

教育資料與圖書館學

*Journal of Educational Media & Library Sciences*

<http://joemls.tku.edu.tw>

---

Vol. 51 , 特刊 (2014) : 57-89

圖書館電子資源使用分析模式

之研究—以醫學電子資源為例

The Study of Analytical Model of

Library Electronic Resources Usage:

A Case of Medical Electronic Resources

游 忠 諺 Chung-Yen Yu

Ph.D. Candidate

謝 建 成\* Jiann-Cherng Shieh\*

Professor

E-mail : [jcshieh@ntnu.edu.tw](mailto:jcshieh@ntnu.edu.tw)

**[English Abstract & Summary see link](#)**

**[at the end of this article](#)**





# 圖書館電子資源使用分析模式 之研究—以醫學電子資源為例

游忠諺<sup>a</sup> 謝建成<sup>b\*</sup>

## 摘要

隨著網路興盛與蓬勃發展，電子資源的重要性與時俱進。實務上，圖書館所能獲取之COUNTER統計報表缺乏電子資源實際使用者與所屬單位等相關詳細資訊，因此無法針對使用者或實際使用單位作進階使用分析。為解決圖書館在電子資源使用分析時所遭遇之困難，本研究提出一電子資源使用分析模式，透過網際網路紀錄檔，以系統化方法轉換、清理、整合、分析巨量電子資源使用數據。為聚焦及簡化分析模式之探討，本研究以某醫學大學圖書館所訂購醫學電子資源為研究個案，依訂購年度、期刊、學院、特定學科領域，及特定文獻等常見需求項目進行統計分析。其結果除可作為圖書館館藏政策制訂與使用者行為分析參考，彙整資料亦可應用於資訊計量、資訊行為、書目探勘等研究領域，提供更多元議題之探討。

**關鍵詞：**電子資源使用分析，電子期刊，資訊計量，資訊行為，書目探勘

## 前 言

自1990年代起，電子期刊以其跨越傳統載體，不受時空限制，傳播快速等因素而興起，隨著網路興盛與蓬勃發展，電子期刊的重要性早已超越了紙本期刊，學術圖書館徵集期刊的方式，也不再完全以紙本期刊為主，轉而從紙本化朝向電子化，明顯改變學術出版、組織、保存的方法，學術圖書館的館藏發展政策也隨時代更迭持續轉變。但發展至今，圖書館的經費受限於學校或研究機構之總預算逐年減少，除紙本書籍外，電子資源的費用昂貴，圖書館因考量其成本效益，須透過有效且系統化的方式進行評估，以達成有效管理之目的。

<sup>a</sup>國立臺灣師範大學圖書資訊學研究所博士候選人

<sup>b</sup>國立臺灣師範大學圖書資訊學研究所教授

\* 通訊作者：jcshieh@ntnu.edu.tw

電子期刊隨著資訊科技的發達與網路的普及性增加，如今已成為圖書館服務的重要項目，但其經營模式與紙本期刊大不相同，電子期刊之訂購多以商業資料庫系統套裝方式進行販賣，且來源多元可供圖書館選擇，如出版社、期刊全文匯集資料庫廠商、書目供用中心等，其管理模式相較紙本期刊則複雜許多(註1)。而採以使用為主(use-centered)之評估時，從館藏評估觀點的角度，圖書館須著重在電子資源的利用情況，以及是否符合使用者需求程度之探討，其資料來源是根據各種館藏利用之統計資料加以分析，但電子資源具有多變性高且數量持續增長之特性，過去常用的分析方法，未能完全適用於電子資源，同時圖書館難以從代理電子資源的廠商取得統計資料與使用之原始數據，故在重疊的複雜環境中，如何發展其有效測量的方法，在進行電子資源之各項評估時，為圖書館所面臨的重要課題。

電子資源文獻使用(use)探討，可依據使用者的類型分為兩種觀點(註2)：(一)圖書館讀者，與(二)文獻作者等兩種身份觀點。從圖書館讀者身份的觀點，認為只要讀者透過圖書館網站於電子資源進行瀏覽、檢索、下載等行為時，即被認定為讀者「使用」該電子資源或文獻；由於圖書館網站會記錄讀者連線至電子資料庫伺服器後的各項行為，圖書館可依據伺服器網際網路使用紀錄檔，判斷某特定電子資源是否被瀏覽、檢索或點擊下載等。此項觀點是以量化數據為主要角度，僅依據網際網路使用紀錄檔是否儲存特定電子資源或文獻相關使用訊息，以判定電子資源是否被使用。但對於「讀者為何要使用該電子資源？」或「讀者檢索該篇文獻的動機與目的為何？」，甚至「如何得知讀者是否真實閱讀下載完成的文獻？」等，相關讀者資訊尋求行為等問題，並未重視。

而從文獻作者之觀點，亦為引文分析之角度思考，其意指為當文獻作者在其學術寫作時，參考他人之觀點或其他資訊來源，並在其文章內容或參考書目註釋中表達其引用之聲明，代表其所引用之文獻或其資料來源是有被「使用」的。事實上，無論一般讀者或文獻作者，其兩種角色是從不同情境下所賦予其意義，就情境而言，因寫作情境的產生，從讀者身份而進階篩選出文獻作者身份的觀點，讓文獻使用的範圍更加明確，此亦為書目計量學中引文分析的概念。

就圖書館管理層面而言，電子資源的使用是從讀者為出發點進行衡量，其常見延伸發展議題為：(一)館藏使用評估：透過流通研究、館內使用研究、使用者調查、館際互借分析、文獻傳遞測試等方式，評估館藏使用的狀況；(二)發展電子期刊評鑑，以利於決策支援、預算編制、策略規畫、品評鑑等管理議題；(三)使用者行為探索：從使用者的連線(Login)次數、下載次數、連線時間、參考諮詢次數等項目，針對使用者進行統計分析與行為動機方面之探索；(四)書目探勘：從使用者的電子資源下載紀錄為基礎，結合文獻書目資訊，依據圖書館

提供的各項服務為情境區隔，發掘各類型的資料樣態，以得知使用者的各項主題偏好等相關議題。

從上述另一觀點來探討文獻使用，常見的探討角度是從引文分析來看整個學術交流體系的變化，是從每年由索引摘要資料庫所推出的引文指標(Citation indicator)，來分析各種研究主題的趨勢變化，由此可看出學科領域中，其所收錄期刊的引用情況與變化，易言之，都是從文獻本身為出發點，探討期刊文獻如何被「使用」，並從作者與文獻角度進行研究與評量，但仍有其不足之處，原因在於索引摘要資料庫本身為商業公司所建置，如Thomson Reuters公司的Web of Science，Elsevier公司的Scopus，其所收錄的期刊雖達萬種，但未必圖書館所訂購電子資源皆有收錄，部分開放取用(Open Access)期刊或由出版社自行建置的電子期刊皆有可能未必收錄其內，若要得知其書目資訊則須從其他管道得知。

