

## 國家文化資料庫詮釋資料格式設計理念

陳昭珍

國立臺灣師範大學圖書資訊學研究所教授

### 一、國家文化資料庫建立緣起

為配合政府推動知識經濟方案，文建會於八十九年十一月召開之「全國知識經濟會議」上提出「建置文化藝術知識庫計畫」並獲通過。為貫徹此施政目標，擴大執行效果，並執行九十年元月「全國第六次科學技術發展會議」結論中文建會應辦理之「建置國家影音資料庫」之決議，而將「建置文化藝術知識庫計畫」更名為「建置網路國家文化藝術院」計畫。在此計畫下，文建會開始建置「國家文化資料庫」系統。此系統主要之任務有：

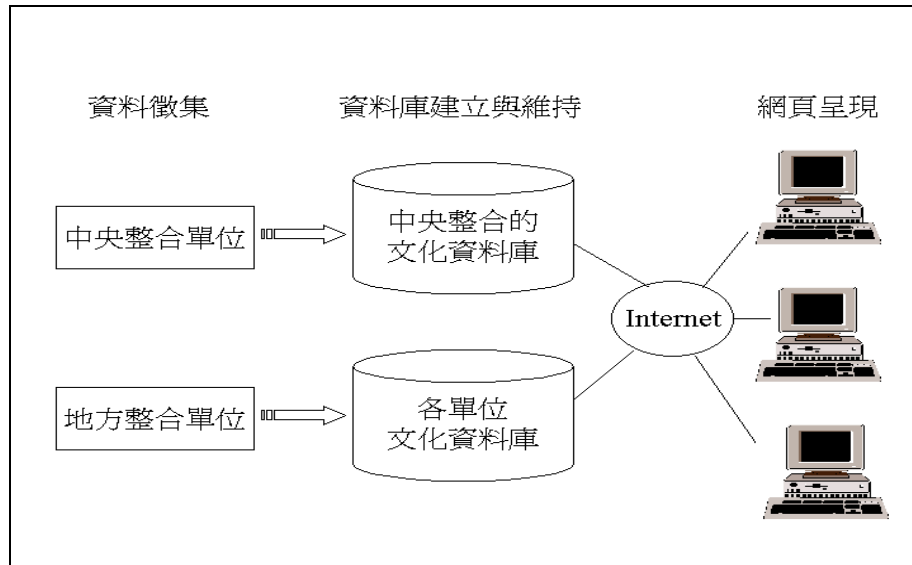
1. 將各文化單位徵集到的文化資源加以描述及數位化，再轉載至文建會之國家文化資料庫，集中運用；
2. 文建會建置及維護國家文化資料庫之運作；
3. 透過網際網路使全民共享國家文化資源。

國家文化資料庫開始建立之初，所包括的資料類型有：音樂資料、美術資料、戲劇資料、舞蹈資料、傳統藝術資料、文學資料、電影資料、建築資料等八大類。其原始資料的來源有兩種：一為已經由中央與地方文化機關／單位收藏的公藏文化資料，二為尚藏於民間的文化資料。

國家文化資料庫之基本架構包括中央整合資料庫與分散典藏資料庫二種：(如圖一所示)

1. 中央整合資料庫(文建會會本部)：中央整合資料庫是指全國性的管理系統，收集各地送來的數位化文化資料，作為國家數位文化資料保存中心，並提供未來各種不同的用途。收藏由分散典藏資料庫所送來的數位化文化資料，並供全國民眾檢索使用。資料庫亦可接受來自於網際網路的查詢請求，透過資料庫的查詢機制，提供外界各種不同的文化資料相關資訊。

2. 分散度藏資料庫：分散度藏資料庫是由文建會轄下各機關／單位，及縣級文化單位，建置各類型文化資料的數位資料庫，將文化資料予以適當的數位化處理，並賦予詮釋資料，形成一個具有強烈語意描述的數位化文物物件。並透過網際網路傳遞的形式，將各式文物的數位化資料上載到中央整合資料庫。



