

表格文本轉換工具(新版)

操作手冊

目錄

壹、	工具網址.....	3
貳、	表格文本格式.....	3
參、	操作程序說明.....	3
一、	上傳 CSV/EXCEL 檔案.....	4
二、	為轉換的文本命名與紀錄編號.....	5
三、	將表格文本的欄位對應至規範欄位.....	7
四、	安排 DocMETA 裡所要顯示的欄位.....	8
五、	添加文件內文.....	10
六、	DocuXML 建庫檔下載與直接建庫.....	11
肆、	策略運用範例.....	13

壹、 工具網址

<https://docusky.org.tw/DocuSky/devtools/CSVXLS2DocuXML/>

工具頁面如圖 1 所示：



圖 1 表格文本轉換工具(新版)首頁

貳、 表格文本格式

本工具目的在於將表格內的資料逐步轉製成含有 Metadata 的 DocuXml 檔，以利後續進行建庫與分析。使用者應備有一份由 EXCEL 編輯儲存的表格資料檔案。可使用的表格形式包括 Excel(.xls, .xlsx) 檔案或 .csv 格式的文本。

參、 操作程序說明

新版的表格文本轉換工具透過圖形化介面，便利使用者直覺操作表格文本的轉換流程。本工具操作共分成五個步驟，每一步驟的畫面皆顯示「進度條」、「注意事項」、「操作區域」、「使用說明」以及上一步、下一步的指示符號，使用者能藉由這些人性化的引導介面，將表格文本轉換為 DocuXML 格式的檔案，並隨後將其上傳至 DocuSky 平台建構成資料庫。

一、上傳 CSV/Excel 檔案

以下將使用一份以 Excel 整理《歷代名畫記》所存成的 xlsx 檔案，上傳至本工具中，示範表格文本的轉換流程。(圖 2) 在步驟一畫面中 (圖 3)，可以使用兩種方式將表格文本上傳至工具頁面，第一種方式是點選想要上傳的檔案，將其拖曳至步驟一「新增檔案」的框格中(①)；第二種方式為點選「新增檔案」，從浮現視窗中選擇欲上傳的檔案(②)。(上傳文本之前，請先確認文本內容為 UTF-8 編碼格式，以免轉檔後出現亂碼)

上傳成功後，原本的「新增檔案」框格會變為綠色(③)，將游標移入框格內，框格內將顯示「垃圾桶」、「眼睛」以及「綠色勾」三種符號，前二者代表：刪除這份表格文本、即時檢視表格文本內容。若按下「綠色勾」，則框格會轉為紅色，表示這份文件將不會進入轉換流程；這意味著，使用者可以同時上傳好幾份表格文本，並且可以選擇哪些文本將一起進入轉換流程。完成文本上傳後，即可按下畫面右方的「>」符號，前往步驟二。

