

附件

轉置入本部圖書及影音出版資訊網檔案格式

文 化 部
圖 書 及 影 音 出 版 品
數 位 化 及 系 統 建 置
數 位 化 規 格

MOC

1. 交付儲存媒體	1
1.1. 儲存媒體規格.....	1
1.1.1. 圖書類.....	1
1.1.2. 影音類.....	1
2. 出版品數位化規格.....	1
2.1. 圖書資料	1
2.2. 影音資料	1
2.3. 精華片段	2
2.4. OCR 辨識	2
3. 數位化檔案命名原則.....	2
3.1. 圖書類	2
3.1.1. 典藏用	2
3.1.2. 展示用	2
3.1.3. 編碼說明.....	2
3.2. 影音類	2
3.2.1. 典藏用	2
3.2.2. 展示用	2
3.2.3. 精華片段.....	3
3.2.4. 編碼說明.....	3
4. 光碟燒錄作業	3
4.1. 光碟編號命名(LABEL NAME)規則	3
4.2. 目錄(DIRECTORY)存放規則	3
4.2.1. 光碟目錄.....	3
5. 詮釋資料.....	4
5.1. 圖書出版品詮釋資料格式(PUBLICATION METADATA FORMAT)	5
5.2. 影音出版品詮釋資料格式(VIDEO METADATA FORMAT).....	7

1. 交付儲存媒體

1.1. 儲存媒體規格

1.1.1. 圖書類

繳付儲存媒體規格			
圖書出版品典藏檔	圖書出版品 PDF 瀏覽檔	圖書出版品 JPEG 封面圖檔	圖書出版品 OCR 文字檔
單面 4.3GB DVD-R	單面 4.3GB DVD-R 或 120GB 以上硬碟	單面 4.3GB DVD-R 或 120GB 以上硬碟	單面 4.3GB DVD-R 或 120GB 以上硬碟

1.1.2. 影音類

繳付儲存媒體規格		
影音 MPEG2 典藏檔	影音 1500Kbps WMV 檔	影音 1-3 分鐘 WMV 精華檔
單面 4.3GB DVD-R	雙面 4.3GB DVD-R	單面 4.3GB DVD-R 或 120GB 以上硬碟

1.2. 出版品數位化規格

1.2.1 圖書資料

出版品數位化規格					
出版品類型	掃描規格	掃描解析度	掃描模式	檔案大小	轉檔規格(區域展示規格)
美術類作品	1:1	300dpi/inch	tif/RGB/24bit(不壓縮)	約 27MB	1:1 轉 Pdf(一文一檔)
攝影作品	1:1	300dpi/inch	tif/RGB/24bit(不壓縮)	約 27MB	1:1 轉 Pdf(一文一檔)
一般出版品(含圖)	1:1	300dpi/inch	jpeg/RGB/24bit(75%壓縮)	約 2MB	1:1 轉 Pdf(一文一檔)
期刊論文集(含圖)	1:1	300dpi/inch	jpeg/RGB/24bit(75%壓縮)	約 2MB	1:1 轉 Pdf(一文一檔)
一般出版品(文字)	1:1	300dpi/inch	Tif/B&W(ccittG4 壓縮)	約 100k	1:1 轉 Pdf(一文一檔)
期刊論文集(文字)	1:1	300dpi/inch	Tif/B&W(ccittG4 壓縮)	約 100k	1:1 轉 Pdf(一文一檔)
書的封面頁(展示用)			JPG(長邊 800pixel)	約 100k	長邊 800pixel

1.2.2 影音資料

影音數位化規格				
轉檔類型	檔案大小	傳輸 Bit Rates	Video 規格	Audio 規格
MPEG2 (典藏保管規格)	720 x 480 pixel	8 Mbps Video: 7,609,200 bps Audio: 192,000 bps	Aspect Ratio: 4:3/16:9 Frame Rate: 30fps	Layer: Layer2 Mode: Stereo Sample Rate: 44.1/48 Khz
WMV (區域展示規格)	640 x 480 pixel	1500 Kbps Audio: 96 Kbps	Aspect Ratio: 4:3/16:9 Frame Rate: 30fps	Sample Rate: 44.1/48 Khz
WMV (*精華片段規格)	640 x 480 pixel	1500 Kbps Audio: 96 Kbps	Aspect Ratio: 4:3/16:9 Frame Rate: 30fps	Sample Rate: 44.1/48 Khz