圖一：國家文化資料庫概念架構

## 二、國家文化資料庫詮釋資料格式設計理念

為執行國家文化資料庫計畫，除了必需設計可以集中管理的資訊系統外，為描述中央與地方文化典藏之數位化檔案，尚須建立各種詮釋資料格式及描述規範。以下即說明詮釋資料之意義與功能，國家文化資料庫詮釋資料設計情境及使用規範：

### (一) 詮釋資料的意義主要有下列兩點：

#### 1. data about data

將作品數位化後(如一幅畫之影像檔、一首歌轉錄之聲音檔等)放在電腦資訊系統中，如果沒有文字資料的描述，將無法被檢索查詢，也無從了解該數位資料相關的資訊，如作者是誰、作品的名稱為何、何時何地所創作、作品的特色為何、與其他作品的關係，甚至其數位化檔案的特性(如壓縮格式、解析度等)若無文字的描述，也無從得知。所以，簡單的說，所謂的詮釋資料(metadata)，就是以文字來說明原始作品或數位化資料相關屬性的資料，也就是所謂的“data about data”，如圖書館的書目記錄(MARC records)，即為一種 metadata。

#### 2. structured data about data

metadata 最重要的概念，即其結構性，更進一步可將 metadata 定義為“structured data about data”，使得採用 metadata 做組織資訊的方式和全文索引(full-text indexing)有所區隔。一般而言，詮釋資料格式大都為欄位化的格式來描述作品，而非以一段很長的全文來說明作品，若為全文文字，在系統中也會以標誌(mark-up)技術來處理，將非結構性的全文資料轉為結構性的資料，以提高檢索的精確度與效率，及全文資料的加值價值。

## (二) 詮釋資料的功能

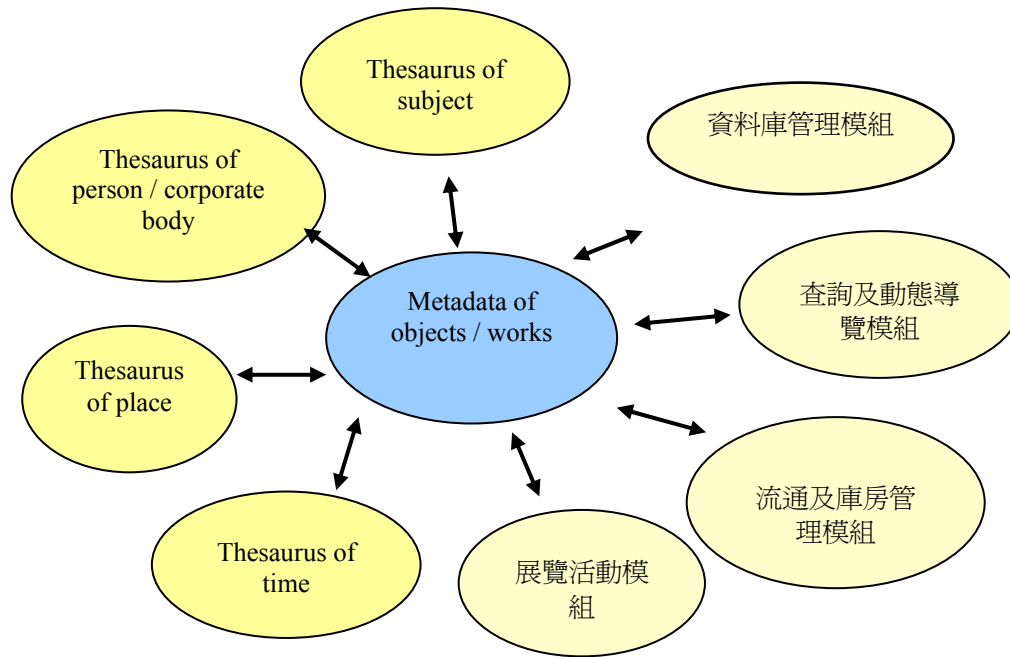
詮釋資料之功能具體而言包括下列各項：

1. 「描述」：概述資料之內涵、說明資料之歷史(如說明其原始資料為何，及其他的改變)、資料之聯絡人(如擁有者)、資料使用的合法情況...等。
3. 「檢索」：讓使用者查到該資料
4. 「選擇」：讓使用者決定該資料是否是他所要的
5. 「管理」：控制資料的管理，如禁止兒童使用某類資料
6. 「應用」：指示應如何解譯該資料(例如說明資料之格式、編碼、加密的情況)；指示該資料與其他資源的關係

