

從資源建置及交流觀點探討中文人名權威紀錄消歧 議題

Issues Related to Chinese Name Authority Disambiguation in View of Resource Establishment and Exchange

許靜芬 Hsu Ching-fen

國家圖書館書目中心主任

Editor, Bibliographic Information Center, National Central Library

Email: tacoma@ncl.edu.tw

【摘要 Abstract】

區分同姓名不同人物，不僅是圖書館人名權威控制重要的任務，也是學術人員正確評估學術產出重要的助力之一。本研究目的為歸納名稱權威建置主要之消歧欄位，從資料建檔及資源交流合作角度，探討消歧資料著錄及應用之跨語言共通議題，及中文環境之特殊議題。透過文獻探討圖書館及 VIAF、ORCID、ISNI 等名稱資源採用的消除人名歧異方式，進而針對名稱、年代等權威紀錄重要欄位，抽樣比較「臺灣書目整合查詢系統」(SMRT)之中文人名權威與 VIAF 相對應紀錄之著錄異同。研究結果提出之建議包括改善消歧欄位之著錄，建請臺灣學術研究人員及早善用國際標準人名識別符號如 ORCID 或 ISNI，採用分階段名稱消歧作業程序，將有助提升名稱消歧精確度及資源交流或鏈結資料之效益。

To disambiguate identical personal names is not only an important task of authority control at library but also a fundamental key to precision of assessing academic productivity. The purpose of the study is to sort out the major elements for name disambiguation and explore the issues common in different languages and issues pertaining only to Chinese names. This study reviews literature to examine how library, VIAF, ORCID and ISNI differentiate personal name records, and furthermore

compares name and date information in sample authority records both from the Synergy of Metadata Resources in Taiwan and from VIAF. The research concludes with suggestions on enhancing information related to personal names, encouraging use of international standard identifier and applying multi-phase processes to improve precision of name disambiguation and linked data.

【關鍵詞 Keywords】

人名權威、名稱消歧、權威控制

Personal Name Authority; Name Disambiguation; Authority Control

一、前言

長久以來權威控制在圖書館目錄中扮演提升書目控制及查詢效益的幕後工作，自動化系統透過各類型權威紀錄，維持書目紀錄中各類型標目的標準化及標目之間的參照關係，以便提供符合需求的查詢結果。目前個人名稱標目是書目紀錄中為數最多的標目，因此，個人名稱權威紀錄(簡稱人名權威)成為圖書館在人力經費有限情況下，通常會優先建置的權威紀錄類型。在中文姓名重複性極高的情況下，區別名稱相同的權威記錄是否屬同一人物，成為建置多重來源權威資料庫的重要程序。這種消除歧異的過程通稱為「disambiguation」，本文將以「消歧」一詞敘述相關過程。

國家圖書館(簡稱國圖)和國立臺灣大學圖書館(簡稱臺大圖書館)自1998年開始進行權威資料整理合作，將兩館各自建立的人名稱權威紀錄彙整為單一資料庫，目前以MARC21格式儲存於「臺灣書目整合查詢系統」(Synergy of Metadata Resources in Taiwan, 簡稱SMRT)提供查詢下載。這項合作除了增補紀錄內容外，將兩館對於同一人物的重複紀錄進行辨識、確認、整併等消歧程序，也成為資料彙整重要的一環。

臺灣地區中文名稱權威資源合作之外，國圖於2004年受邀加入兩岸三地「中文文獻資源共建共享合作」的子項「中文名稱規範聯合協調委員會」，與中國大陸中國國家圖書館、北京大學CALIS聯機編目中心、香港JULAC-HKCAN等三地成員館，共同建立「中文名稱規範聯合數據檢索系統」(簡稱CNASS)，提供四成員館上傳其

各自建置的人名權威紀錄，目前以分散式架構分別儲存及管理成員館上傳的資料，前台則提供使用者以單一介面同時查詢各館上傳的資料。由於現行政策不整併重複紀錄，因此，查詢結果也會顯示各館彼此重複的紀錄。

在兩岸三地進行華文權威資源共建共享合作的同時，美國國會圖書館、德國國家圖書館及 OCLC 於 2003 年 IFLA 研討會期間，簽訂「虛擬國際權威資料檔」(Virtual International Authority File, 簡稱 VIAF) 聯盟合作協議，法國國家圖書館在 2007 年加入後，成為 VIAF 四大主要推動單位 (OCLC, 2012)。VIAF 打破各地區記錄建檔格式、語文的差異，建立同一人物之不同來源紀錄之間的關聯，不但鏈結權威紀錄也連結相關的 WorldCat 書目紀錄。VIAF 資料收錄目前已擴及 30 國之 35 所國家圖書館的權威紀錄檔，其中個人名稱權威紀錄超過 3 億 5 千萬筆 (OCLC, 2014)。即便如此，其中除了仍在測試階段的新加坡國家圖書館資料之外，尚缺華文地區提供的中文人名權威資源。國圖在所述與國內及兩岸三地的權威建置合作基礎上，也期望加入 VIAF 國際社群合作，將臺灣的中文人名權威資源推廣至國際圖資界及更廣大的語意網社群。

有鑑於上述權威建置及推動資源交流的過程中，人名權威消歧顯然是不可或缺的過程。因此，希望經由本研究達到下列目的：

- (一) 從建檔及應用的角度，針對現有編目相關規範 RDA、MARC 21 Format for Authority Data (簡稱 MARC21 權威格式)，及較具規模的 VIAF、ORCID 等人物資料檔，歸納出與人名權威紀錄消歧相關的重要欄位。
- (二) 依據建檔實務作業的難易情形，針對能有效區別人物異同的中文人名權威資料欄位，提出資料建置及維護建議。
- (三) 從跨資料庫鏈結及資源交流合作的角度，針對前述有助消歧的欄位資料，提出應用方式相關建議。
- (四) 在參與跨資料庫鏈結合作的資料庫仍維持各自更新的情況下，提出個別及鏈結資料維護更新的可行方式建議。
- (五) 透過實際樣例資料比較，評估有助消歧欄位資料蒐集建檔之難易度，考量各欄位用於消歧之時機及優先序，模擬建構消歧作業程序，作為後續撰寫消歧程式或人工作業之參考。

二、文獻探討

辨識作者名稱不但在提升圖書館館藏管理及目錄服務品質扮演重要角色，對版權管理、出版社等機構而言，更有商業利益考量的必要性，以便藉此確立作者與作品的關聯性，避免因名稱混淆造成的糾紛。Smalheiser 與 Torvik (2009) 指出作者名稱消歧的四大挑戰為：(1) 單一作者可用不同名稱發表或出版作品；(2) 有相同名稱的作者有多人；(3) 出版社或書目資料庫裡欠缺重要的作者資料，如完整的作者名字、所在地、學歷或頭銜等；(4) 許多學術著作是由多位跨領域作者共同完成。為了釐清人名與作品實體的關係，研究者從不同角度探討名稱消歧方式。