1.2.3 精華片段

*精華片段：1500kWMV 檔案中擷取出每小時 3 分鐘的精華片段，不足 20 分鐘的 1500kWMV 檔案，則固定擷取 1 分鐘。

1.2.4 OCR 辨識

將針對 TIFF 的 B&W 與 JPEG 格式進行辨識，辨識結果存成.txt 檔案。

2 數位化檔案命名原則

2.2 圖書類

2.2.1 典藏用

檔案種類	範例	備註
TIFF (RGB)	cca100001-rf-007222519x-rgb-0001-u.tif	
TIFF (B&W)	cca100001-rf-007222519x-bw-0002-u.tif	
JPEG	cca100001-rf-007222519x-0003-u.jpg	

2.2.2 展示用

檔案種類	範例	備註
JPEG	cca100001-rf-007222519x-cp-i.jpg	書的封面
PDF	cca100001-rf-007222519x-0003-i.pdf	
TXT	cca100001-rf-007222519x-0003-i.txt	OCR

2.2.3 編碼說明

cca100001：文建會組織代碼。

rf：書目類別資料。

007222519X：是 ISBN。

rgb/bw：是圖檔的色彩標示。rgb 表全彩；bw 表黑白。

0001：是序號。

-u/-i：用途，-u 表典藏用；-i 表展示用。

tif/jpg/pdf/txt：是副檔名。

2.3 影音類

2.3.1 典藏用

檔案種類	範例	備註
DVD	cca100001-mv-影帶編號-0003-u.mpg	

2.3.2 展示用

檔案種類	範例	備註
WMV	cca100001-mv-影帶編號-0003-m.wmv	

2.3.3 精華片段

檔案種類	範例	備註
WMV	cca100001-mv-影帶編號-0003-t.wmv	

2.3.4 編碼說明

cca100001：文建會組織代碼。

mv：電影類別資料。

影帶編號：需要訂定，將採用“分類號”與“序號”結合的方式編訂。

0003：是序號。

-u/-m/-t 是用途。

mpg/wmv 是副檔名。

3 光碟燒錄作業

3.2 光碟編號命名(LABEL NAME)規則

媒體類型	光碟編號命名規則	說明
300DPI TIF RGB(美術,攝影)	A0000001~A9999999	以 A 開始，後 7 碼序號
300DPI JPG RGB(含圖圖書)	B0000001~B9999999	以 B 開始，後 7 碼序號
300DPI TIF B&W(文字書籍)	C0000001~C9999999	以 C 開始，後 7 碼序號
MPEG2(影音典藏)	D0000001~D9999999	以 D 開始，後 7 碼序號

WMV(區域展示)	E0000001~E9999999	以 E 開始，後 7 碼序號
WMV(精華片段)	F0000001~F9999999	以 F 開始，後 7 碼序號， 或以硬碟交付
PDF(展示用)	G0000001~G9999999	以 G 開始，後 7 碼序號， 或以硬碟交付
JPG(圖書封面 800 pixel)	H0000001~H9999999	以 H 開始，後 7 碼序號， 或以硬碟交付
Metadata(圖書及影音)		
TXT(圖書 OCR)	I0000001~I9999999	以 I 開始，後 7 碼序號， 或以硬碟交付

注意：每片 DVD-R 燒錄容量以不超過 4GB 為原則！

3.3 目錄(DIRECTORY)存放規則

3.3.1 光碟目錄

每個目錄存放最多 1000 個子目錄或檔案。如一片 DVD 存放的檔案數目 1000 以下，不用建立目錄，直接存放在根目錄即可，如果超過 1000 個檔案，需要區分目錄。目錄命名 001,002,003,004,...,999

範例：以 A00000001 這片 DVD 光碟為例，如果存放的檔案個數

(1)檔案數 5 個時,直接放在根目錄

//A00000001/file1

//A00000001/file2

//A00000001/file3

//A00000001/file4

//A00000001/file5

(2)檔案數 2500 個時,建立 3 個子目錄, 分別是 001, 002,003

//A00000001/001/file1

//A00000001/001/file2

...

//A00000001/001/file1000

//A00000001/002/file1001

//A00000001/002/file1002

...

//A00000001/002/file2000

//A00000001/003/file2001

//A00000001/003/file2002

...