但從圖書館衡量電子資源是否續訂與停訂的角度，雖然是從使用統計報表的數據為考量依據，索引摘要資料庫僅為參考來源之一，但仍有其困難點：(一)主要因其資料來源仍來自於訂購的資料庫供應商，對於每種期刊的使用數據，每家資料庫供應商皆有提供各自的使用報表，但所提供的數據報告僅為機構本身購買的資料庫或期刊，在特定期間所成功下載的全文總數，需要圖書館連結至各廠商提供之使用平台進行資料匯出，再重新整理以得到統一的報表格式；(二)圖書館得到之統計數據，僅能得到該館對於該期刊的總下載量，並依據全文檔案的格式—網頁(HTML)或PDF兩種格式，按照下載年月進行統計，無法針對機構內的次級機構，進行明確的區隔與統計；(三)從資料庫廠商取得使用原始數據相當困難；(四)無法分析巨量的使用紀錄資料。

另外，學術圖書館為控管電子資源的使用，會自行建置代理伺服器進行使用者權限的驗證，但使用紀錄的資料量為動態增加，資料量之龐大與複雜，以少數主機進行統計分析的傳統思維，已無法因應日新月異之龐大使用數據，故至目前為止，針對電子資源使用之評估模式，未能發展出適當的評估方法與相關研究。

綜觀目前圖書館對於電子資源使用分析之困境，本研究為解決實務上所遭遇的問題，為電子資源使用提出一系統化之分析模式，並以醫學電子資源為研究個案進行模式驗證，說明模式建構的過程中，探討如何分析原始使用數據，建立書目及使用紀錄明細資料庫，以作為圖書館分析探討電子資源使用分析之參考。未來並可將使用資訊整合出如書目計量學、書目探勘與電子資源使用者行為等圖書資訊學相關研究的資料來源，提供更多的整合應用。

故本研究之研究問題為：(一)如何建構醫學電子資源下載使用分析模式？(二)如何分析網際網路紀錄檔之原始數據？(三)如何依據不同學院系所單位進行使用

統計？本研究之研究範圍以醫學電子資源為主要分析對象；實際使用資料來源是以某醫學大學圖書館提供之醫學電子期刊網路服務，並以自行建置代理伺服器之使用下載全文數據為依據，在研究限制方面，雖可從原始數據之使用者識別欄位與圖書館自動化系統讀者進行比對，並針對使用者進行點擊（含下載）次數之分析，然因個人資料保護法之法律規範，必須取得使用者之同意始能進行個案撰寫，故本研究個案設計以科系學院做為案例說明。

## 二、文獻探討

本研究試從(一)電子期刊評鑑，(二)電子資源評估，(三)E-Metrics評估電子資源使用狀況，(四)COUNTER計畫的報表規範等不同觀點項目，說明過去相關研究提出的觀點與實際上所遭遇的困難，以下分述之。

### (一) 電子期刊評鑑

電子期刊(E-Journals)，與傳統紙本期刊的首要區別，是從傳播載體進行區隔，從廣義上的解釋，是以連續性方式出版，並透過電子媒體發行的期刊，包括各種電子形式，如微縮、光碟、線上資料庫、網路資源等(註3)。電子期刊雖源自於紙本期刊，但無論傳播載體或傳播方式，兩者仍有其差異，對於館藏政策的發展，Mercer(註4)認為在新時代需要重新定義館藏的品質與價值，特別在電子資源方面，由於傳統的期刊評鑑方式無法完全反映電子期刊的使用全貌，且圖書館電子館藏的使用資訊並不充分揭露，圖書館需要因應館藏性質的變化，重新定義電子資源的價值。

從電子期刊評鑑的角度，在資料取得方面，圖書館與代理商網路伺服器記錄保存許多電子期刊使用資料，主要是了解使用者如何使用電子資源，但若從圖書館管理角度進行思考，可分為幾個層面：1.決策參考：幫助圖書館如何發展電子期刊館藏；2.預算編製：讓購置預算達到最佳使用；3.策略規畫：提供各項電子期刊未來推廣策略之參考；4.品質評鑑：從使用紀錄得知電子期刊與資料庫的質量優劣，協助做為評鑑之參考(註5)。就台灣的學術圖書館而言，王梅玲(註6)認為，學術圖書館極需要網路服務評鑑，特別在電子期刊層面，但相對而言，對於圖書館員要如何進行電子資料庫與電子期刊之評鑑，仍有其技術上的困難，需要有計量方法與工作協助之需求，藉以達到能採以微觀或鉅觀的不同角度，進階探討單一電子期刊的評鑑、使用者偏好等相關議題。

### (二) 電子資源評估

從使用評估角度分析圖書館的電子館藏，詹麗萍(註7)分析電子館藏的使用評估，對於使用或使用法(usage)，認為需要多方面的資訊及實際數據，並提出1.資料收集途徑，與2.以使用為主的分析方法等兩層面的論述。

在資料收集途徑方面，除了請廠商提資料庫使用統計數據外，圖書館自行收集的方法包括：1. 建立連結電子資源的網頁 (Intermediate HTML pages)；2. 在網頁伺服器使用特殊軟體 (pass-through scripts)；3. 透過線上目錄使用特殊軟體 (Using the library's online catalog with pass-through scripts)；4. 使用獨立的資料庫系統 (Database-driven access systems)；5. 使用代理伺服器 (Proxy servers)。

透過上述途徑，其主要目的在於1. 透過建立連結電子資源的網頁，透過點擊該網頁的次數，計算每一項連結電子資源的的點擊次數，即可估算使用者對於該電子資源的連線次數。2. 在網頁伺服器或圖書館的線上目錄使用特殊軟體，如點擊流 (clickstream) 側錄軟體，推知使用者在電子資源或瀏覽圖書館網站時，其網站之瀏覽歷程；3. 透過獨立的資料庫系統，當使用者透過其檢索機制，紀錄每次查詢之檢索詞，以得知使用者之檢索偏好；4. 使用代理主機，結合IP或自動化系統之身份驗證機制，驗證使用者是否為合法使用者，同時結合上述概念，紀錄使用者之電子資源瀏覽與下載歷程，進而分析使用者之使用行為，如瀏覽、下載、列印、電子郵寄全文等。

在評估方法層面，詹麗萍認為實務上圖書館在進行館藏評估時，以使用或使用者為主的角度，有六種常見的分析方法，分述如下(註8)：

**1. 引文分析 (citation analysis)：**引文分析法為書目計量學 (bibliometrics) 的研究範疇之一，主要利用各種數學及統計方法，將文獻的被引用率進行分析與探討，藉以探究文獻、作者、期刊等之間的引證關係。作者在發表其著作時，若引用某種期刊之文獻，而該期刊又恰為圖書館所訂購之館藏，透過引文分析，圖書館可獲取使用者對其資料使用的線索，進而可協助圖書館員在館藏發展與服務策略層面的政策制定與調整，以滿足使用者的資訊需求。

**2. 流通研究 (circulation studies)：**分析館內各項流通資料，依不同使用者族群、地點、出版日期、主題分析、使用者在自動化系統所紀錄之交易類型(如：借閱、續借、催還、預約)等加以分析，其結果可用於改善館藏發展實務與修正資源分配，但該方法僅能觀察流通資料之使用，而禁止流通的館藏則無法進行分析統計，且未能提供讀者借閱失敗或館藏無法滿足讀者需求之相關資訊。

**3. 館內使用研究 (in-house use studies)：**館內使用研究可透過上架統計、閱讀調查表單、觀察、訪談或問卷調查等研究方法搜集相關資訊。此法常用於非流通性資料的使用研究，結合前述流通研究方法，可提供館藏使用更正確的資訊，但此法受限於使用者能否充分合作與願意表達意見，否則影響調查結果之正確性。