Taffet (2004) 將名稱與人物實體的關係分為一名對多人、多名對一人兩種關係，以 *Biographies of Notable Americans* 的一萬四千餘則人物描述著手，試著分析、註記文章中與人物相關的各種形式名稱、事件，並建立其關聯，歸納人物紀錄主要元素，再透過使用者研究修正消歧程式，以便試驗發展跨文件的人名消歧程式。

張菲菲等人 (2010) 提出以擷取訊息為基礎的中文人名消歧系統，透過實體及實體關係識別的跨文件自然語言分析，加上日期及地點的格式化，聚合實體相關識別訊息。對相同人名進行消歧，最重要的就是需要選出足以區別不同人名的主要個人特徵，包括姓名、別名、出生日期、出生地點、居住地、Email、職位、家庭成員等，另選出實體關係特徵，包括與人物相關的其他實體、出現在同一文件的人名、地名、團體名稱等。研究結果顯示個人資訊雖能表示一個人的身份，但由於語彙中的個人資訊特徵較少，所以造成查全率較低；加入實體關係特徵時，查全率提高了 44 個百分點，說明使用實體關係特徵能改善個人的辨識度。

除了從個人相關資訊研究名稱消歧之外，也有從分析文獻書目資料方式探討名稱消歧議題的相關研究，使用的書目資訊包括共同作者名稱、文獻題名、期刊或會議論文集題名 (Han et al., 2004)，而陳光華與謝其男 (Chen & Hsieh, 2011) 除了 Han 等人用的三種資料之外，另增加了未曾被使用過的出版年和頁數，相較於作者名稱和題名，出版年和頁數在格式上的出入較少，雖然有助於提升消歧的正確率，但因書目內容的變數較多，對名稱輔助仍有其限度。

目前最早建置且現今最具規模的國際合作權威資源為 VIAF，其資料來自各國不同格式的各類型權威紀錄檔，其中名稱權威包括個人、團體、地名三種類型，而個人名稱權威紀錄累計超過 3,800 萬筆 (Hickey & Toves, 2014)。VIAF 在處理 30 個國家

的 36 個機構提供的權威紀錄檔時，先將每筆權威紀錄與 WorldCat 超過 1 億 400 萬筆書目紀錄建立實體 (entity) 的關聯，以便確立權威紀錄所描述的實體與其相關作品的關係，再依據權威與書目的關係，擷取用於比對的書目紀錄資訊，包括各種形式名稱、與名稱相關的日期、與出版品相關日期、題名、共同作者、ISBN 及其他標準識別符號，再將重要書目資訊整併於原始權威記錄中，另產生「加工權威紀錄」(processed authority record)。加工權威紀錄通過後續層層比對、匹配、消歧等複雜程序後，將不同來源紀錄對於同一人物的權威紀錄區分為相關聯的群組 (group)，程式再依據關聯度聚集相關群組，形成包含多種來源紀錄的群集 (cluster)，並給予各群集一組固定的 VIAF 識別符號 (VIAF ID)。VIAF 每月重新執行所有資料之比對、匹配、消歧程序，隨著權威紀錄的匯入或內容更新，加上程式不定期改進調整，不同群集中的紀錄也可能隨之重新排列組合，在不斷精進資料處理過程中，讓群集透過日常微調，提高紀錄群集的準確度 (Hickey, 2014)。

2007 至 2009 年間進行的 Name Project 目標為研究英國典藏社群所需之名稱權威服務，並發展一雛形機制。計畫包括分析現有名稱權威機制，並找出相關資料來源。研究團隊訂定服務所需的詮釋資料，並決定現有規範標準對於雛形機制是否適用。該雛形包含 (1) 從期刊文獻或研討會論文資料庫預先擷取相關書目訊息，(2) 允許個人管理並優化不轉入的資料。該計畫擷取用以作為消歧參數的書目資訊，包括作者名稱、文獻題名、主題分類及關鍵字 (Hill, Needham & Danskin, 2009)。該計畫成果已納入 ISNI 資料庫。

研究人員的學術成果對個人及其所屬機構皆相當重要，在學術領域資源中若能準確查詢名稱，對研究人員、所屬機構和經費提供者進行學術產出評估都相當重要，此外，專任職務、經費、協作等也都需仰賴這些評估。透過精確辨識各研究者的產出，個人引用度量比起具爭議的期刊影響係數更為有效及可靠 (Cals & Kotz, 2008)。

ORCID 設立宗旨是創建個別研究者的識別符號，但 ORCID 的資料來源同時包含個人建置及機構批次提供的紀錄檔，因此不可避免會收錄同一人物的重複紀錄。Bilder (2011) 認為要探討偵測重複紀錄的電腦運算規則時，必須了解能讓運算規則最佳化的特定詮釋資料，因此，提出對於不同來源的相同人物紀錄加上「與……相同」(same_as) 的宣告 (claim)，再依據此宣告建立紀錄之間的關聯，透過分析這些宣告的唯一、重複及衝突性，就可建立基本但有效的信任度量。善加利用無法避免的紀

錄重複和衝突特性，反倒能建置並逐步釐清紀錄出處、建立信任機制、清楚建立紀錄所屬的關聯。

ISNI 於 2012 年 3 月正式發布國際標準 ISO 27729，給予各類媒體內容貢獻者的標準識別符號。此識別符號可視為描述人物詮釋紀錄的速寫，以唯一且簡單的全球通用識別符號區分人物。ISNI 蒐集 VIAF、全球 300 個版權管理團體及 2 萬 6,000 所圖書館提供的名稱紀錄，經過資料處理程序後，集中給予個別人物唯一識別符號，也希望這組 ISNI 識別符號能被全球更多機構採用於其人物紀錄中，以便透過網路串接同一人物跨時空及語種的相關資料。

ISNI 的核心紀錄欄位，包括公眾認知的名稱、其他形式名稱、創作類型、創作職責、資料來源 URI，其他輔助辨識的資料包括：共同作者、所屬機構、出版社、作品題名、生卒日期等。ISNI 的各種來源資料以 VIAF 紀錄為比對基準，並依據比對符合程度註記相對的信任級別。在 ISNI 查獲及顯示的內容通常來自 VIAF，以確保其他資料提供者的資料保密特性 (Gatenby & MacEwan, 2011)，ISNI 比對符合 VIAF 的紀錄或發現 VIAF 紀錄錯誤時，皆會回饋給 VIAF。

為了辨識及區分各領域創作者，圖書館界、版權管理機構及出版界各自建置不同規範的人名紀錄檔，這些檔案的建立及維運過程有其困難度，而且不同機構也花費重複人力經費維護同一人物的資料 (Gatenby & MacEwan, 2011)。圖書館以人工辨識、區分不同作者的策略，在數量日增且相互連結的數位圖書館時代，已趕不上圖書館迅速且大量增加的數位資源。由特定機構給予作者單一識別符號的做法雖然簡單，技術也可行，但對人為因素欠缺考量，包括作者即時更新資料的意願和持續性，在全球性資料庫公開個人資料的接受度，而此類資料庫系統的維運人力經費也需考量 (Smalheiser & Torvik, 2009)。