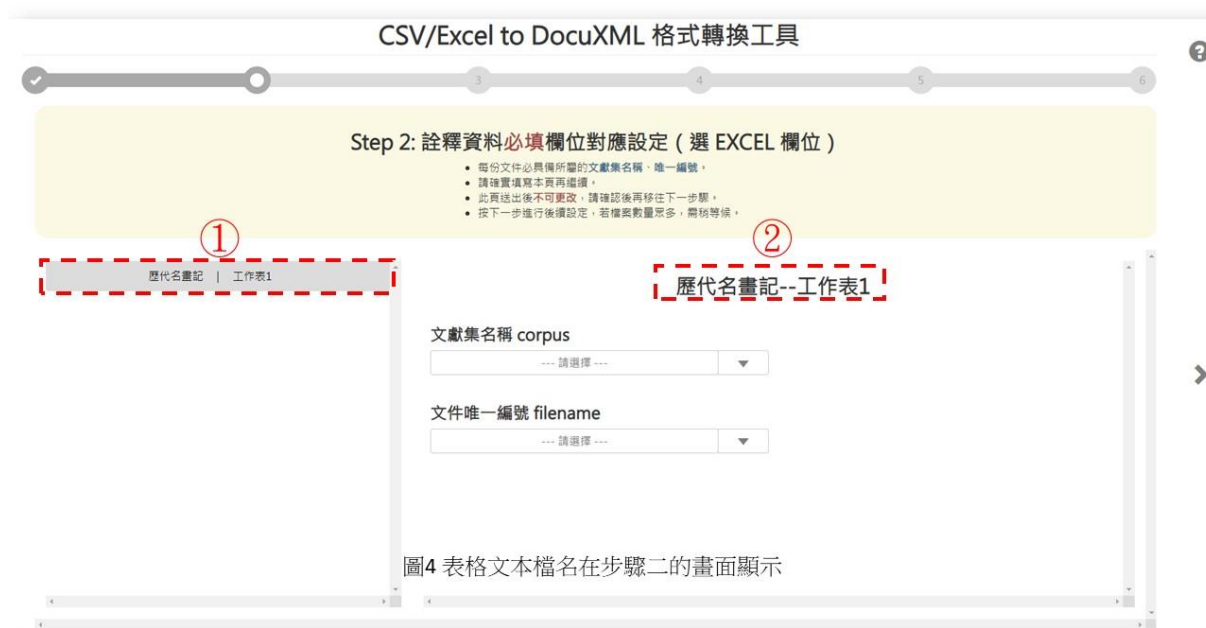
1	人物出場次序	朝代	畫家姓名	身分	畫作品評	擅長類型	畫作內容	本文
301	出場-288	020 唐朝下	畢宏	給事中京兆少尹;左庶子	善畫	山水、雜物	樹石	畢宏，大曆二年為給事中。畫松石於左省廳壁，好事者留許之。改京兆少尹，為左庶子。樹石擅名於代，樹木改步謝
302	出場-289	020 唐朝下	楊公南	中書門下侍郎;左僕射	善畫	山水		楊公南，名炎，華陰人。李肇三載，門樹六開。風骨俊秀，神情奕邁。善山水，高奇雅勝。大曆四年為中書門下侍郎
303	出場-290	020 唐朝下	史瓘	省郎	善畫	人物、動物	鞍馬、人物	史瓘，官至省郎。善畫鞍馬人物。
304	出場-291	020 唐朝下	裴諝	吏部侍郎兼御史大夫;太子賓客;左散騎常侍;兵部侍郎;河南尹;禮部尚書		山水		裴諝，字士明，河東人。以明經進。畫山水極有思。貞元中為吏部侍郎兼御史大夫。四年為太子賓客，左散騎常侍。
305	出場-292	020 唐朝下	韋鑾		工畫	動物	龍、馬	韋鑾，工龍馬，妙得精微。
306	出場-293	020 唐朝下	韋嗣		工畫	山水、雜物	山水、松石	韋嗣，工山水松石，發有其名，未免古拙。
307	出場-294	020 唐朝下	韋曉		工畫	山水、人物、傳教相關、動物	高僧、奇士、老松、異石、小馬、牛羊、山原	韋子驥，工山水，高僧、奇士、老松、異石，筆力勁健，風格高華。善小馬、牛羊、山原。俗人空知顧曹馬，不知韋
308	出場-295	020 唐朝下	張瑄	檢校禮部員外郎;鹽鐵判官;衡州司馬;忠州司馬	工畫	山水、雜物	樹石、山水	張瑄，字文通，吳郡人。初相隱劉晏知之，相繼王處榮檢校司馬員外郎。鹽鐵判官，坐事，貶衡州司馬，移忠州司馬
309	出場-296	020 唐朝下	陳雲	參議;兵部;衡州刺史;鳳賓縣令;使兼御史中丞;江南判官;著作佐郎;絳州司戶	善畫	山水		陳雲，字玄成，國初丞相陳之孫。明經出身。河南尹嚴武薦為參軍，昭義軍節度使李抱真辟為從事。貞元十四年官
310	出場-297	020 唐朝下	顧況		善畫	山水		顧況，字孟高，吳興人。不修檢操，頗好詩歌，善畫山水，初為韓弘公江州判官，又為著作佐郎，久次不遷，乃罷歸
311	出場-298	020 唐朝下	鄭審		善畫	山水、雜物	樹石、山水	鄭審，事貞元進士所撰《彩箋詩集》。
312	出場-299	020 唐朝下	沈寧		善畫	山水、雜物	樹石、山水	沈寧，亦善樹石山水，有雄偉，師張璪而少劣。
313	出場-300	020 唐朝下	劉商	檢校禮部郎中;汴州觀察判官	工畫	山水、雜物	山水、樹石	劉商，官至檢校禮部郎中，汴州觀察判官，少年有篇，跡高慎工，畫山水樹石，初師於張璪，後自造真為意。自張
314	出場-301	020 唐朝下	劉整		善畫	山水		劉整，任秘書省正字，善山水，有氣象，時有劉之奇，亦能山水。
315	出場-302	020 唐朝下	邊範	右衛長史	善畫	雜物、動物	花鳥、山、花、園、蔬	邊範，善畫花鳥，精妙之極。至於山花園蔬，亡不備寫，為右衛長史。花鳥冠於代，而有筆跡。
316	出場-303	020 唐朝下	子孫		善畫	動物、雜物	花鳥、蔬	子孫，善畫花鳥及蔬。
317	出場-304	020 唐朝下	孫頊		善畫	動物	花鳥	孫頊，善水鳥。
318	出場-305	020 唐朝下	梁廣		工畫	動物、雜物	花鳥	梁廣，工花鳥，善設彩，筆跡不遺毫髮。
319	出場-306	020 唐朝下	陳德		工畫	山水、人物、動物、雜物	花鳥	陳德，揚州人。師邊範花鳥，尤善布色。
320	出場-307	020 唐朝下	陳德		工畫	山水、人物、動物、雜物	人物、鞍馬、蟲禽	陳德，工山水，師張璪，有異調。人物、鞍馬、蟲禽并精於畫，善山水，妙過於父。《一云陳德
321	出場-308	020 唐朝下	戴雲		工畫	人物	子女	戴雲，工子女，極精微。
322	出場-309	020 唐朝下	周昉	宣州長史	善畫	佛、神、雜物	菩薩	周昉，字景玄，官至宣州長史。初效張萱畫，後則小異，頗極思慮。全法衣冠，不近閭里。衣裳勁麗，彩色柔麗，善
323	出場-310	020 唐朝下	趙傳文		善畫	人物、動物	子母、犬、兔、狐、鹿	趙傳文，尚書左丞之子也。畫子母犬兔，善寫貌，應物不窮。兄傳宣亦解畫。
324	出場-311	020 唐朝下	李嗣	太原王;劍南刺史	善畫	人物、傳教相關	子女、菩薩	太原王嗣，終劍南刺史。師趙傳文，善畫，但不及昉之精微。余大父高平公嘗未提筆之。