**4. 使用者調查 (user surveys)：**使用者調查為常見圖書館在取得使用者意

見反映的一種方式，透過口頭訪談、電話訪談、電子郵寄問卷、網頁問卷或傳統的書面問卷實施調查。以其所獲得的資訊作為評定館藏是否有效滿足讀者需求的依據，藉此幫助圖書館得知讀者興趣和改變的趨勢。

5. 館際互借分析 (interlibrary loan analysis)：館際互借紀錄可反映館藏的使用狀況，其意義在於表示當讀者在查詢館藏後發現該館缺乏所需要的資料，可透過館際互借取得所需要的資料。透過館際互借資料的分析，可確認圖書館本身館藏領域之不足，藉以提供修訂館藏發展政策的參考。

6. 文獻傳遞測試 (document delivery test)：用於圖書館在讀者需要資料時能否即時供應的能力，最常使用的方式是找到一份最能反映使用者資訊需求的清單。先檢視館藏目錄，若顯示無館藏時，則另外搜尋其他管道能否檢索到這些資料；其次則是測試取得所需資料須耗時多久。文獻傳遞測試主要是提供客觀衡量館藏能否滿足讀者需求的方法。

上述六種評估方式對於圖書館的館藏評估與使用模式，皆以紙本館藏的思考為出發點。詹麗萍認為，電子館藏評估雖重要，但遲遲未能發展出適當的評估方法，傳統的評估方法未必適用於電子資源，因過去的評估方法是以紙本為基礎發展，電子資源並非實體有形之物質，故其認為電子資源無法以出借流通、館際互借或文獻傳遞服務等方式取得，另也未予以分類編目、也無須排架，故傳統的流通研究、館際互借分析、文獻傳遞測試、館藏直接分析等方法均不適用於電子資源評估。

從借閱流通角度而言，紙本期刊尚可採用借閱次數做為是否使用之依據，但電子資源並非用傳統借閱的方式進行衡量，同一期刊同時可供多讀者下載閱讀，使用次數並非限制於圖書館的開館時間與區域，任何具有合法身份能取用 (access) 圖書館電子資源的使用者，透過網路並通過身份驗證，即可下載期刊文獻，對於「使用」狀態的產生，並不需透過館員即可進行，因此電子資源的使用是任何時間都在發生的狀態，必須透過系統性的方法進行分析，始能將複雜繁瑣的原始使用資料轉換為可加值利用的資訊。

### (三) E-Metrics 評估電子資源使用狀況

E-Metrics 最初應用於電子商務，為網路企業所用的網站績效評量法，被認為是以科學方法檢驗網站通行數據的一種工具。其定義為：「測量網站成功程度的計量法」，其目的在了解網路使用者行為，確定網站是否吸引適當的使用者來訪、評定網站經營策略是否得當。其目的包括：1. 了解網站使用者的行為；2. 確定網站是否吸引適當的使用者來訪；3. 評定網站經營策略是否得當 (註9)。

E-Metrics 應用於圖書館，依據圖書館服務與館藏評估兩個層面，再細分四個項目，每一個項目都有各自所屬的元素，分述如下 (註10)：

1. 以使用者為中心 (user-centered)：以網站使用者做為評估重點，如包括使用者區別惟一性 (unique visitor)、何時離開網站 (turnaway)、連線時間區段 (session) 與瀏覽頁面 (visit) 等元素。

2. 以服務為中心 (service-centered)：以網站頁面的服務點擊次數做為衡量重點，如頁面的檢索、觀看 (view)、下載與列印等提供功能之執行次數。

3. 電子資源 (E-resources)：以資源類型為統計重點，如電子期刊、電子資料庫、電子書與數位檔案 (digital document) 等資源類型的數量。

4. 電子服務 (E-services)：以提供之網站功能為重點，如圖書館之檢索目錄 (WebOPAC)、數位參考服務 (digital reference)、整合查詢 (MetaSearch) 與分散式協定—開放式定位址 (OpenURL) 等元素。

E-metric 從上述不同類型進行評量，透過量化或質化等不同的資料收集與分析方式，藉以得知圖書館服務與館藏的現況，並結合圖書館服務與館藏評估，做為一個聯結平台的分析方式。

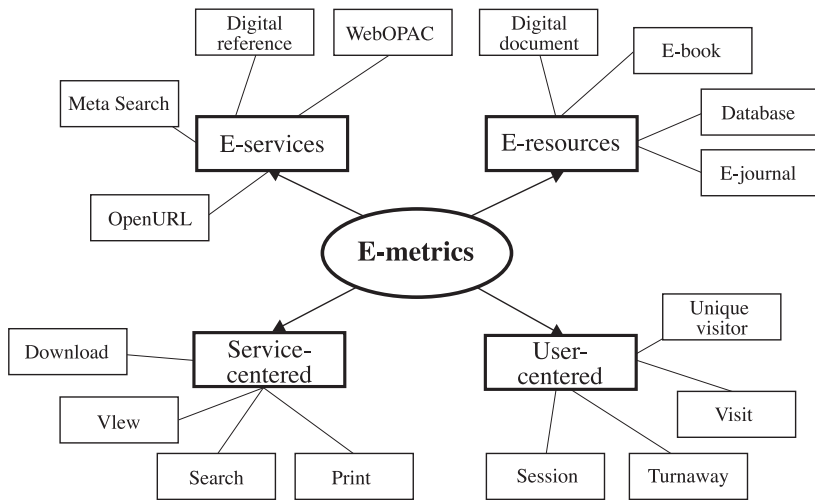


圖1 E-Metrics應用於圖書館評估面向

資料來源：Andrew White and Eric Djiva Kamal, *E-Metrics for Library and Information Professionals* (New York: Neal-Schuman, 2006), 6.

雖然電子期刊是 E-metrics 的評量項目之一，但因電子資源的使用不限定任何時間點，在即刻發生的狀態下，所產生的交易紀錄 (transaction log file)，遠比圖書館自動化系統儲存的館藏資料與書本借閱紀錄量更為龐大，過去採取人工分析做法，僅能分析小部分的資料集，對於資料集龐大的使用紀錄根本無法透過傳統人工分析進行分析，且在合理時間內達到擷取、管理、處理，並整理成為人類所能解讀的資訊。事實上，圖書館在進行電子資源使用分析時，勢必處理使用紀錄等原始資料，透過系統化的分析模式，將原始資料進行清理、整理與整合，始可拆解成各種不同的獨立資料集，可得出許多額外的資訊和資料



關聯性，藉以可察覺讀者、單位機構內的更多潛在資訊與使用趨勢。

另外，圖書館員需要分析使用數據的原因：1. 衡量不同的電子資源與其服務價值；2. 協助在訂購期刊時的決策；3. 規畫圖書館資源的基礎建設與分配模式；4. 支援圖書館行銷與服務。另一方面，對於電子資源廠商，使用數據的意義在於：1. 電子期刊與傳統紙本期刊不同，紙本期刊透過複製的份數決定計價，而電子期刊則是以下載次數反映電子期刊的現有價值，因而廠商藉此嘗試訂定新的計價模式；2. 評估透過不同資訊來源來得知該電子資源在市場上的相對重要性；3. 為編輯者提供開發新產品的系統需求與評估依據；4. 改善網站的使用設計與導覽功能，增加收集使用數據的能力，進而能在市場分析時，得到最佳的推論與使用分析（註11）。