三、研究方法與執行步驟

本研究主要採用文獻分析法，透過研讀近幾年人名權威建置及處理相關文獻，著重 MARC21 權威格式配合 RDA 增訂欄位之討論，並蒐集探討同名不同人之權威紀錄消歧作法，再經由現行名稱權威資料庫維運相關文獻之研析，探討實務運作可行方向。另透過蒐集 VIAF、Open Researcher and Contributor ID (簡稱 ORCID)、國際

標準名稱識別符號 International Standard Name Identifier (簡稱 ISNI) 資料庫對人名權威紀錄著錄情形、消歧方式相關資料，再從 SMRT 選取具備若干有助於消歧欄位的樣本紀錄，分別與 VIAF 相對應的人名權威紀錄進行著錄差異之比較。

樣本紀錄的蒐集先採用國圖內部使用之批次書目處理工具進行權威記錄資料篩選，而後再採用 Excel 應用軟體之篩選、小計等功能進行樣本分析，最後則以人工檢視確認，執行方式分別說明如下：

(一) 選定消歧重點欄位

能用以辨識人物異同的資料項目相當多樣，本文將針對 RDA、MARC21 權威格式對人名權威記錄元素的相關說明，再檢視 VIAF、ORCID、ISNI 及 SMRT 資料實際著錄情形，選擇各資料庫較共通的元素作為樣本資料篩選及著錄比較的重點欄位。

(二) 資料篩選

先擷取已經由人工修訂的較完整 SMRT 人名權威紀錄，再依據前項選定的消歧重點欄位，分別擷取出重點欄位完備程度不一的資料集，篩選方式為僅含 1 個重點欄位、含多個不同組合之重點欄位。

(三) 第 1 次樣本查找

前項篩選出的各資料集中，選擇與 VIAF 資料欄位較可能有交集的資料集之前 6,000 筆為第 1 次樣本，擷取這些記錄的紀錄號碼、主要中文名稱（主標目）、生卒年、專長領域，製作成 Excel 工作表。以該表的主要中文名稱逐一在 VIAF 查找相對應紀錄，而且僅透過單純檢視名稱及年代，判別是否查獲同一人物紀錄。確認查獲者，在 Excel 清單記錄 VIAF 群集的 VIAF ID；無法立即判別異同或無法查獲者，亦分別註記。

(四) 第 2 次樣本對照檢視

取第 1 次樣本的前 500 筆作為第 2 次樣本，逐一詳細檢視 VIAF 相對應記錄的內容，整理並記錄確定相同、不確定異同及無法查到的原因。並比較 SMRT 與 VIAF 相同人物紀錄的主要名稱、生卒年欄位著錄的差異。

四、研究資料分析

本文研析國際通用編目規範對於消歧相關欄位之說明，及現有國際權威資源共通的消歧欄位，再據以擇定篩選 SMRT 及 VIAF 樣本的消歧重點欄位，並進一步比較兩種資料庫樣本對於這些欄位的著錄情形，結果如下。

(一) 相關編目規範說明

FRAD 的使用者工作指出，權威資料的工作在於幫助使用者尋找 (find)、識別 (identify)、釐清實體之間的關係 (contextualize)、了解選定權威名稱及形式的理由 (justify)。FRAD 實體的組成包括名稱 (names)、識別符號 (identifier)、受控檢索點 (controlled access point) (陳和琴，民 100)。國際圖聯規範記錄的功能編號工作組 (2009) 研擬 FRAD 出於模型的需要，在「邏輯」層次上定義屬性 (即定義實體的特徵，而不是具體定義資料單元)。某些情況下，給定實體的特徵會不時地發生變化 (例如，某人從事的活動領域)。這個模型未對隨時改變的屬性與不改變的屬性之間做出嚴格的區分。針對個人所定義的屬性，包括與個人相關的日期、頭銜、性別、出生地、死亡地、國別、居住地、機構、地址、個人的語言、活動領域、專業/職業、傳記/歷史、與個人相關的其他訊息。這些屬性中除了特定日期 (如出生、死亡)、出生地、死亡地之外，其餘皆可能發生變化。

RDA 依據 FRAD 概念，在第 9 章 Identifying Persons 為辨識個人定義的元素包括首選 (preferred) 及其他形式名稱、出生/死亡/特定活動等日期區間、頭銜、完整形式名稱、其他相關標示、性別、出生/死亡/居住等地名、國別、地址、機構、語言、活動領域、專長或職業、自傳資料、識別符號。其中首選名稱、出生日期、死亡日期、頭銜、完整形式名稱、其他相關標示、專長或職業、識別符號被列為描述個人實體的必備核心元素。(American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals, 2014a)

針對 RDA 定義的個人屬性元素，MARC21 在 01X-09X 段之號碼及代碼欄位，以及 1XX 和 3XX 段標目相關訊息欄位，各有所新增或重新定義相對應的欄位 (如表 1)，對於原本可記載多種訊息的分欄，在保留現有分欄情況下，另外為各種訊息分別定義新的專屬分欄，例如最常用以辨識人物異同的欄位 100/400/500/700 分欄 d，維持現有分欄 d 的生、卒日期，另於欄位 046 分別定義專屬分欄 f (出生日期) 及分欄

g (死亡日期)；前述 RDA 為個人定義的元素，除了個人相關名稱、日期及識別符號之外，幾乎都在 3XX 段有其專屬欄位，包括頭銜 (368 分欄 d)、其他標示 (368 分欄 c)、活動領域 (372 分欄 a)、專長職業 (374 分欄 a) 等核心元素，及性別、出生地、死亡地，以及相關國別、地名、機構、語言、活動領域等非核心元素。(American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals, 2014b)

