

解讀ESI基本科學指標

文／黃慕萱

國立臺灣大學圖書資訊學系教授兼系主任
高等教育評鑑中心兼任研究員

自1963年美國科學資訊研究所(Institute for Scientific Information, 簡稱ISI)發表《科學引文索引》(Science Citation Index, 簡稱SCI)以來, ISI又陸續於1973年發表《社會科學引文索引》(Social Sciences Citation Index, 簡稱SSCI)、1978年出版《人文與藝術引文索引》(Arts & Humanities Citation Index, 簡稱A&HCI), 收錄各學科的優質期刊文獻。1997年ISI配合線上資料庫發展趨勢, 而將此三專輯由光碟版轉為網路版本, 合稱Web of Science(WOS), 提供收錄期刊的「單篇文章」被引用狀況, 可藉此得知各篇文章的引用文獻(Cited Reference)與被引次數(Times Cited)。

論文發表量的排名統計工具

ISI後來又以「科學家、機構、國家、期刊」的角度, 分析WOS中文章的引用及被引用狀況, 於2001年推出《基本科學指標》(Essential Science Indicators, 簡稱ESI), 將SCI/SSCI所收錄超過11,000種期刊的最近十年文章, 分為22個學門, 針對被引次數世界前1%的科學家與機構、前50%的國家與期刊, 分別統計其論

文數、被引用總數、平均被引用數與此三項指標之世界排名榜。ESI同時提供論文數、被引用總數、平均被引次數的各學門平均值, 及高被引文章(Highly cited papers)、熱門文章(Hot Papers)與研究趨勢(Research Fronts)等資訊, 提供使用者快速了解當今科學論文發表現況、優質文獻與前瞻議題。

故於ESI中, 我們可針對最近十年的文章, 進行不分學門及各學門的科學家、機構、國家與期刊之排序, 排序方法可選擇以論文數、被引用總數, 或是平均被引用數進行。亦可於ESI上查詢各國、各機構或各科學家所發表的高被引文章、熱門文章數量, 或比較各領域間平均論文或引文量之特性與差異。然在實際使用上, ESI的設計和指標選擇存在一些問題, 應用時須相當小心。本文將針對ESI資料庫, 介紹其收錄內容與時間、功能、經常被使用的指標, 及需特別留意之問題。

收錄內容與收錄時間

ESI收錄近一千萬篇來自SCI/SSCI收錄期刊之文章, 含定期科學性文章(regular scientific articles)、評論文章(review

articles)、會議論文(proceedings papers)、研究紀要(research notes), 但不包含期刊中的讀者來信(letters to the editor)、更正通知(correction notices)、摘要(abstracts)等內容。資料庫中隨時保持最近至少十年的文章, 而以每二個月更新之速度累積十年二個月、十年四個月...之資料量, 待時間達十年十個月後, 下次更新將會把起始年份向後移一年, 亦即將總年份減為十年。例如若目前ESI收錄之年份為1996年1月1日至2006年10月31日共計十年十個月之資料量, 當下次更新時, 起始年將往後移, 改為1997年1月1日至2006年12月31日共十年。

ESI所收錄的期刊會被分為22個學門, 再依學門進行各項統計。這些學門包含: **工學**領域之電腦科學(Computer Science)、工程(Engineering)與材料科學(Materials Sciences); **生命科學**領域之生物與生化(Biology & Biochemistry)、環境/生態學(Environment/Ecology)、微生物學(Microbiology)、分子生物與遺傳學(Molecular Biology & Genetics); **社會科學**之一般社會科學(Social Sciences ,General)、經濟與商業(Economics & Business); **理學**領域之化學(Chemistry)、地球科學(Geosciences)、數學(Mathematics)、物理學(Physics)與太空科學(Space Science); **農學**之農業科學(Agricultural Sciences)、植物與動物科學(Plant & Animal Science); **醫學**之臨床醫學(Clinical Medicine)、免疫學(Immunology)、神經科學與行為(Neuroscience & Behavior)、藥理學與毒物學(Pharmacology & Toxicology)、精神病學/心理學(Psychology/Psychiatry); 及跨學科

(Multidisciplinary)。

在ESI資料庫中, 每種期刊只會被分入一個學門; 只有被歸類為跨學科學門(Multidisciplinary field)的Science、Nature與PNAS期刊, 會被按照各篇文章的參考文獻(reference)與引用文獻(citation), 重新為每篇文章單獨分類, 但每篇文章仍只會被分類到一個學門。

ESI的功能

ESI可視為協助學術評估的重要參考工具之一, 在進行機構或國家的學術評量時, 若需要統計這些單位的論文發表量或被引量, ESI能快速提供參考指標。尤其它已將文章分類為科學家、機構、國家、期刊的排名狀況, 使用者能迅速獲得想要比較對象的資料, 也能分別統計各領域進入被引次數1%的科學家與機構, 與前50%的國家與期刊之入圍狀況。ESI所提供的高被引文章、熱門文章與研究趨勢, 更是方便的指標, 讓使用者能隨時得知目前領域的前瞻主題或優質文章。由於ESI已將收錄文章做過完整的分析統計, 對於使用者而言, 最大的優點就是節省時間, 能馬上使用這些統計數據, 配合加值運用獲得更高成效。