因此，電子資源統計報表為圖書館與出版商所採用的一種評量方式，其功能對於圖書館管理方面，可以協助館員得知圖書館所訂購電子資源在訂購年度之讀者點擊（含下載）狀況、評估電子期刊對讀者的價值、比較各出版商的產品及文獻傳遞方式，藉報表提供之資訊，淘汰不合時宜之電子資源及制定長期訂購計畫等（註12）。也因電子資源使用之統計數據對於圖書館在制定館藏發展的重要性極高，圖書館如何制定採用標準化的方式，持續蒐集與分析數據，乃屬於長期計畫的範疇，故圖書館陸續提出多項評估電子資源的研究計畫，以E-Metrics的評估計量方法應用於電子資源的評估，並積極從事研究開發，其中如COUNTER（Counting Online Usage of Networked Electronic Resources）（註13）、ARL E-Metrics Project（註14）計畫等，另外，以COUNTER所制定出的報表格式為目前實務上最常使用之類型。

#### （四）COUNTER 計畫的報表規範

COUNTER 計畫規範出版商提供各類型的統計報表，在其計畫進行過程中，自2003年公布第一版的統計報表後，分別於2005年、2008年、2012年公布新的報告類型，截至2014年7月止，最新版本為2012年4月公布的第四版（註15），報告以資源類型為區隔，分為期刊（Journal）、資料庫（Database）、圖書（Book）、多媒體（Multimedia）、標題（Title）五類，另2014年3月，COUNTER另公布了文獻（Article）的統計報表類型（註16），而對於功能層面，則分為「強制要求出版商提供」與「非強制要求提供」兩層級，其資料來源是源自電子資料庫伺服器的網際網路使用紀錄（transaction logs），其紀錄根據COUNTER的標準轉化成可利用且易讀的資訊，產生適合圖書館解讀及利用的電子資源使用報表。其報表類型參見表1。

從上述COUNTER的電子資源統計報表格式得知，主要是以期刊、資料庫、書籍、多媒體內容、文獻等做為分類依據，再配合時間、平台、載具等其他變數進行統計分析。對於出版商而言，透過交易紀錄，足以提供學術單位

圖書館之統計數據，且逐年公布報表類型，從期刊、資料庫、書籍，進步至行動載具、多媒體與文獻為統計區隔單位的多種類型報表。但對於各圖書館而言仍有其不足，圖書館在制訂續訂政策時，常需要多種類型的數據以茲佐證。COUNTER 提供了不同類別的統計報表，但對於學術機構編制內下層單位之細項統計，卻囿於紀錄檔之欄位限制而無法進行分析，亦為其在報表格式制定時的先天限制。在實務上，亦有廠商為解決圖書館員須自行下載各不同出版社或電子資料庫的使用數據，結合 SUSHI 協定（註 17）開發整合平台，提供報表

表 1 COUNTER 公布報表類型－期刊、資料庫、圖書、多媒體、標題、文獻

名稱	意義
Journal Report 1 (JR1)	依據月份及期刊名稱統計單本期刊內文章全文的有效需求數 (successful request)
Journal Report 1 GOA	依據月份及期刊名稱統計 Open Access 單本期刊內文章全文的有效需求數
Journal Report 2 (JR2)	依月份及期刊名稱統計單本期刊被拒絕使用(turnaways)的總次數
Journal Report 3 (JR3)	依月份、期刊名稱與內文格式(Page-Type)統計單本期刊內全文的有效需求數與被拒絕使用的次數
Journal Report 3 Mobile	依月份、期刊名稱與內文格式(Page-Type)統計單本期刊內，透過行動裝置(Mobile)取得全文的有效需求數與被拒絕使用的次數
Journal Report 4 (JR4)	依月份及服務統計被搜尋的總次數
Journal Report 5 (JR5)	依據出版年(Year-of-Publication, YOP)與期刊名稱統計單本期刊內文章全文的有效需求數
Database Report 1 (DB1)	依月份及資料庫名稱統計資料庫被檢索、檢索結果的點擊與紀錄閱覽的次數
Database Report 2 (DB2)	依月份、資料庫名稱、類別統計資料庫被拒絕使用的總次數
Platform Report 1	依月份、資料庫名稱、平台統計資料庫平台被檢索、檢索結果的點擊與紀錄閱覽的次數
Books Report 1 (BR1)	依月份及題名統計單本圖書的有效需求數
Books Report 2 (BR2)	依月份及題名統計單本圖書中章節的有效需求數
Books Report 3 (BR3)	依月份、題名與類別統計單本圖書被拒絕使用的次數
Books Report 4 (BR4)	依月份、平台與類別統計單本圖書被拒絕使用的次數
Books Report 5 (BR5)	依月份及標題統計被搜尋的總次數
Multimedia Report 1	依月份及多媒體內容集合統計多媒體的有效需求數
Multimedia Report 2	依月份、多媒體內容集合與項目類型統計其有效需求數
Title Report 1	依月份與標題統計全文與書本章節的有效需求數
Title Report 1 Mobile	依月份與標題統計在行動裝置內全文與書本章節的有效需求數
Title Report 2	依月份、標題與類別統計被拒絕全文使用的次數
Title Report 3	依月份、標題與內文格式統計標題的有效需求數
Title Report 3 Mobile	依月份、標題與內文格式統計在行動裝置取得的有效需求數
Article Report 1	依文獻標題與出版社收集與統計使用資訊
Article Report 2	依文獻作者、月份與 DOI，統計不同來源的有效需求數
Article Report 3	依文獻作者，統計個別文獻有效取得次數的報告綜整

資料來源：COUNTER, "The COUNTER Code of Practice for e-Resources: Release 4", COUNTER, April 2012, accessed May 28, 2014, <http://www.projectcounter.org/r4/COPR4.pdf>;  
COUNTER, "The COUNTER Code of Practice for Articles: Release 1", COUNTER, March 2014, accessed May 28, 2014, [http://www.projectcounter.org/documents/counterart\\_cop\\_MAR2014.pdf](http://www.projectcounter.org/documents/counterart_cop_MAR2014.pdf).