表 1

RDA 第 9 章 Identifying Persons 規則與 MARC21 相關欄位對應

RDA 規則編號	RDA 元素名稱	MARC 21 欄位	MARC 21 分欄代碼	MARC 21 欄位或分欄名稱
9.2.2	Preferred Name for the Person	100, 400, 500 1st indicator 0 or 1	a	Personal name
9.2.2	Preferred Name for the Person	100, 400, 500 1st indicator 0 or 1	b	Numeration
9.2.3	Variant Name for the Person	400 1st indicator 0 or 1	a	Personal name
9.2.3	Variant Name for the Person	400 1st indicator 0 or 1	b	Numeration
9.3	Date Associated with the Person	100, 400, 500 1st indicator 0 or 1	d	Dates associated with a name
9.3.2	Date of birth	046	f	Birth date
9.3.3	Date of Death	046	g	Death date
9.3.4	Period of Activity for the Person	046	s	Start period
9.3.4	Period of Activity for the Person	046	t	End period
9.4	Title of the Person	100, 400, 500 1st indicator 0 or 1	c	Titles and other words associated with a name
9.4	Title of the Person	368	d	Title of person
9.5	Fuller Form of Name	100, 400, 500 1st indicator 0 or 1	q	Fuller form of name
9.6	Other Designation Associated with the Person	100, 400, 500 1st indicator 0 or 1	c	Titles and other words associated with a name
9.6	Other Designation Associated with the Person	368	c	Other designation
9.7	Gender	375	a	Gender
9.8	Place of Birth	370	a	Place of birth
9.9	Place of Death	370	b	Place of death
9.1	Country Associated with the Person	370	c	Associated country
9.11	Place of Residence, Etc.	370	e	Place of residence/headquarters

表 1 (續)

RDA 規則 編號	RDA 元素名稱	MARC 21 欄位	MARC 21 分欄 代碼	MARC 21 欄位或分欄名稱
9.12	Address of the Person	371	a, b, c, d, e, m, z	Address
9.13	Affiliation	373	a	Affiliation
9.14	Language of the Person	377	a	Language code
9.14	Language of the Person	377	l	Language term
9.15	Field of Activity of the Person	100, 400, 500 1st indicator 0 or 1	c	Titles and other words associated with a name
9.15	Field of Activity of the Person	372	a	Field of activity
9.16	Profession or Occupation	100, 400, 500 1st indicator 0 or 1	c	Titles and other words associated with a name
9.16	Profession or Occupation	374	a	Occupation
9.17	Biographical Information	655	a	History reference
9.17	Biographical Information	678 1st indicator 0	a, z	Biographical or historical data
9.18	Identifier for the Person	010	a, z	LC control number
9.18	Identifier for the Person	016	a, z	Record control number
9.18	Identifier for the Person	024	a, z	Standard number or code
9.18	Identifier for the Person	035	a, z	System control number

Note. Adapted from “RDA Elements Mapped to MARC 21 Format for Authority Data,” by American Library Association, Canadian Library Association & CILIP: Chartered Institute of Library and Information Professionals, 2014. Retrieved from <http://access.rdatoolkit.org/jscmap3.html>. Copyright RDA Toolkit.

(二) 權威資源的消歧作法

1. VIAF

VIAF 資料處理過程中，除了名稱之外，日期在很多項程序中都是重要的比對資訊。在匹配同一實體的權威紀錄時，主要先以姓氏聚集待比對的紀錄，再依據不同欄位初步給予不同級別的關聯度，若名稱/題名相似且人物的生卒年符合的話，將被視為關聯度最高的紀錄。某些比對結果若無歧異，則會採單一日期比對，例如若出生年一致就視為比對符合 (Hickey, 2014, p.7)。另外，將匹配成對的紀錄進行群組的過程中，如果有 2 筆相同來源的紀錄連結到另一來源的某筆紀錄，其中一筆是因為個人名稱及作品題名相符，另一筆則因名稱及 2 種日期相符，就這兩種情況而言，會優先認定名稱及 2 種日期相符者確實相同，而忽略名稱及作品題名相符者 (Hickey, 2014, p.8)。

完成比對的權威記錄，則將相關書目資訊納入，結合成加工權威記錄，這些紀錄以 MARC21 為基礎，除了 MARC21 權威資料定義的欄位之外，VIAF 利用自行定義的 9XX 段欄位（如表 2），儲存擷取自 UNIMARC 及 MARC21 權威及書目紀錄之人物屬性或作品相關資料，包括：連結不同來源的相同人物權威紀錄、權威所屬人物相關之作品、辨識號碼等用於消歧的資料，例如：欄位 945 儲存職業代碼、欄位 946 儲存性別、欄位 947 儲存取自 UNIMARC 欄位 102 或 MARC21 權威紀錄欄位 370 的國別、欄位 948 儲存取自 UNIMARC 欄位 101 或 MARC21 權威紀錄欄位 377 的語言（OCLC, n.d.）。

表 2
VIAF 加工權威紀錄自訂欄位說明

欄位	資料	資料來源
901	ISBN	(020\$a)
902	ISSN	(022\$a)
903	LCCN	(010\$a)
910	Title from bib	(or \$t in an auth record)
912	Title not used during matching (e.g. a very common title)	
919	Title from authority record note field	The 'a' subfield is the normalized form used in matching, the 'A' shows the raw text
920	Publisher id	(020\$a)
921	Publisher	(260\$b, 533\$c)
922	Country of publication	
930	Alternate form of author's name	From the statement of responsibility
940	Language	(041 \$a or 008)
941	Relator code	
942	NATC	(050 \$a)
943	Decade of publication	From dates 1 & 2
944	RecType + bibLvl	
945	Occupation code	
946	Gender	
947	Nationality	From UNIMARC 102 or MARC-21 370
948	Language of entity	From UNIMARC 101 or MARC-21 377
950	Joint personal author	
951	Joint corporate author	(110 710)
955	Personal name added entry	100 & 700's from records with this name as a subject
960	Personal name subjects	From 600's, if they don't match main author
969	'Subject'	Present if person occurred as a subject (600) in a bibliographic record

表 2 (續)

欄位	資料	資料來源
996	Authority IDs compared, but not matched	
997	Coded version of dates	Birth date and death date, along with indication of type of date. A date of 0 indicates unknown. Date types can be 'lived', 'circa', or 'flourished'
998	Authority link	\$a name \$2 Linked-to record ID \$3 Type of link
999	Bibliographic record link	

Note. Adapt from “VIAF Tag Documentation,” by OCLC. Retrieved from <http://viaf.org/viaf/viaftags.xml>. Copyright OCLC.

2. ORCID

ORCID 紀錄之建立大致分為本人申請或由第三方代理申請，申請的基本資料至少須有名稱及 email，若由第三方代理申請，必須提供相關授權資訊，才會給予 ORCID 辨識號。ORCID 蒐集的個人資訊除了名稱和 email 之外，另有個人學經歷、研究計畫、專利及出版品等。(ORCID, 2013) 依據 ORCID 網站在建立個人紀錄的指引說明，個人資訊分為名稱 (Names)、關於我 (About me)、網站 (Websites) 三部分。名稱部分的欄位，包括姓氏、名字、在 ORCID 紀錄公開名稱 (published name)、其他名稱 (如名字縮寫、婚前姓氏、其他文字名稱等)。在「關於我」部分則涵蓋 1,000 字元以內的自傳、256 字元以內的研究領域關鍵字、從選單選取工作所在的國家。網站部分則可輸入個人、服務單位、Wikipedia 或社群網站的網址。(ORCID, n.d.)