圖2《歷代名畫記》詮釋資料表



二、為轉換的文本命名與紀錄編號

步驟二的畫面（圖 4）左側部分以灰底反白標示上傳檔案的名稱(①)，以及該份表格文本在 Excel 裡歸屬的頁籤名稱；此處顯示為「歷代名畫記|工作表 1」（該標題在主畫面框格裡以更大的字體再度提示(②)）。本步驟需要使用者為這份準備轉換成 DocuXML 的文本亦即文獻集命名，而為使將來資料庫的每一筆記錄保持唯一性，表格文本裡的每一筆記錄必須被賦予唯一的編號，以避免出現資料混淆的結果。



點選「文獻集名稱 corpus」項目，其下拉式選單中可以選擇「檔案名稱 | 歷代名畫記」(③)作為文獻集的名稱，或選擇「自訂」功能，在空白欄位中輸入自訂文獻集名稱(④)。(圖 5) 為文獻集命名後，點選「文獻唯一編號 filename」項目，於選單中列出的諸多欄位裡，選擇表格內容沒有重複的欄位以作為唯一編號；此處，由於表格文本中設有一欄「人物出場次序」，為每一筆資料唯一且不重複的編號，可當作文件唯一編號(⑤)。若表格文本缺乏唯一性質的欄位內容，則可選擇「自動產生檔名」的方式，在空白欄位中填入「檔名前綴」(⑥)，由系統自動產生以前綴字彙為先導的編號。(圖 6) 完成第二步驟後，按下畫面右方的「>」符號，前往步驟三。