指標介紹

ESI只會列出被引次數達前1%的科學家與機構, 及前50%的國家與期刊, 我們可於ESI中查詢並排序這些入選單位十年的論文數、被引次數以及平均被引次數。除此之外, ESI另提供高被引文章、熱門文章與研究趨勢等資訊, 以下僅針對本次評估所使用的幾項指標進行說明:

◆ 論文數 (Papers) :

論文數代表在該段ESI收錄時間中，以該單位名義發表之論文總量，可表現出該單位學術研究成果的產量。

◆ 被引次數 (Citations) :

被引次數代表在該段ESI收錄時間中，以該單位名義發表之所有論文，於此段時間內被引用之次數加總。被引次數可反映該單位文章被利用、對日後科學進展造成影響的程度。

◆ 平均被引次數 (Citations Per Paper) :

平均被引次數為上述該單位之被引次數除以論文數後之平均數，意即估算每篇文章平均會被引用多少次，目的在評估該單位每篇論文的被利用程度或影響程度。平均被引次數若單獨使用，較易產生偏頗，須與論文數及被引次數配合使用。

◆ 高被引文章 (Highly Cited Papers) :

在該段ESI收錄時間中，將文章分年分學門評估，被引次數達前1%的文章即為高被引文章。因發表越久的文章被引次數越有可能較高，且應用科學類文章的被引次數也較可能高於基礎科學，此定義可排除時間與學門對引用次數造成的影響。值得注意的是，ESI中所有計算或比較的年度，都是以文章被收錄進系統的「資料庫收錄年」(database year)為準，並非文章的原始出版年(publication year)。

ESI的問題

然而，ESI設計和指標選擇上仍存在一些問題，本文將簡略整理幾項ESI較不為眾人討論之問題所在，期能藉此供大眾了解使用ESI之注意事項，以避免貿然比較而引發爭議。

一、人文社會科學收錄問題

由於ESI資料庫僅收錄SCI與SSCI之期刊，未包含A&HCI收錄期刊，故資料庫中所涵蓋人文社會科學文獻相當有限；且人文社會科學有地域屬性和語文限制等獨特發表特性，不見得適合直接以國際期刊評量。因此ESI資料庫可能無法代表人文社會學者的學術產出，較不適合用來評估以人文社會科學為主的大學。

二、缺乏彈性

ESI資料庫設計為統計十年至十一年的各項數據，無法變更資料年度如只統計五

年或單年資料；加上ESI主要統計項目固定為科學家、機構、國家與期刊，若使用者想以其他項目，如特定關鍵字分析論文或被引次數，亦無法如願。由於ESI將年代與主要統計項目固定，此項設計較無彈性，導致使用者無法依個人需求獲得更多資訊。

三、總論文數排名與平均被引次數排名

在ESI資料庫中，僅呈現總被引次數前1%之科學家與機構、前50%之國家與期刊之統計數據；意即即使該科學家或機構或國家或期刊之總論文數或平均被引次數已達世界前1%或50%，但若十年總被引次數未能達到門檻值(1%或50%)，則該單位或該人／期刊及其論文狀況仍不會出現於ESI資料庫中。因此ESI所得出的總論文數或平均被引次數的排名，無法說是「世界排名」，僅能說是針對總被引次數前1%的科學家與

機構、或50%的國家與期刊，再次進行之總論文數排序或平均被引次數排序。

四、不同學門各有文章發表特性

在科學家、機構、國家與期刊的分類架構下，還可再依22個學門與不分學門(All field)23個類別分別排序論文數、總被引次數與平均被引用數。事實上，不同學門文章本就會有不同的特性表現，如分子生物及遺傳學(Molecular Biology & Genetics)十年來的平均被引次數高達24.41次，而數學(Mathematics)的平均被引次數僅2.71，二者落差甚大。再加上ESI所分的22學門分類標準不同，如生物與生化(Biology & Biochemistry)、微生物學(Microbiology)、分子生物與遺傳學(Molecular Biology & Genetics)細分成三學門，工程(Engineering)則為一個學門數，土木和機械等學門均包含其中。且ESI的學門規模又各有大小，如臨床醫學(Clinical Medicine)十年來的論文數共為1,830,704篇，遠多於太空科學(Space Science)的112,183篇；臨床醫學(Clinical Medicine)的被引次數為19,681,810次，亦與電腦科學(Computer Science)的574,206次有相當大的差距。由於學門之間分類標準不同且其規模大小不等，再加上學門又可細分為性質各不相同的次學門，因此ESI論文數與被引次數排名，有時不能反應出學術表現的全貌。