## (三) 國家文化資料庫詮釋資料格式設計情境

完整的國家文化資料庫所需建立的詮釋資料格式包括：作品(收藏品)之詮釋資料格式、藝術家(人名)詮釋資料格式、主題語詞詮釋資料格式、地名詮釋資料格式、時代詮釋資料格式、展覽及活動詮釋資料格式。這些詮釋資料主要乃從文化資訊保存與分享的角度而建立，其中藝術家、主題詮釋資料、地名、時代，除可單獨被查尋使用外，亦可作為作品詮釋資料中，人名及主題詞的權威控制檔；而地名、時代詮釋資料的建立，則另有助於文化資源時間軸及空間軸的建立。

除了文化相關資訊的描述外，以一個文化機構而言，實體或數位文化資產可能會被借出、展覽、辦活動，因此也需設計展覽及活動之詮釋資料格式，以及借出人等檔案，如圖二所示：



圖二：詮釋資料類型及系統模組關係圖

由於國家文化資料庫主要是一集中式的資料庫，各單位向文建會申請的計畫，其詮釋資料及數位檔都會存放一份在文建會的系統，因此國家文化資料庫的所訂定的，主要是為各單位繳交到國家文化資料庫之詮釋資料格式，而不是為各單位內部作業及系統設計而用，換言之，國家文化資料庫所訂定的詮釋格式是交換格式，而不是各機關內部使用的格式。由於各主題資料的屬性不同，國家文化資料庫所設計的詮釋資料格式包括：音樂資料詮釋資料格式、美術資料詮釋資料格式、戲劇資料詮釋資料格式、舞蹈資料詮釋資料格式、傳統資料詮釋資料格式、文學資料詮釋資料格式、電影資料詮釋資料格式、建築資料詮釋資料格式、老照片詮釋資料格式、古文書詮釋資料格式等；而在支援檔部份則有：藝術家詮釋資料格式、主題語詞詮釋資料格式、地名詮釋資料格式、時代詮釋資料格式、展覽及活動詮釋資料格式。

