



臺大醫圖利用WOS及JCR資料庫進行機構 研究表現評估之經驗談

應家琪（推廣服務股）

學術機構研究表現的評估方法有許多種，其中一種方式便是以Web of Science資料庫（以下簡稱WOS）為依據，提供SCI論文表現的相關指標，以供長官瞭解機構的研究表現及定位。筆者自民國90年起開始接觸此項業務，從早期以人工方式進行論文篇數的計算，到「臺大醫學校區研究競爭力分析系統」（以下簡稱研究力系統）的開發及運作，再再顯示長官們的資訊需求已非人工計算的方式可以滿足，因此，走向自動化方式乃必然之趨勢。（註1）

筆者每年會主動從WOS資料庫檢索相關資料後，再匯入研究力系統中，透過研究力系統的運作（每年更新3次），將WOS及Journal Citation Reports（以下簡稱JCR）兩個資料庫的資源加以整合，以提供相關統計數據。在提供資料方面，原則上有兩種方式，一種是主動提供；一種是被動配合院方或各單位的需求提供資料。每年主動製作並呈上呈長官的資料包括：臺大醫學校區（含醫學院、醫學院附設醫院及公衛學院）各個單位的論文發表篇數、投稿期刊的平均IF值及每位教師平均發表的論文篇數。此外亦會針對國內外醫學相關機構進行論文發表情形的比較，其項目包括：論文篇數、被引用次數、平均被引用次數、刊登期刊的IF值、刊登期刊的平均IF值。至於被動配合需求所提供的資料，則更多樣化，假使從研究力系統中取得的資料即可滿足所需時，就比較單純，若長官所需的資料並非研究力系統中的資料可以滿足時，則必須依據個案需求另外進行檢索，例如：直接檢索WOS或Essential Science Indicators（簡稱ESI）資料庫。在進行前述作業時，曾經遇到一些狀況，包括：

1. 作者在發表論文時，未清楚註明單位名稱。例如作者只註明臺灣大學，未進一步註明第二層級（醫學院、醫院或公衛學院）及第三層級單位（如內科或內科部）以致無法檢索到該篇論文；或者單位名稱寫法有問題，例如：「Department of Urology, En Chu Kong Hosptial, National Taiwan University, Taiwan」，此寫法係將恩主公醫院寫在臺灣大學之下，必須進一步找出原文才能判斷是斷詞錯誤，或者是作者寫法有問題，且須確認該作者是否屬於醫學校區後，方能加以歸類或刪除。
2. 在WOS資料庫曾經發現應該收錄卻漏收錄的情形，以致無法檢索到該篇論文。
3. WOS資料庫中作者任職機構（即Address欄位）著錄錯誤的問題，包括建檔錯誤、斷詞錯誤等，以致研究力系統無法自動辨識，必須輔以人工方式加以釐清及歸類，方能將該論文歸入到作者所屬的單位中；例如：原文中的寫法為「^aDepartment of Internal Medicine, Far-Eastern Memorial Hospital, and ^bDepartment of Internal Medicine, National Taiwan University Hospital, College of Medicine, National Taiwan



- University, Taipei, Taiwan」，但WOS資料庫中則著錄成「Nat'l Taiwan Univ, Far Eastern Mem Hosp, Dept Internal Med, Taipei 10764, Taiwan.」，誤將遠東紀念醫院著錄成臺灣大學之下的二級單位。
4. WOS資料庫有延遲收錄的狀況，所以無法及時提供完整的收錄篇數。例如：Epub Ahead of Print（為紙本期刊出版前的電子版）、Accepted（已被接受刊登，但尚未正式出版）及In Press（正在出版中的論文）等論文，因為在WOS資料庫中，尚無法查得任何資訊，以致無法提供相關統計數據。
 5. 對於其他機構掌握有限，因此在進行WOS資料庫檢索時，檢索策略不容易下得很周全，擔心有所遺漏或疏失。

基於上述狀況，本館的因應之道便是建議院方發函給各單位，宣導同仁於發表學術論文時，應該以統一的格式及英文名稱註明任職機構，以避免因任職機構註明不清楚而無法檢索到該篇論文。如發現問題是屬於資料庫製作者時，則會將問題反應給資料庫代理商，以使其解決相關問題，然而，我們也發現，因為WOS資料庫中的資料龐大，所以要期待該公司在短期內清查或解決資料庫中所有的問題似有困難，也因此本館委請廠商開發研究力系統時，便考量到人工判斷的機制，以系統自動歸類為主、人工判斷為輔，釐清一些有問題的書目資料，將錯誤在可控制的範圍內降至最低。