整合與資料匯出等服務，雖可自行設定電子資源與學院科系之間的對應關係，但仍無法完整解決上述圖書館內部單位細項統計分析的問題。

總而言之，從圖書館管理的立場，能夠透過圖書館自動化系統的使用數據，進行使用評估，但取得的數據皆為圖書館借閱與流通的部分，藉以支援館內流通、館際互借、使用者調查、館內使用研究等項目，但目前圖書館提供的服務，電子資源的需求性已日漸增加，過去讀者到圖書館內尋求館員面對面服務的需求已大為減少，相形之下，透過圖書館網站提供的服務，即可完成其資訊需求，但時至今日，較少分析圖書館網站服務的相關研究。整合過去的分析方法，過去詹麗萍提出以使用為主的館藏評估方法（參見圖2），未提出分析圖書館服務網站研究為主的評估方式，而COUNTER提出的電子資源使用報表，也無法以機構或學校內下層單位與使用者做為評估單位，故透過圖書館自行建置代理伺服器的方式，雖可補其不足，但資料量龐大，且須專人進行長期的資料收集、處理、分析與解釋，對於圖書館本身，則又受限於人力資源之困窘，故需要一系統化的分析方法以解決人力處理上的困難。

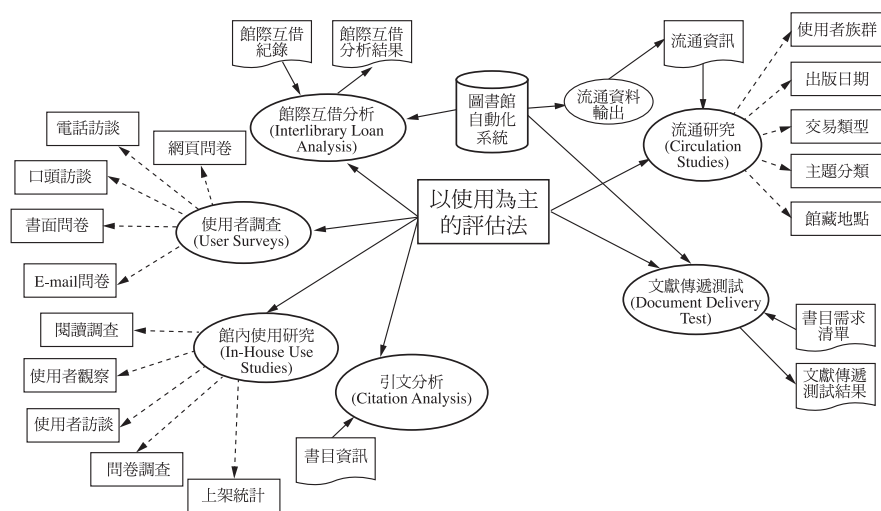


圖2 圖書館資源常見評估方法

資料來源：本研究自行繪製，其內容整理自詹麗萍，電子資源與圖書館館藏發展（台北市：麥田，2005），170-74。

此外，由於現今網路資源為圖書館提供服務的重要來源，故圖書館網站的使用相關研究也相形重要，特別在電子資源使用層面，是過去較少提出分析方式的一環，因而本研究希冀提出一系統化的分析方法，將原始數據進行加值處理，藉以提供更多的整合應用研究。

### 三、研究方法設計

為解決以往電子資源使用分析的問題，本研究提出一電子資源使用分析模式，首先必須自行收集使用數據，故須自行建立代理伺服器，其後資料處理與分析的過程，分為四個階段，分別為實體資料來源、資料預處理、資料呈現與應用領域，最後為易於保存與分析數據，須自建使用明細資料庫儲存處理完成的資料，以利於透過自動化方式在最短時間能獲取所需的統計數據。以下針對本研究之資料來源、分析模式內容、系統建置與資料收集和資料處理等過程進行描述。

#### (一) 使用記錄檔資料來源

本研究採用之記錄檔為某醫學大學圖書館電子資源使用之代理伺服器使用紀錄，在說明個案範例，其時間範圍涵蓋自2010年10月至2014年4月。

#### (二) 模式說明

本模式分為四個概念：1. 實體資料，2. 資料預處理，3. 資料呈現，4. 應用領域，以下分述之，請參見圖3。

**1. 實體資料：**在使用資料紀錄分為兩個來源，即代理伺服器網際網路使用紀錄檔與圖書館自動化系統。前者是使用者在下載電子資源的行為紀錄，而後者是圖書館所紀錄的所有館藏、讀者資訊、借閱紀錄等資料。在本研究中，圖書館自動化系統可提供讀者資訊，以利於在分析使用資訊時，建立各使用者的區隔。

**2. 資料預處理：**本概念主要結合上述兩資料來源進行資料的預處理與整合，從前一階段處理完成並依目的之不同。其資料處理之過程與步驟較為繁複，區分為：(1)原始資料轉換；(2)文獻書目資料取得；(3)使用者識別與資料取得；(4)處理資料匯出。

**3. 資料呈現：**使用資料在經由資料清理、預處理等過程後，可依據不同目的，而篩選出所需要資料，分為五種：(1)文獻原始書目；(2)文獻參考書目；(3)文獻主題詞；(4)使用者與其下載文獻明細資料；(5)使用者下載電子資源使用統計。本研究並建構一文獻書目暨使用明細資料，為不同領域需求，提供使用數據或書目文獻，以做為資料分析之用。

**4. 應用領域：**本研究除建構醫學電子資源使用模式外，其所處理後的資料，可應用於書目計量學、書目探勘與電子資源使用者行為之用。

此四項概念的重點在於，如何透過系統化的分析方式，清理與整理使用數據資料，匯出成可再進階利用的分析資料，為不同的應用領域提供使用明細與相關書目資訊，為圖書資訊相關研究領域提供新的資料來源，後述說明實際資料預處理與整理之過程。