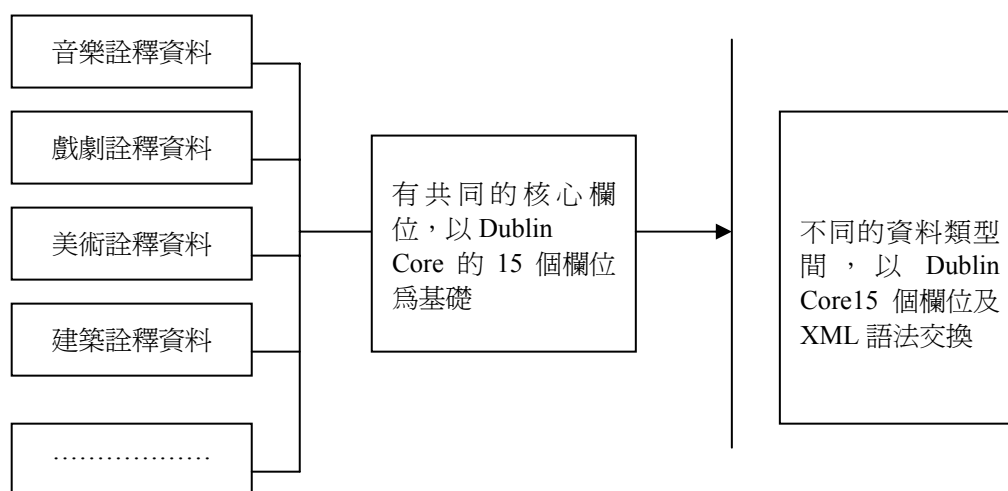
#### (四) 國家文化資料庫的設計原則

國家文化資料庫詮釋資料的設計原則包括下列諸項：

##### 1. 以 Dublin Core 為核心欄位

詮釋資料包括語意及語法二部份，國家文化資料庫所使用的語法為 XML，而語意部份，為顧及各類資料之特色與屬性，我們分析了各主題資料之特性及使用需求，設計符合各主題特性的詮釋資料欄位，不過為使各類詮釋資料格式得以互通(interoperability)，各種詮釋資料集中到文建會的國家文化資料庫後，其處理情況不致太複雜，無法整合，因此我們採用簡單的 Dublin Core(DC) 15 個欄位作