ORCID 以現代研究人員為主，實際抽樣查詢 ORCID 的紀錄，可發現中文紀錄公開的內容完備程度落差相當大，許多華人名稱紀錄的查詢結果顯示並無中文名稱，亦無其他足資辨識的資訊，例如查詢「chen hua」的結果有 12 筆名稱相符，其中 7 筆無其他資料，其餘 5 筆僅有 1 筆提供學歷，另 4 筆則有作品書目。

3. ISNI

ISNI 是目前廣為圖書館、出版界、版權管理組織採用的識別符號 ISO 標準，並已成為語意網資料鏈結環境中重要的元素。該資料庫收錄全球數百個跨領域權威資料庫的名稱紀錄，其中 VIAF 也是資料提供者之一，而 ISNI 也正進行資

料匯入 VIAF 之測試。由於該資料庫資料來自各會員，因此，會員也必須承擔資料源頭品管的責任。在法國國家圖書館及大英圖書館的品管團隊，則負責評估新的來源資料、檢視統計抽樣、修正需要人工檢視的紀錄等作業。透過程式及人工品管，以便有效判斷批次匯入的紀錄是否與現有紀錄重複，避免對於同一實體重複給號。此外，也開放請使用者提供資料之釐清或增修，使用者提供的修定會經過品管團隊確認後才決定是否顯示於前臺。(ISNI International Agency, n.d.)

由於會員來自文化、研究、出版、創作、演藝界等各類型機構，因此 ISNI 的紀錄元素包括名稱、生卒年、性別、城市、國家、國籍、創作類型、創作職責、其他相關人物名稱、作品題名、資料可信度、資料來源代碼、樂器及聲部、來源識別符號的類型、文字等。另有針對名稱定義的欄位 Name use person，註記該名稱為公開、內部或虛構性質 (ISNI International Agency, 2013)。

4. SMRT

自 2012 年開始，SMRT 人名權威紀錄從原本的 CMARC 轉換為 MARC21，並因應 RDA 規則著錄相當多與個人資訊相關的欄位，為了提升名稱辨識度，除了儘量蒐集個人相關名稱之外，也力求記載其生卒日期、學科專長、重要學經歷及相關團體機構、相關地名、資料來源等。

在轉換 MARC21 的過程中，試圖以程式將過去 CMARC 紀錄以段落敘述所涵蓋的生卒年、學經歷等資料進行格式化處理。生卒年的格式較固定，因此較易於將原欄位 100 分欄 d 的生、卒年拆解，分別轉換至 046 記載出、卒年的專屬分欄 f、g；至於學經歷資料段落雖以「;」區隔各階段學經歷，但學經歷常會穿插於其他背景資料敘述中，難免產生誤判情況，因此須投注相當多人力進行修訂。

(三) 樣例資料篩選及分群

由於 SMRT 人名權威資料先經程式比對整併，再由人工陸續整理修訂，為了選用較完整且正確的資料為研究樣例資料，因此以經過人工修訂、整併、新建的 262,737 筆紀錄為樣本母體。承上述國際通用規範相關說明及名稱資料庫消歧作法，依據編目規範選定 VIAF、ORCID、ISNI 及 SMRT 等資料庫較普遍具備的欄位，以 MARC21 權威記錄欄位而言，亦即欄位 100 分欄 d 之生卒年（或朝代）、欄位 100 分欄 c 之專

長學科領域、欄位 370 之相關地名、欄位 373 之相關團體、欄位 670 之資料來源等 5 項，此 5 項資料將作為後續討論的消歧重點欄位及樣本資料篩選條件。

為了解 SMRT 資料經過人工修訂後的消歧重點欄位分布情形，再從二十六萬餘筆樣本母體中，依據記錄中具備一或多項消歧重點欄位進行資料集篩選，篩選結果統計如表 3。

表 3
SMRT 權威紀錄樣例資料集篩選統計

資料集編號	資料集篩選條件	數量 (筆)
A	有 朝代	15,780
B	有 生卒年	17,384
C	有 學科領域	118,474
D	有 相關地名	111,111
E	有 相關機構	171,198
F	有 資料來源	244,534
G	有 生卒年 或 朝代	33,164
H	有 學科領域 或 生卒年	138,257
I	有 相關機構 且有 學科領域 或 生卒年	108,155
J	有 資料來源 但無 學科領域 或 生卒年	124,224
K	5 項條件皆具備	5,534

資料集篩選結果顯示 SMRT 現有經過人工修訂的記錄中，具備 5 項消歧重點欄位之一者，以具備「資料來源」的資料集 F 為數最多 (24 萬 4,534 筆)，但由於此欄位資料與書目資料著錄相關，不擬於本文討論。而後資料集數量多寡依序為資料集 E「相關機構」(17 萬 1,198 筆)、資料集 C「學科領域」(11 萬 8,474 筆)、資料集 D「相關地名」(11 萬 1,111 筆)，但此三種資料在 VIAF 中文紀錄中並不多見，難以進行著錄比較。為考量選用與 VIAF 較互有交集的消歧資料欄位，所以選用具備「生卒年 (或朝代)」的資料集 G (3 萬 3,164 筆) 做為查找 VIAF 相對應紀錄及著錄比較的樣本。

(四) 查找 VIAF 相對應紀錄結果

由於生卒年是 VIAF 人名權威紀錄最主要的消歧依據，因此從前述篩選的含生卒年或朝代的資料集中，擷取前 6,000 筆紀錄，再逐一至 VIAF 查找相同人物紀錄，查

找結果無相同名稱的紀錄計 3,004 筆，有相同名稱者計有 2,996 筆，但其中有 327 筆因無其他資料，而無法確定是否為相同人物。

由於 VIAF 為以群集形式消歧，查詢結果通常有多種來源紀錄，系統回應的查詢結果中，各來源紀錄內容也有不同程度的差異。再進一步檢視時，發現人物有多個名稱時，VIAF 群集紀錄中主要名稱雖不同，但經過消歧處理，仍能在同一人物的不同來源紀錄之間建立群集關聯，例如以作者筆名「一信」查詢 VIAF 紀錄，美國國會圖書館為作者之筆名及本名「徐榮慶」各建一筆紀錄，但主要名稱不同的紀錄已被歸為同一群集。

再擷取前述 6,000 筆中的前 500 筆，做為第 2 次樣本資料，進一步檢視相對應的 VIAF 紀錄的生卒年著錄情形，並將年代著錄異同、辨識是否相同遭遇的問題，記錄於 Excel 中（如圖 1）。這階段記錄的著錄比較差異，包括 VIAF 是否以中文為主標目、中文主標目與 SMRT 的差異、生卒年著錄差異等。