//A00000001/003/file2500

4 詮釋資料

1. 以下表格內欄位名稱英文字串後之「?」、「*」、「+」、「空白」為 DTD 所使用之符號，其所代表的意義為：

「?」表非必備且不可重複

「*」表非必備且可重複

「+」表必備且可重複

「空白」表必備且不可重複

2.  表「必備」欄位

3. 詮釋資料內容值著錄規則為：『中文字及英數字者、空白鍵、“-”(dash)符號』一律使用半形；『符號者』一律使用全形。

4.2 圖書出版品詮釋資料格式(Publication Metadata Format)

欄位名稱		欄位說明	範例
Type 作品類型		說明作品的類別。 預設值為「出版品」	出版品
Format 資料格式		說明所描述的物件其媒體類型為何。 ※內容值代碼請參見： http://km.cca.gov.tw/download/rule/format_table.pdf	圖書
CallNumber 索書號		請填入該資料的索書號(分類號+作者號)	980
title 書名	mainTitle 主要書名	該資料的主要名稱，為正式名稱、或多數人所熟知的名稱、作者所命名的名稱等。	光影・歷史・人物
	alternativeTitle* 其他書名	副題名、封面題名、劃一題名、全集題名、翻譯名稱、其他常用或少用的名稱...等， 著錄時需加入前導語 。	副題名：歌仔戲老照片
	isPartOf * 叢書名	叢書之題名與叢書號。	
subject* 主題標目	subjectMatter 分類	請根據索書號填入該資料的分類詞彙	戲劇
	subjectArea? 標目地區	地區複分	台灣
	keywords* 關鍵詞	請填入被描述資料之相關重要詞彙(如：人、事、時、地、物)，以利日後檢索之用。	
creator* 作者	creatorRole? 角色/頭銜	該書目資料之著者、編者、譯者等，包括個人、團體或公司。	編者
	creatorName* 姓名	該書目資料的作者姓名。	行政院文化建設委員會
description? 簡述	version? 版次	資料之版本項，如圖書之初版。	初版
	pages 頁數	總頁數	
	publishedPlace 出版地	出版地點	台灣
	abstract+ 摘要	資料之主題、內容大綱或意念描述，至少 300 字以上的說明，書有序採用序言。若有其他語言之摘要，請分別著錄。	
	notes* 附註	其他說明事項，例如：有法文譯本。如有多個附註時， 著錄時需加入前導語 。	
publisher* 出版者		該資料之出版者，包括個人、團體或公司。	行政院文化建設委員會
date?		以西元紀年，格式為 YYYY-MM-DD。	2002-11

欄位名稱			欄位說明	範例
出版日期			(若月日不詳者可僅著錄至西元年)	
identifier 識別資料	ISBN? 國際標準書號		International Standard Book Number 國際標準書號共 10 碼，不需著錄中間的 "- "	9570180609
	coverPage? 封面的物件檔名		存放封面的數位物件(jpg 800pixel)的檔名	cca100001-rf-007222519x-cp-i.jpg
	contentPage+ 內文的物件檔名		填入該圖書內容之典藏檔案 TIF(RGB)、JPG(RGB)、TIF(B&W)之起迄頁數，以”~”(半形)為分隔。(連結至數位檔)	cca100001-rf-007222519x-bw-0001-u.tif~cca100001-rf-007222519x-bw-0005-u.tif
	OCRPage* OCR 的物件檔名		填入該圖書內容之展示用檔案 TXT(OCR)之起迄頁數，以”~”(半形)為分隔。(連結至數位檔) ※ 只針對 TIF(B&W)、JPG(RGB)之典藏檔案進行 OCR 辨識。 ※ 展示用檔案 TXT(OCR)之檔名須與典藏檔檔名對應。	cca100001-rf-007222519x-0001-i.txt~cca100001-rf-007222519x-0005-i.txt
	systemID 系統識別碼		代表本書的識別碼(數位物件名稱去掉-0001-u.jpg 等附檔名)	cca100001-rf-007222519x
relation* 關連	chapterID 章節序號		填入本段落在整本出版品的序號	1
	chapter 章節名稱		著錄目次第一層名稱	歌仔戲的歷史
	pageNo 頁碼		資料之起迄頁數，以”-”(dash)為分隔	1-25
	fullText+ 全文影像		填入該篇章節內容之展示用檔案 PDF 的檔名。(連結至數位檔)	cca100001-rf-007222519x-0003-i.pdf
language+ 語文			該資料內容所使用的語文。 例如：中文、英文、法文等	中文
rights 權限 範圍	Copyright Restrictions 著作權/使用限制	holderStatement 著作權授權狀態	該出版品著作權授權使用狀態	永久授權
		accessRestrictions 使用限制	做為網路瀏覽用(internet)或會內下載用(intranet)之判斷依據	會內下載用(intranet)
	owner 典藏者	ownerCountry 典藏單位國家	典藏單位國家	中華民國