為詮釋資料之核心欄位，也是國家文化資料庫與其他系統統的交換欄位，各主題詮釋資料主要乃就這 15 個欄位加以擴充，如下圖所示：



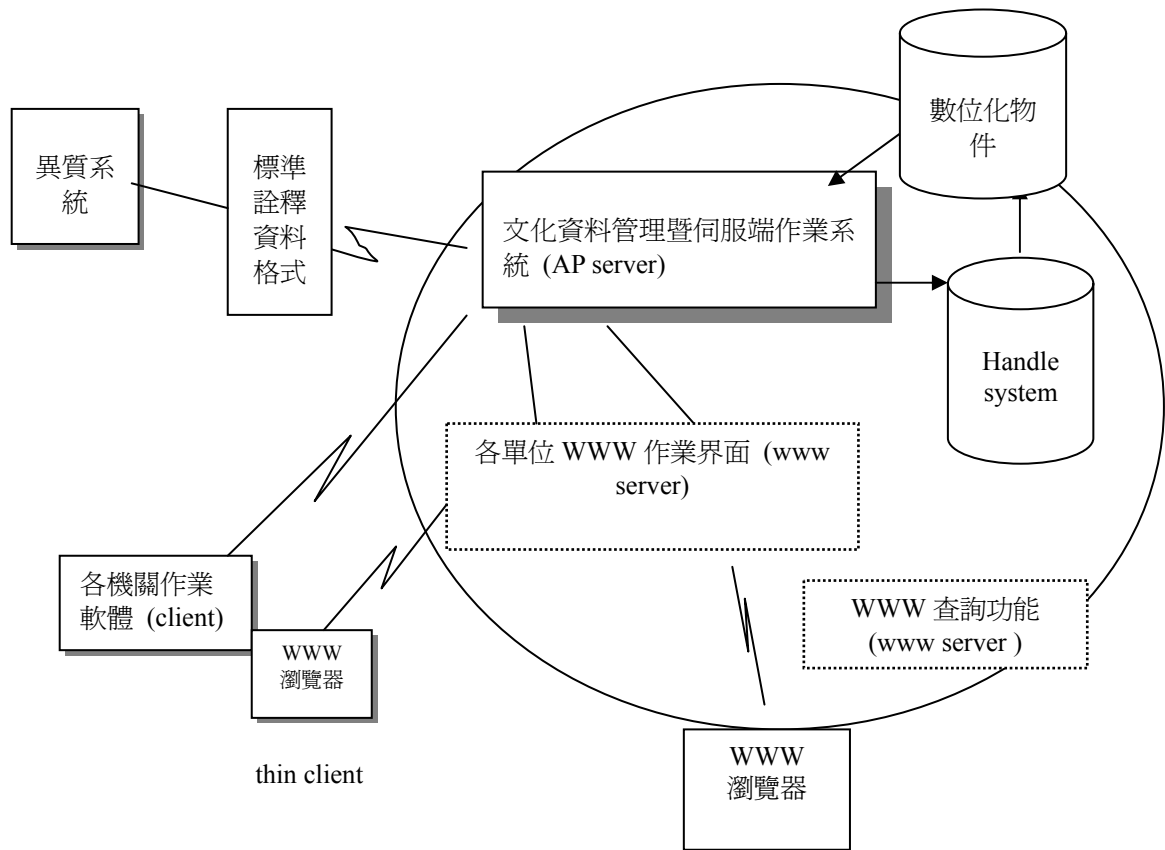
圖三：以 DC 為核心之設計理念

## 2. 必備與非必備欄位

在 Dublin Core 並未規定那一個欄位一定要填寫，但在實際的情境下，一定要定義那些欄位是必備，那些欄位是非必備。詮釋資料格式中有少數欄位是必備欄，一定要填寫資料，如作品名稱、資料類型等，若各單位之系統中原來無此欄位，轉出後必須再補充；大部份的欄位是非必備(optional)，端視描述者對該資料掌握與了解的狀況而填寫。

## 3. 詮釋資料上載原則

詮釋資料就建檔、查詢、顯示與交換而言，可有不同之呈現，但在底層實則皆來自最大集合之欄位。各單位原則上可參考文建會所提供之各主題詮釋資料格式設計自己系統的建檔、查詢、顯示欄位，但交換格式一定需遵守文建會要求之各主題詮釋資料格式。換言之，若各單位已有資訊系統，其資料的描述欄位，除非修改系統否則皆已確定，因此已無設計各單位詮釋資料格式之需求，而只需另外擴充詮釋資料的轉出功能，並遵守文建會要求之各主題詮釋資料格式，以 XML 語法將資料上傳至文建會。若各單位尚未有資訊系統，則建議採用文建會所頒布之各主題詮釋資料格式之欄位建檔，則所描述之欄位及其完整性將與文建會所要求之交換格式一致。(如圖四所示)



