook_200/toc.html