主要名稱	生卒年	專長學科	Tag996_VIAF_ID	說明	中文主標	中文主標note	VIAF年代符合	VIAF無年代	VIAF年代說明
蔣復璁	1898-1990	圖書館學	Sa60530451\$uhttp://viaf.org/viaf/60530451				1		另有1898-?
姚雪垠	1910-1999	中國文學	Sa22152833\$uhttp://viaf.org/viaf/22152833				1		
姚朋	1926-	中國文學	Sa95680573\$uhttp://viaf.org/viaf/95680573				1		
何建章	1926-2004	經濟	Sa121927557\$uhttp://viaf.org/viaf/121927557				1		20th century
李再鈞	1928-	雕塑	Sa39659002\$uhttp://viaf.org/viaf/39659002				1		
喬啟明	1897-1970	農業經濟	Sa95107457\$uhttp://viaf.org/viaf/95107457				1		
劉其偉	1912-2002	書畫	Sa98373630\$uhttp://viaf.org/viaf/98373630				1		無卒年
林天從	1924-2001	繪畫		1					
桂裕芳	1930-	法語	Sa282713076\$uhttp://viaf.org/viaf/282713076	無法確認(無相同著作)			1		20th century
馮承鈞	1887-1946	歷史	Sa51743750\$uhttp://viaf.org/viaf/51743750			1 無卒年	1		另有1885-1946
易夢虹	1916-1991	經濟		1					
張鐵生	1904-1979	政治	Sa70803344\$uhttp://viaf.org/viaf/70803344	無法確認(無相同著作)					1
張靜廬	1898-1969	出版學	Sa10135290\$uhttp://viaf.org/viaf/10135290				1		
申學庸	1929-	音樂		1					
薩孟武	1897-1984	政治	Sa112091256\$uhttp://viaf.org/viaf/112091256		1 無卒年		1		另有1898-1984
安作璋	1927-	歷史	Sa5527997\$uhttp://viaf.org/viaf/5527997				1		
傅斯年	1896-1950	歷史	Sa37753908\$uhttp://viaf.org/viaf/37753908		1		1		
李羨林	1911-2009	文學	Sa111257457\$uhttp://viaf.org/viaf/111257457		1 無卒年		1		
劉鴻喜	1923-	地理	Sa308746757\$uhttp://viaf.org/viaf/308746757	同領域,但無相同著作					1
周聯華	1920-	基督教	Sa13784131\$uhttp://viaf.org/viaf/13784131						1
馬空群	1909-2007	政治	Sa77795876\$uhttp://viaf.org/viaf/77795876	年代相距大			1		另有1913-
戴笠	1897-1946	政治	Sa122169518\$uhttp://viaf.org/viaf/122169518				1		另有1896-1946
蔣作賓	1884-1942	外交	Sa23583897\$uhttp://viaf.org/viaf/23583897				1		
劉鴻儒	1930-	金融	Sa30866509\$uhttp://viaf.org/viaf/30866509				1		20th century
姜龍昭	1928-2008	戲劇	Sa72744841\$uhttp://viaf.org/viaf/72744841				1		無卒年

圖 1 SMRT 樣本於 VIAF 查詢結果紀錄表樣圖。資料來源：本文作者整理。

經過詳細比較結果，其中有 92 筆記錄確認為同人物，但生卒年與 SMRT 有所出入，包括：缺卒年、著錄為「19th century」或「20th century」、生年或卒年有 1-3 年的差距等。由於紀錄中有其他輔助辨識的資料，因此，即便年代有所差異，仍可判

定為相同人物。然而 VIAF 無年代者，則須進一步檢視相關名稱或書目資料，才能判斷兩者是否屬同一人物資料。

(五) 消歧常用欄位之建置情形

從前述篩選結果來看，SMRT 最常用的人名權威紀錄消歧欄位為生卒年及專長學科領域，由於生卒年資料取得因人而異，因此，紀錄中含專長學科領域者為數較含生卒年者多。目前專長學科之著錄依據國圖公布的「國家圖書館人名權威檔個人專長學科領域表」(簡稱「學科領域表」)控制詞彙，在無法查得其他個人資料情況下，編目員通常會依據作者的作品主題，從學科領域表擇定較接近的詞彙著錄之。

由於學科領域日異專精及多元，學科領域表的詞彙並無細分的學科名稱，以致在建立諸如學位論文作者的權威紀錄時，造成詞彙不足的情形，在缺乏其他區別資訊情況下，常見同姓名且同學科領域，但又缺生卒年的權威紀錄。

生卒年是人名權威消歧最直接的依據，雖然同姓名人物也可能有同年出生或死亡的情形，但在機率相當低的情況下，生卒年長久以來是歐美圖資界個人權威紀錄最常用的消歧欄位。由於生卒年絕大部分取自二手資料來源，因此，針對同一人物在不同資料來源獲取的生卒年代也可能有不一致的情形。此外，依據中國編目規則，民國之前的中國人物以朝代區分，若無其他據以區分的資料，同姓名且身處同朝代的人物，就會產生權威名稱相同的紀錄。

權威紀錄建置過程中，人物生平資料的蒐集相當重要，這些資料通常包括重要的生老病死的時空、學歷、經歷、專長技藝、重要著作等資料。這些資料在過去是以段落敘述形式著錄，近兩年開始則改採 MARC21 的欄位 373，將人物相關的各個團體機構及年代，以重複欄位 373 的方式分別著錄，讓生平經歷資料格式化呈現。

五、結語

國際間較具規模的名稱權威資源 VIAF 的合作對象主要為各地國家圖書館，而 ORCID 則以研究人員或機構為資料提供來源，ISNI 則收錄各機構已彙整的批次檔為主。看似獨立的資源，資料卻交互再利用，因為 VIAF 是 ISNI 的資料基礎，而 ISNI 也正進行匯入 VIAF 測試，ISNI 與 ORCID 於 2014 年 1 月簽訂策略合作備忘錄，雙

方達成共識進一步於互通技術進行合作 (ISNI International Agency, 2014b)。雖然 VIAF 建置較早，但由於 ISNI 已成為 ISO 標準，ISNI 目前收錄 750 萬筆個人或團體權威紀錄，個人或團體權威紀錄是圖書館最需要的權威控制資源，法國國家圖書館於 2014 年 1 月成為辦理 ISNI 給碼的圖書館之後，ISNI 未來是否會成為圖書館權威控制的名稱權威標準號，值得臺灣圖書館界密切注意。

由於 VIAF 的合作對象為國家圖書館，且該資料庫以 MARC21 為基底，目前國圖 SMRT 系統的中文人名權威採用 MARC21 建檔，匯入 VIAF 在資料處理上較無問題，因此，資料匯入 VIAF 是國圖近程目標。然而，不論針對國內或國際合作，基於提升人名權威消歧的成效，國圖與其合作單位積極商討下列中文人名權威紀錄建檔相關議題：