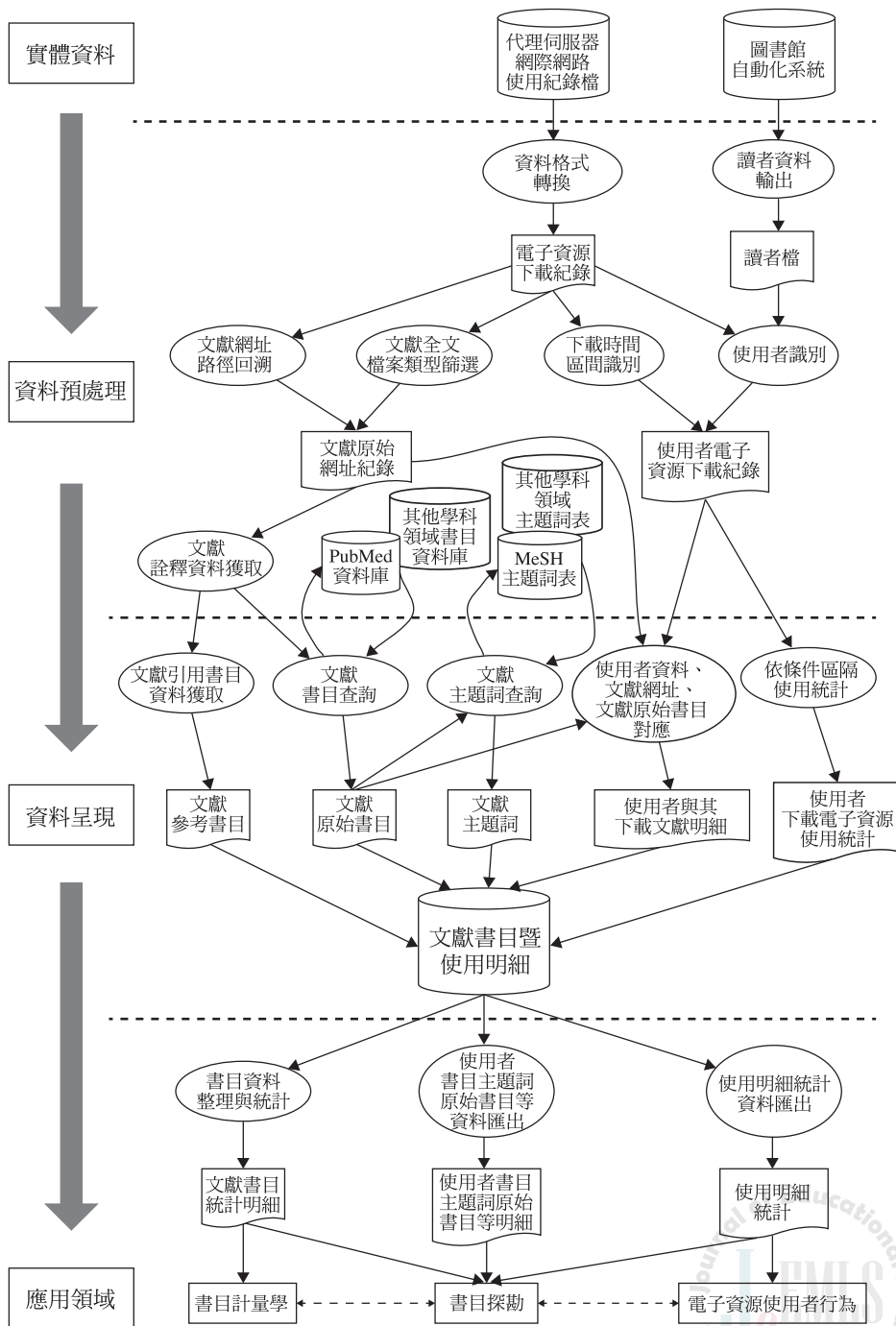


圖3 電子資源使用分析模式

資料來源：本研究自行整理

### (三) 系統建置與資料收集

圖書館提供電子資源服務，如何判斷是否為合法使用者，始能允許文獻的取用，實務上的做法，一般多採取使用1.讀者身份，2.讀者來源IP位址等兩種驗證方式。前者主要目的是識別是否為該單位機關具合法存取權限之使用者；後者目的是在進行文獻取用時，廠商會根據來源的IP位置識別是否為合法單位(註18)，判定是否可取用該文獻。但電子資料庫廠商並無法得知取用文獻之原始使用者為何，僅能得知是否為具取用權限之單位，故在電子資源使用分析方面，廠商提供的統計報表，僅能得知全體使用統計數據，圖書館無法針對機關之下屬單位進行細項統計分析。為達成進行細項使用者統計分析的目的，圖書館須自行建置一電子資源代理伺服器，其主要目的有二：1.判定機構的合法使用者與所屬權限；2.與電子資料庫廠商伺服器進行網路介接。換言之，圖書館建置之代理伺服器為一中介窗口，提供身分驗證與取用文獻之服務，透過此方式收集單位內的使用紀錄，進而與原有的圖書館自動化系統結合，獲知使用者身份與連線來源位置，始能進行細項的資料處理與分析等後續階段。

實務上，圖書館在建置代理伺服器，提供給使用者有數種連線方式：1.使用者須在連線瀏覽器進行額外的參數設定，將代理伺服器的IP位置填入，經身份驗證後，才能正確的連線與取用所需文獻；2.使用者不須事先填入代理伺服器的IP位置，但須連線至圖書館提供的電子資源管理系統進行身份驗證，再連線至所需文獻的網站；3.透過虛擬個人網路(Virtual Private Network, VPN)連線方式。

另外，在收集使用紀錄之後，由於代理伺服器提供的紀錄檔，其格式未必統一，在進行系統分析與判讀前，須得知該伺服器軟體採用的紀錄檔格式，始能依據紀錄檔的欄位樣式，讀取所需資料進行分析。

### (四) 資料處理階段

從資料探勘的觀點，取得原始資料後，須進行資料預處理(data preprocessing)的動作，包括資料清理(data cleaning)、資料轉換(data transformation)、資料縮減(data reduction)、資料整合(data integration)等過程(註19)。本研究經下列數個步驟進行處理：

**1. 資料清理：**資料清理主要是清除原始資料的雜訊(noise)，因任何收集到的原始資料，易有如遺失值、雜訊與資料不一致等情況。本研究在資料清理階段，由於網際網路紀錄檔的記錄格式，特別在網址紀錄的部分，資料庫廠商或期刊出版社，除了文獻網址外，可能會插入特殊識別碼，讓網址資料的長度無法完全記錄，容易造成網址欄位之紀錄不完整，故在資料清理階段，須清除無法識別的資料。另因本研究主要以期刊全文檔案，且以PDF檔案格式為區隔方

式，故在網際網路紀錄檔中若非上述 PDF 檔案格式之全文檔案紀錄，則在此階段皆一併處理。

**2. 資料整合：**本研究因處理跨年度之紀錄資料，必須將各年度與不同月份的原始紀錄資料進行整合，本步驟主要將不同時期的紀錄檔，整合成一個跨年度與月份的原始紀錄資料檔。

**3. 資料轉換：**取得代理伺服器網際網路使用紀錄檔，至轉換為電子資源下載紀錄，為求在資料處理的便利，須將原始紀錄進行資料格式的轉換，其理由為：(1)紀錄檔若非在紀錄時間的格式，多採用毫秒 (microsecond) 的格式儲存，利於運算處理但不利於人類的辨讀。(2)當紀錄檔資料量發展至海量資料的規模時，以檔案做為儲存方式，會在資料檢索與運算的效能不彰。

在資料格式轉換過程中，分為五個動作，參見圖 4：

**1. 轉換時間戳記：**代理伺服器在時間戳記 (Timestamp) 欄位採取機讀格式進行時間欄位的註記，其註入方式採取 Unix 紀元 (UTC 1970-01-01 00:00:00) 以來的秒數，並採取毫秒方式輸出，在這階段須將毫秒的儲存型式轉換為人類能判讀的格式，本研究採用紀錄檔分析報表軟體 -SARG (Squid Analysis Report Generator) (註20)，該軟體提供格式轉換功能，易於在處理原始資料時間戳記。

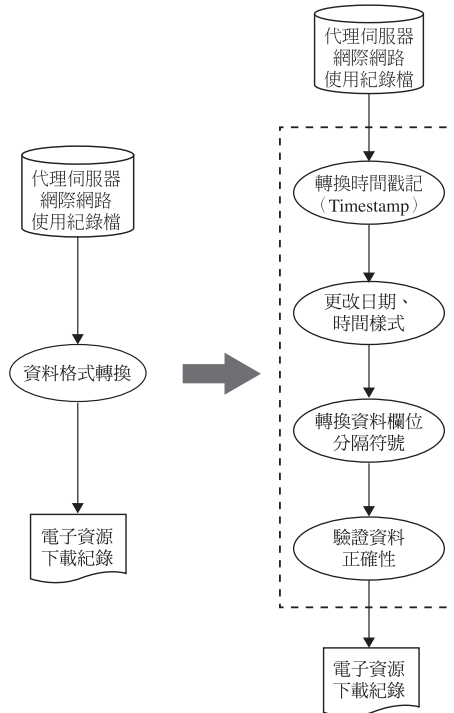


圖4 網際網路使用紀錄檔資料格式轉換流程

資料來源：本研究自行整理



**2. 更改時間、日期樣式：**第一次轉換時會以歐美常用的「MM/DD/YYYY」輸出，為求與資料庫管理系統中，時間欄位的樣式一致，再進行二次轉換，以「YYYY-MM-DD」方式輸出，在本階段採用Unix系統常見的編輯器sed (stream editor)，將紀錄檔的時間格式透過正則表達式方式輸出。