同時，只要任一個學門的總被引次數進入前1%或50%的門檻值，該單位(或人名／國名／期刊名)便會出現於ESI中，此各學門的門檻值亦有極大差異，例如太空科學(Space Science)機構被引次數1%之門檻值為8,549次，而一般社會科學(Social

Science, General)機構之被引次數門檻值僅為469次。但無論該機構以任何領域超過門檻值而進入ESI，只要有一個以上學門超過，ESI即會統計該機構所有學門的各項數據。

此外，ESI所呈現的各學門機構數亦有相當大的差距，如同樣收錄被引次數前1%機構，經濟與商業(Economics & Business)中僅包含149個機構、數學(Mathematics)為160個機構，而臨床醫學(Clinical Medicine)卻有2,365個機構；事實上，醫學相關的機構數較多，甚至有許多僅發表少數論文的小型研究機構。由於臨床醫學母體數較其他學門為大，進入門檻值的機構數亦較多。

五、權威控制問題

在書目計量方法中，權威控制是確保統計結果正確性的重要因素，因文獻發表時，作者可能以各種文字寫法敘述來源，首先要能判別一個機構會出現多少寫法，如長庚紀念醫院可能會出現長庚婦幼醫院、長庚兒童醫院、長庚大學附設醫院等寫法，但其皆為同一單位，故應集中為同一個控制詞彙。在ESI中，四大分項(科學家、機構、國家與期刊)皆有進行單位的詞彙權威控制，惟仍存在少數問題，例如同名科學家即並未加以區別，所有同名者的表現同時被記錄於該名字中，故會出現如不分學門計算Wang, J十年來有7,002篇論文的現象。也因此ESI中的科學家排名狀況較少被用於評估工作。

其次，在機構項目中，ESI雖已進行每個機構多種寫法的控制，但對於機構從屬關係並未加以控制，因而出現有些機構有將所有附屬機構合併，而有些機構則

無之情況，如德州大學系統(University of Texas System)的九個分校(UT Arlington、UT Austin、UT Dallas等)文章於ESI中全被歸於Univ. Texas中，但加州大學系統(University of California)的十所分校則分別以Univ. Calif Los Angeles、Univ. Calif San Francisco、Univ. Calif Berkeley等單獨分校名義出現於ESI中，並無統一的Univ. Calif項目。如此一來，統計的基準點不同，若要直接進行排名比較易生問題。

權威控制的問題同樣會發生於合併後的機構，例如上海交通大學已與上海第二醫科大學合併，但在ESI中，兩校仍以各自的控制詞彙(Shanghai Jiao Tong Univ.及Shanghai Med Univ. 2)呈現，使用者無法了解合併後的上海交通大學之論文發表狀況及排名。

另外，機構名稱亦可能同時出現保留全稱與縮寫的情形，如大陸東南大學於ESI中分別有Southeast Univ.與Se Univ.兩種寫法，ESI同時保留兩種名稱並未加以合併。以上這些權威控制的特例確實存在於資料庫中，容易造成統計或比較各機構時不夠精確的問題，使用時須特別小心。

六、平均被引次數問題

在評估某對象的論文表現時，質與量是我們希望同時考量之指標，論文數可以表達文章發表的「量」的水準，而被引次數則可以揭示「質」的表現，但總被引次數可能因總論文數較多而因此提高，故ESI同時提供平均被引次數，希望排除文

資料來源：

1. 文中ESI收錄內容與指標定義參考自ESI Frequently Asked questions網頁<<http://thomsonscientific.com/support/faq/wok3new/ESI/#thresholds>> (95年11月)
2. 文中列舉之各項ESI數據實例皆直接檢索自ESI資料庫，收錄年份為1996年1月1日至2006年8月31日，檢索日期為2006年11月27日。

章數量而評估被引用表現。

ESI提供平均被引次數的原意是好的，但若單獨只看平均被引次數則非常危險，例如不分學門(All field)中，平均被引次數排名前五名的國家，分別為Bermuda、Switzerland、USA、Gambia、Panama，第一名的百慕達論文總數僅218篇，與第三名美國的2,831,004篇相去甚遠，令人無法將其視為研究表現較美國良好之國家。

此情況於機構中更為嚴重，不分學門中平均被引次數排名前三名：Hlth Res Network Apache Med Syst、Clin Nancy、Amer Acad Phys Assistants雖皆有千次以上的平均被引次數，但論文數卻都只有1篇；甚至不分學門的前20名中，前19名的論文數皆為4篇以下，第20名亦僅為11篇。故若欲於ESI中使用平均被引次數，筆者建議應設最低論文數的門檻，再針對通過標準之國家，進行平均被引次數的比較較為妥當。

結語

ESI所提供的各項數據，有助於使用者了解世界各領域的文章發表及引用狀況，快速比較不同對象(如機構、國家或研究員)間的文章發表情形、得知不同領域內的文章發表特性，及評估個人或單位於世界學研舞台之位置或差距。但其中仍存在許多問題，在利用其進行評量比較時，應特別注意ESI的各項特性，以避免不必要的爭議。