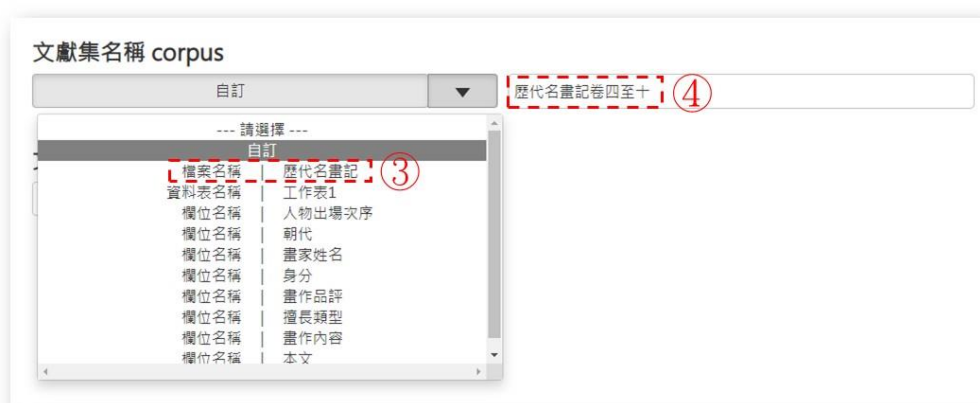


圖5 文獻集命名方式



圖6 文件唯一編號命名方式

三、將表格文本的欄位對應至規範欄位

使用者將文獻整理為許多筆資料，每筆資料由多個欄位組成，這些屬性各異的欄位共同描述了每一筆資料。這樣的表格文本轉換成資料庫時，使用者需要將這些欄位對應至系統已經擘劃且定義清晰的各種欄位，如此，系統得以理解使用者製作的欄位內容，進而轉換為資料庫的格式。進入步驟三的主畫面，使用者在表格文本中制訂的欄位名稱將依次羅列（圖 7），每個欄位名稱下方皆有顯示系統規範欄位名的下拉式選單，使用者需從中選取性質相應的欄位名。例如（圖 8），使用者制訂的「朝代」可對應規範欄位中的「文件時間（年號）| era」（①）；「畫家姓名」可對應「文件作者 | author」（②）。注意選單中各個規範欄位的底色，若該欄位名稱底色呈淺黃色，表示在資料庫建置後，該欄位的內容將於「後分類」顯示。此外，當使用者將游標滑過下拉式選單中的各類規範欄位時，畫面右側會同時出現該欄位的說明與填寫範例，以及是否會在「後分類」出現（③），讓使用者更清楚每個規範欄位的性質，使欄位對應更加契合系統的定義。有關係統制定的規範欄位項目及其定義，在本步驟頁面亦提供連結，讓使用者可以瀏覽完整且依據屬性排列的表單，便於參考每一欄位的定義與中英文名稱。（圖 9）

此外，使用者在此步驟要注意的是，欄位與規範欄位的對應是為了在資料庫建置後，讓這些欄位可以進入「後分類」；若使用者想要使某些欄位具有分類統計的「後分類」功能，則使用者必須在此步驟詳慮對應欄位的性質，亦即自己制定的欄位要與哪些具有後分類屬性的規範欄位相對應。

Step 3: 詮釋資料選填欄位對應設定（選系統預設欄位）

- 詮釋資料系統預設欄位詳細說明請見 [Metadata 規範表](#)。
- 每個系統預設欄位只能被選擇一次。

歷代名畫記 | 工作表1

歷代名畫記--工作表1

人物出場次序

請選擇 ---

朝代

請選擇 ---

畫家姓名

請選擇 ---

身分

請選擇 ---

畫作品評

請選擇 ---

文件主題層層三 | doc_topic_i3

- 後分類顯示：TP
- 說明：該欄文件的第三層主題。
- 在後分類與其他層層主題一同顯示：主題層層一/主題層層二/主題層層三

圖7 表格文本欄位對應至系統規範欄位

Figure 8 illustrates the column correspondence examples for the data selection interface. It shows three screenshots of the interface, each with a red dashed box highlighting the selected item and a red circle with a number (1, 2, or 3) indicating the step.