**3. 轉換資料欄位分隔符號：**將原始的欄位分隔符號，轉換為以定位點 (tab-delimited) 分隔，降低在資料處理時判讀的錯誤機率。本階段採用Unix系統上的樣式掃描與處理語言awk，進行紀錄中的欄位分隔符號轉換工作。

**4. 確認資料轉換前後的正確性：**若在不同的作業系統進行格式轉換的行為，如Mac/Linux/Unix/Win，其轉換過程中會因判斷字元編碼的因素而導致失敗。本研究取得的紀錄檔之字元編碼，為求利於跨平台時的處理，字元編碼皆統一以utf-8格式進行儲存，在轉換後再進行資料筆數的統計與樣式的驗證，以確保資料的完整性。

**5. 輸出電子資源下載紀錄：**將驗證正確後的紀錄輸出至檔案，以做為資料預處理階段中主要的資料處理來源。

由上述資料格式轉換流程中，從原始難以判讀的紀錄檔，整理並匯出為電子資源下載紀錄，下一階段為求(1)使用者身份，(2)原始文獻相關資訊等兩項目的，須分別進行不同的處理流程，並衍生其他相關檔案，以下分述之。

(1)使用者身份方面：主要從圖書館自動化系統的讀者檔取得讀者之相關資訊，在使用紀錄檔中，必須著錄能足以識別讀者身份之資料，始能透過與讀者檔進行比對。實務上，圖書館自動化系統習慣以(i)讀者之學號或職員編號，(ii)個人身份證或居留證號碼做為惟一識別之依據，再著錄其他相關資訊，如科系、所屬單位或部門，若遇多重身份之讀者，則須經由其他欄位之輔助加以識別。

(2)原始文獻網址紀錄：每一出版社或電子資料庫廠商之網址樣式(URL pattern)皆有不同，有的為期刊縮寫，有的為ISSN，有的由電子資料庫廠商或出版社自行編碼，如同讀者身份，在資料清理時必須事先分析區隔原始文獻網址之樣式，並清理其他妨礙辨識文獻網址之雜訊，始能著錄至使用明細資料庫時，增加檢索效率。

(3)使用者電子資源點擊(含下載)紀錄：在代理伺服器之網際網路使用紀錄檔案內，除了記錄上述兩者，另須如來源IP、點擊時間等其他欄位，以資識別與分析。

**6. 資料呈現：**依據需求目的之不同，產出各類型的資料檔，如文獻原始書目檔、文獻參考書目檔、文獻主題詞檔、使用者與下載文獻明細檔、使用者下載電子資源使用統計等，以下分述之。

(1)文獻原始書目：在回溯文獻原始書目的過程，其資料來源有二：(i)電子



資料庫或出版社；(ii)PubMed/MEDLINE資料庫。由於從下載紀錄中回溯的網址，為電子資料庫或出版社為該文獻所著錄之書目內容，因而可做為文獻基本書目資訊的來源，待確認回溯網址至文獻後，其書目資料輸出格式有二：(i)網頁內嵌的詮釋資料 (Metadata)；(ii)提供的引文書目檔案 (citation file)，如以Nature 資料庫為例，Nature 提供書目管理軟體Reference Manager、Endnote 相容的RIS 檔案，但並未針對文獻本身進行主題分類，為求書目資料之完整性，則將文獻標題與DOI等可供識別之欄位資訊，透過PubMed/MEDLINE 資料庫進行查詢，取得該文獻之書目資訊，並以MEDLINE 格式輸出成純文字檔儲存，電子資料庫或出版社與PubMed/MEDLINE 資料庫提供之書目資訊內容最大不同處，即為PubMed/MEDLINE 另提供醫學主題詞 (MeSH)，作為主題分類與檢索用途。

(2)文獻參考書目：透過前述取得文獻原始書目等資訊後，可藉此作為取得該文獻之參考書目，其來源有二：(i)電子資料庫或出版社提供之參考文獻資訊檔案 (reference files)；(ii) Thomson Reuters 公司之Web of Science (WOS) 資料庫，提供之參考文獻清單檔案。兩者皆提供利於書目管理軟體Endnote 之匯入格式，或採用Tab 分隔之純文字檔格式，透過文獻參考書目的匯總，可進行書目計量學的常見研究型式，如引文分析、文獻老化、作者生產力等相關研究。

(3)文獻主題詞：PubMed/MEDLINE 資料庫將所收錄之文獻，皆會從文獻之標題、摘要等內容，給予適當的醫學主題詞。而醫學主題詞的編撰皆由相關領域之專家學者擔任，每年皆發布新的版本。透過醫學主題詞，研究者若對特定醫學領域有其偏好，可透過主題詞檢索相關之參考文獻。若結合圖書館電子資源之使用者分析，從微觀角度，可探索使用者下載文獻之偏好；若從宏觀角度，可從科系、學院、單位、團體為區隔，分析不同單位等對於某特定醫學領域之下載偏好；或從期刊為出發點，得知不同使用者或科系、單位對該期刊之下載狀況；對圖書館管理階層，可做為館藏評鑑之參考，或為未來是否續訂評估之有效證據。本研究從PubMed/MEDLINE 資料庫提供的醫學主題表，自行建置MeSH Tree 資料庫，根據每一文獻中的主題詞，剖析每一主題詞的主題路徑，並建構文獻主題詞對應資料庫。

(4)使用者與下載文獻明細：從圖書館本身建置之伺服器取得使用紀錄，從紀錄檔之原始資料，配合圖書館自動化系統，從讀者檔可解析出紀錄檔中使用者代碼的意義。另從下載網址之回溯，可獲取前述的文獻相關資訊，將兩者資料結合分析，即可針對使用者之下載進行分析，但仍須注意個人隱私與相關法律規範。

(5)使用者下載電子資源使用統計：以往圖書館在進行紙本期刊的使用統計，僅能從上架統計、閱讀調查表單、觀察、訪談或問卷調查等方法搜集資訊，對於電子期刊的使用統計，從資料庫廠商或出版社提供之期刊使用數據平台下載使用數據報告，其數據來自讀者透過圖書館之電子資源管理系統或校園網路等途徑，連結至電子資料庫進行文獻的檢索與下載等行為，最後成功下載文獻之統計數據，但其限制是無法區隔機關單位內的子機構與實際使用者，其報表僅顯能顯示統計數字，若要採取更細微的統計模式，如以文獻為單元，分析由那些使用者所下載的統計報表，就無法進行分析。

7. 資料範例說明：以代理伺服器網際網路使用紀錄為例，如下所列以說明資料轉換過程。

```
1392966826.714 991 60.251.55.209 TCP_MISS/206 33250 GET
http://onlinelibrary.wiley.com/store/10.1002/jbm.a.34466/asset/34466_ftp.pdf?v
=1&t=hrx4h5vc&s=aab0ec248a3a488ab3231e4fd2bcc1f4e77f2b39 A123456789
DIRECT/onlinelibrary.wiley.com application/pdf
```