(一) 資料處理需考量中文特性

中文人名權威紀錄和西文權威紀錄共通的問題包括輸入錯誤、資料缺漏等，而中文因其文字特性，在資料著錄時會因建檔機構的政策而衍生歧異，例如同一人物名稱在臺灣以繁體著錄，在大陸則以簡體著錄；姓名各字元若含有異體字(如「群」與「羣」)，則須考量人物慣用哪一個字形。從資料查詢角度，增列參照名稱可解決這些問題，例如 CALIS 以簡體名稱為主要名稱，但以重複 CNMARC 欄位 100 的方式，同時著錄作者的繁體姓名，不失為一種值得參考的彈性做法。至於檢索、比對相關程式應該要具備簡、繁、異體字碼互查功能，但字碼交互查詢也可能查獲很多不需要的資料，因此，納入字碼交互查詢機制的適切點應多加衡量。

(二) 消歧常用欄位之著錄建議

對於同姓名的人物，消歧常用且較有效的欄位不外乎生卒年月日、專長學科領域、職業、所屬機構等資料，國內圖書館對於這些資料著錄須考量的問題如下：

1. 即時著錄區別資訊

圖書館在建立作者名稱權威紀錄時，通常不著錄區別資訊，直至需建立同姓名的第二位作者權威紀錄時，才加註學科領域、生卒年月日等區別資訊。這些缺少區別資訊的權威紀錄在各館內部系統使用雖不成問題，但若考量進行資源交流合作，應儘可能在編目時，即時將待編資料上有助區別名稱的重要資料著錄於權威紀錄中。

2. 記載可能的生卒年代資訊

從本研究樣例查詢結果，可看出生卒西元年代是名稱消歧最直接的依據，但不同參考來源對於同一人物的生卒年記載不同時，除了要再多方查證之外，如果參考資料皆相當有可信度，在保留所有可能的資訊前提下，雖然不可重複 MARC21 欄位 046 分欄 f (Birth date) 及分欄 g (Death date)，但欄位 046 本身可重複，建議採重複 MARC21 欄位 046 的方式，記載多種可能的生卒年資訊。

3. 朝代之外加註西元年代

國內圖書館對於民國以前的人名權威，通常僅以朝代進行區別，但這種區別資訊無法應用於國際平臺，建議在可獲取生卒西元年代資訊情況下，仍將西元年另著錄於 MARC21 欄位 046。

4. 研訂相關控制詞彙

「職業」將逐漸成為區別人物的資訊之一，目前圖書館界除了中文主題詞表中散置各處的職業名稱之外，並無專門的職業詞彙表，編目員在著錄時相當困擾。建議參考國內勞動部「職業分類資訊查詢系統」的詞彙，結合中文主題詞表已有的職業名稱詞彙，訂定較完整的職業詞彙表。

5. 檢討專長學科領域所在欄位

國內圖書館慣用專長學科領域進行名稱區別，CMARC 轉換 MARC21 對照表將 CMARC 的專長學科領域名稱轉換至 MARC21 的欄位 100 分欄 c，但因分欄 c 原本定義用於著錄人物的頭銜、學位、世代數等資訊，專長學科領域著錄於此與定義有所出入，建議應檢討並評估將此資訊記載於定義較接近的 MARC21 欄位 374 (Occupation) 的可行方式。

6. 強化及運用團體機構名稱權威資源

與個人相關的團體機構名稱在人名消歧過程中，也逐漸扮演重要的辨識角色，相對凸顯團體名稱權威控制的重要。國圖過去曾建立臺灣地區政府及大學校院名稱權威紀錄，但並無人力隨時更新各機構因應組織改制之名稱變更。在專注個人名稱權威建置的同時，也應思考圖書館界共同兼顧中文團體名稱權威資源強化之可行方案。

（三）及早善用國際識別符號

從 VIAF、ISNI、ORCID 的資源來看，識別符號已成為不同資源之間最直接的相互辨識資訊。CMARC 雖然以欄位 7XX 段的分欄 3 著錄「權威紀錄號」，此分欄設計目的是從著錄機構的地區性識別號碼，發展為國際性號碼。但由於權威紀錄號通常是系統為新建紀錄產生的編碼，系統一旦汰換，紀錄號碼也須隨之更改，因此這類號碼在各館內部難保其沿用性，從資料交流觀點，亦難具互通性。VIAF、ISNI 的發展既已獲得國際圖資社群肯定及採用，其公開的識別符號已具相當的穩定度，建議國內圖書館也應及早將此兩項資源的識別符號著錄於 MARC21 欄位 024 其他標準識別符號。未來在語意網環境應用的中文人名權威紀錄若包含國際通用的標準識別符號，不僅增加資料辨識度，也相對提升紀錄的可信度。

（四）鼓勵研究人員申請國際識別符號

從 FRAD 概念模型及 RDA 相關規則來看，人名權威對於個人背景描述較以往更多，在個人資料保護法施行後，圖書館蒐集個人相關資料必定有更多顧慮。在個人資料保護法規定中，國圖以公務機關進行人名權威紀錄整理建檔，蒐集對當事人權益無侵害之個人資料，依據個人資料保護法（簡稱「個資法」）第十五條第三款可合法蒐集。另一方面，目前人名權威紀錄的資料來源絕大多數取自網際網路，依據個資法第九條第二款，應屬當事人自行公開或其他已合法公開之個人資料，因此得免為告知。即便如此，當紀錄所屬當事人提出異議時，圖書館仍須謹慎處理。

由於精確辨識名稱對於研究人員學術產出之蒐集、評估相當重要，而國際識別符號又是個人名稱重要的指引，上述個資蒐集的顧慮在大學校院或學術機構或許有其他因應方式。建議國內大學校院或學術機構圖書館，可透過公開說明方式，讓其母機構及所屬研究人員了解申請國際識別符號的重要性，並鼓勵研究人員個別或由機構集體向 ORCID 或（及）ISNI 申請識別符號，建立研究人員專屬紀錄，並透過 ORCID 或 ISNI 既有的資料授權公開機制，設定同意公開的資料。圖書館在建立研究人員的權威紀錄時，只要著錄基本區別資訊，例如：專長學科領域，並將此國際識別符號及該筆紀錄 URI 著錄於紀錄中，再藉由超連結方式查閱其公開的資料。在進行權威紀錄國內外交流過程中，將輾轉再利用此識別符號，可逐步提高個人名稱消歧結果的精確度。

(五) 採用分階段進行消歧處理

依據 SMRT 樣例紀錄於 VIAF 查詢相同人物資料的結果來看，採用分階段比對消歧較為理想。由於名稱消歧最直接方式是名稱＋生卒年代，但 VIAF 許多華裔人物紀錄並無中文名稱，因此，相當程度須以羅馬拼音名稱＋生卒年的組合進行消歧。又因同一人物的相關年代可能有些微差距，而且因紀錄無法隨時更新之故，年代也可能有所缺漏，因此，以程式比對年代異同需容許若干年代數值的誤差。第一階段以此兩項資料組合比對符合者，應可判斷為相同人物。接下來再針對其餘疑似相同者，另採用職業、專長領域等欄位進行判斷，但因職業、專長領域等詞彙，目前並無國際通用詞彙，較難據以消歧。就 VIAF 處理方式來看，合作機構必須提供權威紀錄相關書目紀錄供其據以進行消歧作業，因此，書目資料著錄是否一致，也攸關權威消歧成效。