Screenshot 1 (Top): The dropdown menu is set to '請選擇 ---'. The selected item is 'era' (文件時間(年號)). The corresponding column is 'era' (文件時間(年號)).

Screenshot 2 (Middle): The dropdown menu is set to '請選擇 ---'. The selected item is 'author' (文件作者). The corresponding column is 'author' (文件作者).

Screenshot 3 (Bottom): The dropdown menu is set to '請選擇 ---'. The selected item is 'author' (文件作者). The corresponding column is 'author' (文件作者).

[illegible]

四、安排 DocMeta 裡所要顯示的欄位

為了讓使用者制定的欄位內容得以顯示於爾後建立的資料庫畫面 (DocMeta) 中，必須透過步驟四，將所欲顯示的欄位給予一個英文的欄位名稱，如此才能於建庫後，將每一筆資料的相關資訊顯示於畫面中。進入步驟四的畫面 (圖 10) 按下「+ 新增一項自訂欄位」(①)，在隨即顯現的視窗中點選「欄位資料」項目(②)，從清單中選取想要在資料庫中顯示的欄位，例如「朝代」，然後在「欄位名稱」欄裡填上任意的英文名稱，例如“DYNASTY” (③)。



圖10 安排DocMeta裡所要顯示的欄位

本步驟的設定除了能夠讓前一步驟對應到的規範欄位在 DocMeta 顯現之外，那些無法對應規範欄位的使用者自創欄位，也能在這裡透過一樣的步驟，使其顯示於 DocMeta 裡。例如（圖 11），表格文本中「身分」一欄，無法在第三步驟找到相應的規範欄位，但仍能透過本步驟的設定，賦予其英文欄位名稱，同樣也能在 DocMeta 顯示畫家的身分資訊。