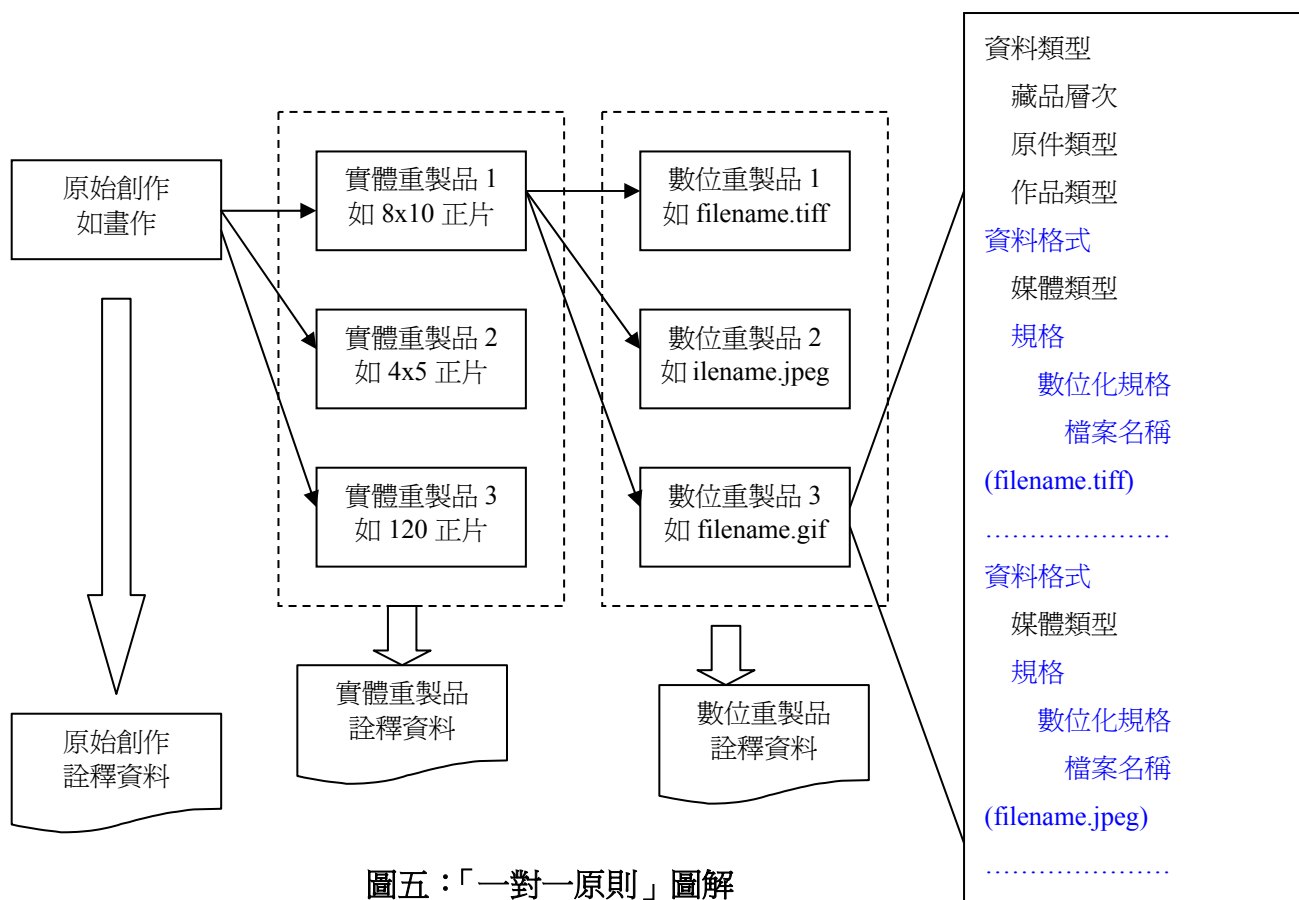
圖四：國家文化資料庫系統架構

#### 4. 一對一原則

- (1) 資料之描述原則上採一對一(121)的原則，即一筆詮釋資料中只描述一種作品，以增加詮釋資料的彈性，並降低描述工作的複雜度，詳見圖五。
- (2) 「一對一原則」是目前國際詮釋資料領域所遵循的原則，不論是原作、實體之重製品與數位化之重製品，每一個物件最好以一筆詮釋資料進行描述。
- (3) 假設有一原始創作，有一實體重製品(如正片) 與一數位化重製品(如數位檔案)，依「一對一原則」會對應到三筆詮釋資料(如上圖)，這三筆詮釋資料在 format(資料格式)欄位的描述會不同；在物件內涵相關欄位(如 subject、description...) 的描述，則大致相同。
- (4) 從典藏管理的角度，原作、實體重製品(如正片、微縮片)、與數位化重製品(數位檔案) 均為應列管的財產，且可能由不同的單位典藏，因此依「一對一原則」對應到不同的詮釋資料有其必要性。
- (5) 原則上原始創作、實體重製品與數位化重製品一定是對應到不同的詮

釋資料。至於多個對應到相同原作的實體重製品，若其內涵相同，僅有格式不同，則只需著錄一筆詮釋資料，在 format 欄位重複多次；若是內涵亦不同，則須著錄多筆詮釋資料。