欄位名稱			欄位說明	範例
		ownerName 典藏單位	典藏單位或機構之名稱	行政院文化 建設委員會
originalConditions+ 原件說明			說明原件的品質狀況，例如：良好、蟲害、霉害、退色、印刷歪斜等等。	印刷歪斜
archiveLocation+ 庫房架位			紀錄典藏光碟於庫房中實體存放位置， 填入光碟編號	A0000001

4.3 影音出版品詮釋資料格式(Video Metadata Format)

欄位名稱			欄位說明	範例
videoID 影音出版品識別號			填入數位檔案命名,去掉 -0001-u.mpg 等的附檔名	cca100001-m v-vg305304
caseTitle? 案名			填入專案名稱	本土電影數 位化專案
programTitle? 節目名稱			填入節目名稱	人親土親鄉 土親
title 片名			填入片名	宜蘭觀光之 旅一
seriesTitle? 系列片名			填入此片是哪一系列	宜蘭觀光之 旅
callNumber 索書號			請填入該資料的索書號(分類號+作者 號)	990
subjectMatter 分類			請根據索書號填入該資料的分類詞彙	休閒
videoDescription 內容描述			將本影片的內容、大綱紀錄在此，且字 數以 100 字以上為原則，但不得與各片 段文字雷同。	
keywords* 關鍵詞			紀錄此片的關鍵詞，若有多個關鍵詞則 須重複填入本欄位	宜蘭
eventDate 事件時間			紀錄此片的拍攝日期	2004-12-16
location* 地點			紀錄此片的拍攝地點,如果有多個地點, 則重複輸入	宜蘭
descriptionLanguage* 描述語言			紀錄此片的描述語言，如：中文、英語、 台語、客語、原住民語、其他	中文
frameRate 秒格數			紀錄此片的秒格數	
frequency 頻率			紀錄此片的取樣率	
color 顏色			顏色,填入 COLOR 或 BW 或 MIXED	COLOR
originalFormat			Betacam-SP ; DigitalBetacam ;	16mm

欄位名稱		欄位說明	範例
原始影帶規格		Betacam-SX；DV；DVCAM：S-VHS：VHS；16mm；35mm；Super-8；Super-16	
originalConditions+ 原件說明		說明原件的品質狀況，例如：良好、蟲害、霉害、影帶跳動、影帶模糊等等。	第四段影帶跳動
archiveLocation+ 數位檔案存放位置		紀錄此片的光碟編號，如 A0000001，配合數位化燒錄光碟的編號名稱，若存放在多個光碟片，則紀錄多個光碟編號	A0000001
copyrightOwner 著作權所有者			行政院文化建設委員會
digitalCopyProvider 數位典藏單位			行政院文化建設委員會
publicAccess 權限使用範圍		做為網路瀏覽用(internet)或會內下載用(intranet)之判斷依據	會內下載用(intranet)
workPeople+ 參與人員	peopleName 人員姓名		黃春明
	organization? 組織	請填入該參與人員的所屬組織，如：蘭陽劇團、國家交響樂團、雲門舞集；若該參與人員無所屬組織，則可不填。	
	role 人員角色		受訪者
	roleDescription 角色內容說明		講述宜蘭地方口音
sequence+ 片段	sequenceID 片段序號	填入本片段在整卷影帶的序號	3
	sequenceDescription 片段內容描述	摘要說明本片段的内容，且字數以 100 字以上為原則，且須含括人、事、時、地、物。	
	digitalObject+ 數位物件連結	該片段數位化檔案之檔名(連結至數位檔)。需要填入(1)典藏檔案 MPG 的檔名。(2)展示用檔案 WMV 的檔名。(3)精華片段檔案 WMV 的檔名。每一個檔案需填入一個欄位。如： cca100001-mv-hjdd203902-0003-u.mpg cca100001-mv-hjdd203902-0003-m.wmv cca100001-mv-hjdd203902-0003-t.wmv	
notes* 附註		其他說明事項，如有多個附註時， 著錄時需加入前導語 。	

本詮釋資料格式參考台北藝術大學設計之「台灣社會人文電子影音數位博物館」計畫，所設計之影音 metadata 設置為本次專案之參考主要依據，此部分主要由國立台北藝術大學電影研究所所長李道明教授所發展設計出來，作為此次影音資料核心詮釋欄位，倘若各單位匯送之資料格式無法依循 ECHO metadata 規範時，則建議以 Dublin Core(DC) 15 個欄位作為詮釋資料之交換欄位，以兼具影音資料詮釋之適用性與交換資料之可行性。

----以上資料取自台北藝術大學之「台灣社會人文電子影音數位博物館」計畫報告並經修訂

MOC