圖11 使用者自創而無法對應的欄位也能顯示於DocMeta

要注意的是，由於這類使用者自創欄位無法與規範欄位相對應，因此其用途只能作為純粹顯示資料的相關訊息，不具備後分類功能。

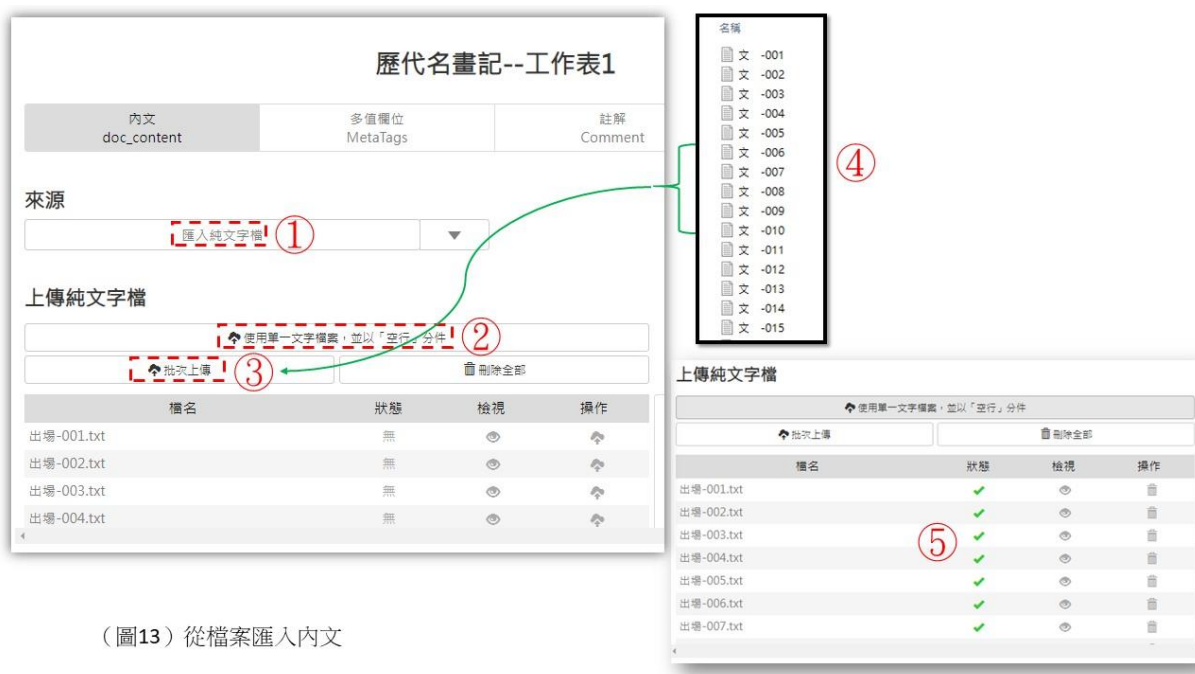
五、添加文件內文

若使用者整理的表格文本中含有或是需要納入較長篇幅的文字內容，則需透過此步驟的設定，使內文能齊備地呈現於資料庫畫面中，以利使用者進行瀏覽、檢索或進階的文本標記等應用。在步驟五的畫面中（圖 12），點選「內文」標籤(①)，畫面下方隨即出現內文「來源」選項；若內文已存在表格文件中的某個欄位裡，使用者可以在清單中選擇「CSV/Excel 欄位」(②)，隨後在「對應欄位」清單中(③)選取含有內文的欄位名稱，範例中的欄位名稱為「本文」(④)。



（圖12）從表格文本的欄位中添入內文

內文也可以存在一份或多份檔案中，以上傳的方式添入資料庫中。(圖 13) 在「來源」選項選擇「匯入純文字檔」(①)，若表格文件中，每一筆資料的內文共存於一個檔案裡，則於「上傳純文字檔」選項下，點選「使用單一文字檔案，並以空行分件」(②)，在隨後浮現的視窗中選取一份存有內文的純文字檔，上傳至工具中。此外，也可以用一件內文存成一個檔案的方式，批次將所有的檔案上傳。欲以此方式上傳檔案，可於「上傳純文字檔」選項下，點選「批次上傳」(③)，於浮動視窗中選取所有文字檔上傳(④)。無論使用單一檔案或是上傳，成功之後，「狀態」欄會由「無」轉成綠色勾(⑤)。使用上傳檔案的方式添加內文，可以隨時點選垃圾桶符號予以刪除個別紀錄裡的內文或是將內文全數刪除，重新來過。也能點選眼睛符號，檢視個別紀錄裡的內文。



(圖13) 從檔案匯入內文

其次，表格文本中的資料隨附註解欄位或是事件欄位，以作為資料的補充用途實屬常見，本步驟亦將上述兩種內文納入資料庫的呈現選項。(圖 14) 若欲表格文本中的註解內容納入資料庫中，可點選「註解 Comment」標籤(①)，於下方出現的「對應欄位」清單中(②)，選擇載有註解內容的欄位，如此，即完成欄位對應；「事件」欄位對應的操作方法與此相同。



(圖14) 增添註解或事件的內文

本步驟另一重要任務為「多值欄位 MetaTags」設定。(圖 15) 表格文本中，使用者制定的欄位大多數都能與規範欄位相對應，但還有另一類內容為多值的欄位，並不在規範欄位的對應範疇內；這類欄位的對應方式需在本步驟執行。在畫面中點選「多值欄位」標籤(①)，下方「對應欄位」清單中選

擇內容為多值的欄位；以歷代名畫記為例，一位畫家擅長的類型往往不只一種，因此，「擅長類型」(②)一欄記載著畫家擅長的各種繪畫類型，屬於多值欄位。一份表格文本中可有多個這樣的欄位並存(圖16)，例如，繪畫內容的要素：花、鳥、車、馬等亦屬多值，因此，「畫作內容」亦為多值欄位；只要點選「+新增一項對應欄位…」(①)隨即出現另一新的對應欄位選項，從「對應欄位」選單中選取「畫作內容」(②)，並於右側出現的空白欄位中填寫該欄位的英文名稱(③)。