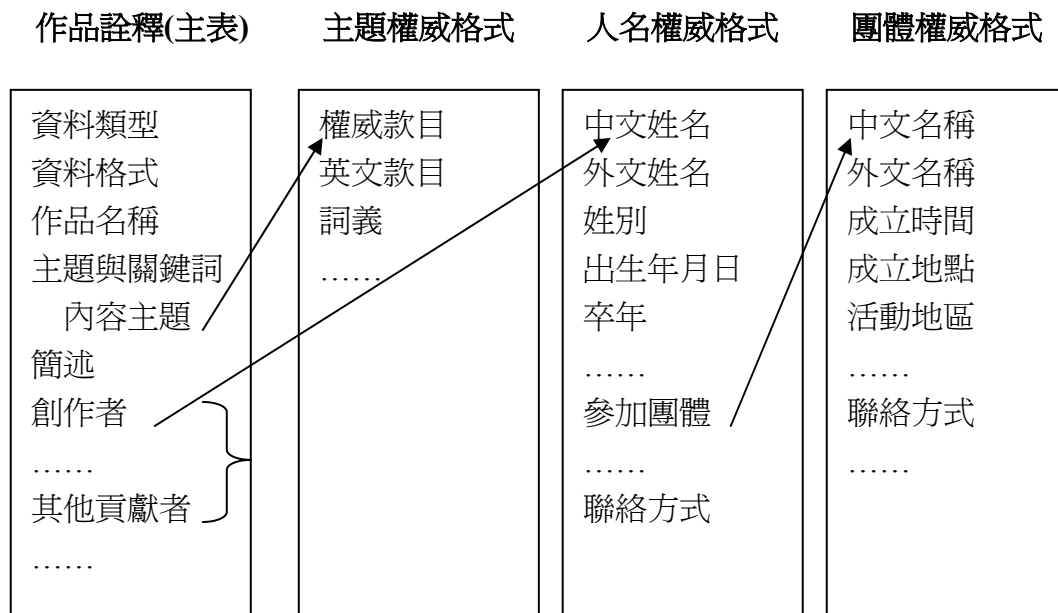
- (6) 例如有一原始創作，拍攝成五張不同規格的正片，每張正片掃描製成五種不同規格的數位檔案，依照徹底的「一對一原則」會有 1 筆描述原始創作的詮釋資料、5 筆描述正片的詮釋資料、25 筆描述數位檔案的詮釋資料。為降低專案執行的複雜度，若五張正片的內涵相同，只是格式不同，可只做一筆描述正片的詮釋資料，在 format 欄位重複五次，分別描述其正片規格；以及只做一筆描述數位檔案的詮釋資料，在 format 欄位重複 25 次，分別描述其數位化規格。
- (7) 各主題詮釋資料皆可能描述不同類型之數位化資料(如文字、影像、聲音、視訊等數位化資料)，但由於要填寫的檔案規格較複雜，因此將數位化規格(DigitalScale)相關欄位由主表中抽出，獨立成附表，以提高可讀性。事實上數位化規格附表中的欄位是包含在各類詮釋資料之 format/scale 欄位之下。



圖五：「一對一原則」圖解

## 5. 採用代碼檔與權威檔維持資料品質

為希望各單位對部份欄位值有一致的填寫方式，以便整合資源，並維持資料品質，所以另外設計一些代碼表，如 format/medium(媒體類型代碼表)、fileType(檔案類型代碼表)，作為填寫時之參考。除了代碼表外，為了提昇檢索的品質，必需另建各種支援檔，這些資源檔包括：主題權威檔(或稱為索引典)、人名權威檔、團體名稱權威檔、地名權威檔...等，在國家文化資料庫中，為了資料的彈性，這些支援檔的詮釋資料格式都是分開來的，在系統中可以互相連結，我們也稱為數表結合原則，一個數位化專案的詮釋資料通常是「數表結合」，不太可能將所有需描述的欄位塞在一個表格之中，或以主表即能完成所有的資料詮釋作業。如下圖所示：



圖六：國家文化資料庫數種詮釋資料連結原則

## 6. 國家文化資料庫系統接受各種 metadata formats 的彈性設計

雖然國家文化資料庫主要的詮釋資料格式是採用 Dublin Core 擴充，這是因為集中是資料庫而必需採取的措施，但國家文化資料庫系統卻是一個相當有彈性的系統，只要將任何詮釋資料格式寫成 XML/DTD 或 XML/Schema，系統就能接受該格式，如電影資料所採用的 ECHO 格式、檔案資料所採用的 EAD 格式、文學資料所採用的 TEI 格式，該系統都可接受。



## 參考資料

1. 文建會，「全國文化資料庫」計畫書（台北市：文建會，民國 90 年）。
2. 陳昭珍主持；項潔，陳雪華共同主持。國家文化資料庫系統架構、詮釋資料規範及數位檔案格式研究計畫報告。文建會委託國家圖書館，民國九十年十二月。
3. 文建會國家文化資料庫知識管理系統，<http://km.cca.gov.tw>