綜觀本次研究，在網路資源隨手可得的時代，諸多當代人物個人相關資料雖然可在網際網路上查得，但基於保護個人隱私之考量，有助消除名稱歧異的資料並非皆可轉載顯示於各系統的查詢結果。此外，個人創作背景、作品有關之網路資訊，雖有助於人工檢視辨識，但若要再利用於一般圖書館之個人名稱自動消歧程式，則需先經過相當程度的程式或人工處理。因此，權威紀錄建置或識別符號管理機構必須不斷精進其消歧程式，另一方面，紀錄所屬人物持續提供或直接修訂紀錄內容，也相同重要 (Gatenby, 2011)。從權威資源建置或識別符號管理機構角度而言，鏈結資料技術雖是強化資源互通的可行途徑，但資源內容的完備度則影響鏈結技術應用的效益，因此，增進資料正確性和即時性，及時隨著資訊處理技術或資料格式演變，適度調整消歧機制，並透過親和易用的介面，提高權威紀錄所屬人物持續更新紀錄的意願，唯有資料管理者及使用者雙方共同努力，才能逐步有效提升權威資源的整體品質。

參考文獻

- 陳和琴 (2011)。FRBR、FRAD、FRSAD 發展介紹。在國家圖書館編目組 (編輯)，*100 年資訊組織進階班研習手冊* (43-80 頁)。國家圖書館，中華民國圖書館學會：臺北市。
- 張菲菲、李宗海、周曉輝、李曉戈 (2010)。基於層次聚類的跨文本中文人名消歧研究。*電腦工程與應用*，50(6)，106-111。

- 國際圖聯規範紀錄的功能需求與編號 (FRANAR) 工作組 (2009)。 *規範資料的功能需求：概念模型*。檢自 http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-zh.pdf
- American Library Association, Canadian Library Association, & CILIP: Chartered Institute of Library and Information Professionals. (2014a, October 14). *9 identifying persons* [in RDA Toolkit]. Retrieved from <http://access.rdatoolkit.org/rdachp9.html>
- American Library Association, Canadian Library Association, & CILIP: Chartered Institute of Library and Information Professionals. (2014b, October 14). *RDA elements mapped to MARC 21 format for authority data*. Retrieved from <http://access.rdatoolkit.org/jscmap3.html>
- Bilder, G. (2011). *Disambiguation without de-duplication: Modeling authority and trust in the ORCID system*. Retrieved from http://0-labs.crossref.org.library.utulsa.edu/wp-content/uploads/2013/04/disambiguation-deduplication_wp_v4.pdf
- Cals, J. W. L., & Kotz, D. (2008). Researcher identification: The right needle in the haystack. *The Lancet*, 371(9631), 2152-2153. Retrieved from <http://search.proquest.com/docview/199022661?accountid=7979>
- Carpenter, T. A. (2014, March 12). Name identification using the ISNI: An interview with Laura Dawson. *The scholarly kitchen*. Retrieved from <http://scholarlykitchen.sspnet.org/2014/03/12/name-identification-using-the-isni/>
- Chen, K. H., & Hsieh, C. N. (2011). Ambiguity resolution for author names of bibliographic data. *Journal of Education Media & Library Sciences*, 49(2), 215-240. Retrieved from <http://joemls.dils.tku.edu.tw/fulltext/49/49-2/215-240.pdf>
- Gatenby, J., & MacEwan, A. (2011). ISNI: a new system for name identification. *Information Standards Quarterly*, 23(3), 4-9. Retrieved from http://www.isni.org/filedepot_download/143/419
- Han, H., Giles, L., Zha, H., Li, C., & Tsioutsoulis, K. (2004). *Two supervised learning approaches for name disambiguation in author citations, proceedings of the 4th ACM/IEEE-CS joint conference on digital libraries, Tuscon, AZ, 296-305*. Retrieved from <http://clgiles.ist.psu.edu/papers/JCDL-2004-author-disambiguation.pdf>
- Hickey, T. B. & (2011, July 6). Re: VIAF and other IDs [Web blog message]. Retrieved from <http://outgoing.typepad.com/outgoing/2011/07/viaf-and-other-ids.html>
- Hickey, T. B., & Toves, J. A. (2014). Managing ambiguity in VIAF. *D-Lib Magazine*, 20(7/8). Retrieved from <http://www.dlib.org/dlib/july14/hickey/07hickey.html>

- Hill, A., Needham, D., & Danskin, A. (2009). Names project: Final report. *JISC*. Retrieved from <http://names.mimas.ac.uk/assets/files/Names-phase-one-final-report.pdf>
- ISNI International Agency. (n.d.). *How ISNI works*. Retrieved from <http://www.isni.org/how-isni-works>
- ISNI International Agency. (2013, September 4). *ISNI data element values*. Retrieved from <http://www.isni.org/content/documents-related-data-submission>
- ISNI International Agency. (2014a). *BnF: First national library in the world to become an ISNI registration agency*. Retrieved from <http://www.isni.org/content/bnf-becomes-first-national-library-world-become-isni-registration-agency-5>
- ISNI International Agency. (2014b). *ISNI and ORCID sign memo of understanding*. Retrieved August 18, 2014, from <http://www.isni.org/content/isni-and-orcid-sign-memo-understanding>
- OCLC. (2012, April 4). *Virtual International Authority File service transitions to OCLC; contributing institutions continue to shape direction through VIAF Council*. Retrieved from <http://www.oclc.org/news/releases/2012/201224.en.html>
- OCLC. (2014, August 28). *2014 annual report to VIAF council*. Retrieved from <http://oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf>
- OCLC. (n.d.). *VIAF tag documentation*. Retrieved from <http://viaf.org/viaf/viaftags.xml>
- ORCID Inc. (2013, March 19). *ORCID privacy policy*. Retrieved from <http://orcid.org/footer/privacy-policy>
- ORCID Inc. (n.d.). *Names in the ORCID registry*. Retrieved from <http://support.orcid.org/knowledgebase/articles/142948-names-in-the-orcid-registry>
- Smalheiser, N. R., & Torvik, V. I. (2009). Author name disambiguation. *Annual Review of Information Science and Technology*, 43(1), 1-43.
- Taffet, M. D. (2004). *Looking ahead to person resolution*. Retrieved from http://fht.byu.edu/prev_workshops/workshop04/FHTCD/Session1/Abstract_Mary_Taffet.pdf
- Wikipedia. (2013, June 2). *Disambiguation (metadata)*. Retrieved from [http://en.wikipedia.org/wiki/Disambiguation_\(metadata\)](http://en.wikipedia.org/wiki/Disambiguation_(metadata))