這類稱為 MetaTags 的欄位，不僅內容為多值，更重要的是資料庫建成後，它們將被安排到「後分類」中，具備分類統計的功能。



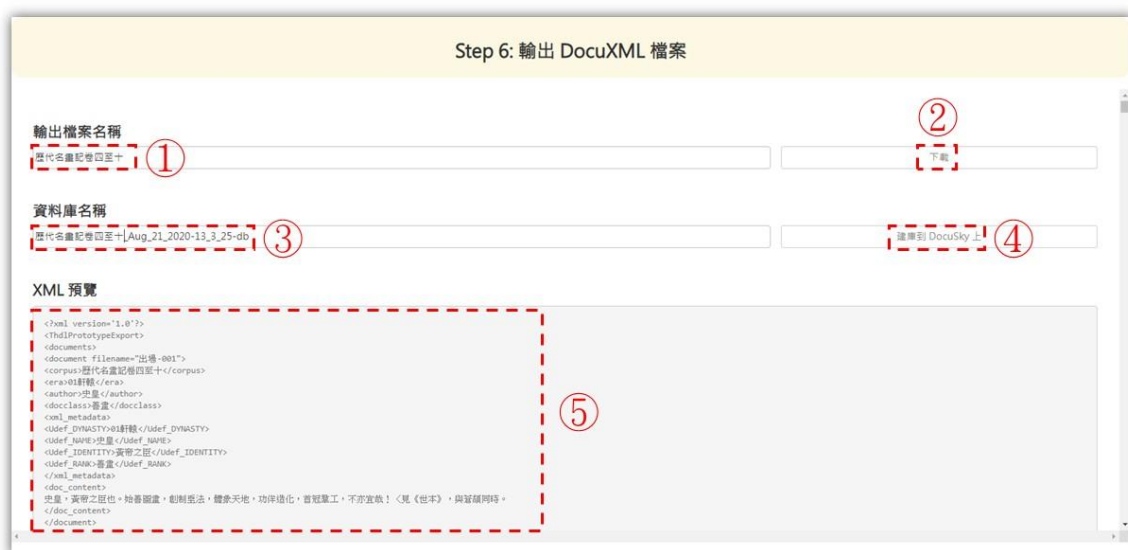
(圖16) 增加多值欄位對應

六、DocuXML 建庫檔下載與直接建庫

步驟六是表格轉換流程的尾聲，使用者將於此處獲得用來建置資料庫的 DocuXML 檔案。(圖 17) 首先在「輸出檔案名稱」欄位中為這份建庫檔命名，例如，“歷代名畫記卷四至十” (①)，然後點選欄位右側的「下載」按鍵 (②)，即可將建庫檔下載到本地電腦上，可供使用者稍後上傳至 DocuSky 平台進行資料庫建置，或使用其他工具對該檔案進行處理。

若使用者不打算即刻對建庫檔進行另外處理，亦可略過下載程序，直接將轉換成果上傳 DocuSky 平台，立即建置資料庫。只要在「資料庫名稱」欄位為資料庫命名 (③)，然後點選「建庫到 DocuSky 上」 (④)，即完成建庫手續。

此外，若使用者嫻熟 XML 語法格式，在下載檔案或直接建庫之前，使用者可以透過畫面上的「XML 預覽」框 (⑤)，瀏覽表格文本轉換後的內容，檢查是否有語法錯誤等問題。



(圖 17) DocuXML 建庫檔下載與直接建庫

肆、策略運用導引

一、表格文本中的數字編號宜於數字前一位添加“0”，例如數字範圍在二位數之內以 01 為始，三位數以 001 為始，以此類推。此舉能使資料庫依據編號呈現資料時，能夠正確羅列資料。

二、使用者將資料整理為表格文本時，用以詮釋資料的各種欄位之制訂，建議參考 Metadata 規範表，這不僅有助於使用者更具結構化地釐清資料的類別與屬性；精確編排的表格文本在轉換之後，有助於使用者獲得資訊更翔實、豐富的資料庫。

三、添加文件內文的步驟，建議使用者盡可能將內文存放於表格文件的欄位中，這是較為便利且避免錯誤的方式。若內文以單一文檔儲存時，需注意每一件對應到一筆記錄的內文，必須以空行相隔，以免系統在對應時造成錯誤。

四、步驟六的下載 DocuXML 建庫檔或直接上傳建庫，並非二則一的選項，使用者可以同時下載建庫檔，並且選擇同時建庫到 DocuSky 上。如此，可以儘速見到資料庫呈現的面貌與功能，若發現資料庫內容有誤，則可透過平台工具編修下載的建庫檔，再重新上傳建庫。