當然，本館在提供統計數據時，也曾遇到科部的質疑，但經查核原因，通常是因為論文資料類型及資料庫收錄時間差的問題。因為我們提供統計數據時，並未過濾文獻的資料類型，而是將所有類型予以統計，包括：Article、Letter、Editorial material、Meeting abstract、Review等，除此之外，前文曾提及有關In Press等論文，因尚未被資料庫收錄所以無法進行統計的問題，也是經常被科部提出質疑。但該類論文日後一旦被資料庫收錄，如無特殊狀況或問題，應是可以順利被檢索得到並加以計算的。此外，因為一篇論文發表後不可能立刻被WOS資料庫所收錄，因此，從論文出版後到真正被資料庫收錄是有時間差的，而且不同時間點進入資料庫檢索，所得的檢索結果也是有差異的，這些狀況皆屬正常，只要向科部說明清楚，即可被接受。

再者，有關其他機構名稱（尤其是國外機構）掌握有限的問題，幾經思考並參考黃慕萱老師所著的「擇優公布國內前20%大學名單」一文後（註2），未來應該會改以學科為基礎進行評比，而非從機構的角度切入進行評比。除了以機構名稱（例如：校名、院名、科系名）進行檢索怕有遺漏之外，不同學校的學院包含的科系亦非全然相同。若改以學科主題進行評比（例如：以校名進行檢索後再設限某些主題），問題就比較小，而且可以將不同機構放在同一基準點上進行比較，結果應較客觀。例如可以針對醫學領域挑選相關學科主題進行比較，而且所挑選的主題，是經過學科專家確認後，並在母機構長官可以認同的主題範圍下進行比較。如此，便能觀察不同學校在某些學科發表論文的情形。不過，在挑選學科主題時，



若挑選太過廣泛或不當，則容易失焦，以臺大醫學院為例，若評比的主題太過廣泛時，則會包含太多校總區（即臺灣大學）的績效，如此，呈現出來的結果，就不夠客觀，必須再針對主題挑選的結果加以調整。

近兩年因為醫院方面引進平衡計分卡（Balanced Score Card，以下簡稱BSC）來衡量各單位的績效，為配合BSC作業，亦提供不同的衡量指標（例如：IF>N或IF學科排名前N%），而不再單純的只看SCI論文篇數。以IF學科排名前N%而言，院方會請各單位提出較常發表的學科領域，並與各單位談妥排名N%後，再由本館計算出各單位發表於學科領域前N%的SCI論文篇數。另因為JCR資料庫每年大約在暑假期間才會出版前一年度的資料（原則上2009年暑假會出2008年版），但BSC作業要呈現出來的是當年度的成效，所以在新版JCR尚未出版前，一律以前一年度的版本為依據，以取得IF值。亦即，在2008年版的JCR尚未出版前，則2008年的IF值是取自2007年版的JCR。此項作業若以人工執行肯定會耗費相當的成本，故因應該項作業之需求，本館特將該作業予以自動化，將相關需求納入研究力系統中，利用系統快速的計算出相關篇數，並確保資料的正確性。

以上，便是本館在利用WOS及JCR兩資料庫，進行機構評估表現的一些經驗。未來當然會隨時配合需求或資料庫的功能加以調整，例如可以考慮針對提供的指標進行修正，亦即在進行國內外機構發表論文情形之比較時，可以加入h-index值，以去除極值的影響，並去除單看論文質或量的偏頗，提供一個簡單易懂、綜合評估質與量的指標供長官參考。（註3）針對BSC作業提供的篇數，未來應該針對文獻類型予以挑選，例如是否要以「Article」為主即可。

文末，想說明的是，雖然筆者長期提供SCI相關統計供長官參考，但筆者仍明白以WOS及JCR做為評量工具，是有其限制的。而且，並非所有重要文獻都是可以量化計算的，就算量化計算後的結果相對較低者，亦不能代表該文獻就不是重要的文獻。最後再強調一點，每個機構計算SCI論文發表篇數的方式或有不同，各機構可能會基於某些情況或政策之考量，決定採用何種計算的標準，並無所謂對與錯，只要能將計算方式的限制說明清楚，應該就是可以被接受的。謹以此文與同道分享之。

註釋

註1：應家琪，「執行SCI統計作業之心得」，中華民國圖書館學會醫學圖書館委員會30週年紀念特刊（民國97年12月）：頁41。

註2：黃慕萱，「擇優公布國內前20%大學名單」，評鑑月刊第7期（民國96年5月）：頁9-22。

註3：黃慕萱，「統計小辭典-h-index指標」，評鑑月刊第11期（民國97年1月）：頁31。