(1)在時間戳記部分，「1392966826.714」為 Unix 系統的時間戳記，在經過轉換之後，可得知為「2014/2/21 15:13:46」。

(2)檔案類型為 PDF，其識別方式為「application/pdf」

(3)在下載網址部分，原始資料為

```
「http://onlinelibrary.wiley.com/store/10.1002/jbm.a.34466/asset/34466_ftp.pdf?
v=1&t=hrx4h5vc&s=aab0ec248a3a488ab3231e4fd2bcc1f4e77f2b39」
```

在進行資料判讀時，可得知網址中包含了期刊的 DOI，在經由 DOI 網址轉換後即可得知為「*Journal of Biomedical Materials Research Part A*」。

(4)進一步從網址中的 DOI，另可對應至文獻篇名

```
「Efficacy of mesoporous silica nanoparticles in delivering BMP-2 plasmid
DNA for in vitro osteogenic stimulation of mesenchymal stem cells」。
```

(5)紀錄檔中另著記使用者識別碼，因個資保護因素，雖可與圖書館原有自動化系統讀者檔進行比對，但僅可針對於未能足以識別個人資訊之項目進行統計分析，如學院所單位等。

(6)依據上述轉換後之資料再匯入自行建置之資料庫。

綜上所述，本研究可依據圖書館自動化系統的讀者檔與網際網路使用紀錄檔進行資料比對與區隔，以找出使用者與文獻之間的關聯性，進而在圖書館進行使用者分析時，即能具備更多的區隔變數，如從點擊（含下載）時間區間、使用者本身，與其所屬學院、科系與文獻等變數，藉以產生更多的加值應用分析。

## 四、研究結果

本研究採用某醫學大學圖書館代理伺服器，篩選Wiley電子資料庫使用數據為研究個案，原始紀錄檔案為1.04GB，共4,181,850筆紀錄，記錄時間由2009年8月17日至2014年4月21日，剔除無法完整辨識之紀錄，並以2013年之期刊清單進行條件篩選，在未進一步分析文獻書目之詳細資料，共有1,216,156篇全文PDF點擊(含下載)次數。

本研究結果說明共有六項，分別為(一)電子資料庫歷年點擊(含下載)次數分析；(二)電子資料庫實際使用分析；(三)年度點擊(含下載)電子資料庫次數分析；(四)跨年度點擊(含下載)期刊之分析—以醫學院為例；(五)單一年度期刊與學科領域科系之點擊(含下載)分析—以公共衛生相關科系為例；(六)特定期刊文獻之點擊(含下載)分析。前兩項之統計分析亦可從COUNTER統計報表得知，而後四項則為COUNTER統計報表所無法提供。以下分別說明之。

### (一)電子資料庫歷年點擊(含下載)次數分析(2010年8月至2014年4月)：

從歷年點擊(含下載)次數分析，2011年至2013年之點擊(含下載)次數皆有超過30萬次。若以各年度為區隔，以2012年總點擊(含下載)次數為最高，有404,111篇文獻點擊(含下載)，其次為2013年有349,358篇文獻，與2011年的302,760篇文獻點擊(含下載)次數；2014年因資料收錄時間，自1月1日至4月21日為止，其平均每月下載次數為29,106次，與前一年度2013年之平均每月下載次數數據相異不大，故平均每月下載次數除2010年偏低外，每月點擊(含下載)次數均超過2萬5千次以上，顯示Wiley資料庫每日的文獻點擊(含下載)量具有近千篇文獻次數以上的規模。

表2 歷年文獻點擊下載次數

年份	點擊下載次數	月均下載次數
2010	43,496	8,699
2011	302,760	25,230
2012	404,111	33,676
2013	349,358	29,113
2014	116,431	29,106

說明：2010年與2014年點擊下載紀錄分別為：

1. 2010年8月9日至12月31日；
2. 2014年1月1日至4月21日。

資料來源：本研究自行整理

進一步探析2010年之點擊下載次數與平均每月下載次數較少的原因，其可能性研判如下：1.因Wiley資料庫變更電子文獻下載網址，在經由網址樣式判斷過程與去重覆之驗證流程後，資料筆數減少；2.期刊需求之差異，因採取2013年的期刊清單做為篩選依據，另收錄時間為2010年8月9日至12月31日止，故

點擊下載次數較其他年度為少。

## (二) 電子資料庫實際使用分析

以訂購 Wiley 電子資料庫 2013 年的期刊清單所載期刊為基準，全數 2,331 種期刊中，僅有 833 種期刊是在 2010-2014 年間至少點擊下載 1 次，而其他 1,498 種期刊，在橫跨 5 個年度的時間區間，則未有任何點擊下載紀錄，亦即只有 35.74% 的期刊有其讀者需求，而多達 64.26% 的期刊則乏人問津，顯示在訂購期刊的方式，採取包裹方式訂購電子資料庫，應可再尋求合理的議價策略；或採取其他推廣策略，讓讀者能增加點擊下載次數，參見表 3。

表 3 期刊跨年度點擊下載次數比例—  
以 2013 年訂購期刊為例

期刊點擊次數(N)	期刊數(J)	比例%(J/2,331)
N>=1	833	35.74
N=0	1,498	64.26
總計	2,331	100.00

資料來源：本研究自行整理

由上述表 2 與表 3 結果，若從 COUNTER 的報表類型 Journal Report 1 (JR1) 為例，可得知依據月份及期刊名稱統計單本期刊內文章全文的有效需求數，由於代理伺服器網際網路紀錄檔之時間戳記欄位，可將系統記錄之時間轉換至一般常見的年月日，而 JR1 報表類型僅能依據到月份，故可依據需求的不同，將報表的時間區隔可從「年」、「月」、「日」亦可區分。

## (三) 年度點擊(含下載)電子資料庫次數分析—以 2013 年為例

若以 2013 全年度為時間區隔變數，計有 349,358 篇文獻點擊下載紀錄，分散於 562 種期刊，期刊之平均點擊下載次數為 621.6 次，另點擊下載次數大於平均值的期刊種數計有 88 種，占 2013 年總期刊數 24.11%，且該期刊群之點擊下載次數合計為 304,270 次，占 2013 全年度總點擊下載次數 87.09%，換言之，該 88 種期刊為 2013 全年度較為重要之期刊群，參見表 4。

表 4 單一年度期刊點擊下載各分列  
項目與次數—以 2013 年為例

項 目	次數/比例
期刊總數(A)	2,331
2013 年點擊(含下載)期刊數(B)	562
2013 年點擊(含下載)總次數(C)	349,358
平均點擊(含下載)次數(D=C/B)	621.6
點擊(含下載)次數>平均值(C)的期刊數(E)	88
>平均值的期刊，其點擊(含下載)次數(F)	304,270
>平均值的期刊，其占期刊總數比例(G=D/A)	24.11%
>平均值的期刊，其點擊(含下載)次數占 2013 年點擊(含下載)總次數比例(H=F/C)	87.09%

資料來源：本研究自行整理

從上列2013年點擊下載期刊再進行分析，以點擊次數多寡為排序依據，顯示*Hepatology*期刊是點擊次數較多的，在38,941點擊次數內，分布於239人。若以學院數為分類區隔，可得知此239人分布於8個學院層級單位，平均每人點擊*Hepatology*期刊文獻次數為162.9次。其次如*The International Journal of Cell Cloning*期刊，點擊次數僅次於*Hepatology*期刊，其點擊人數達180人，分散於9個學院層級單位，其每人平均點擊下載次數雖不如*Hepatology*，仍達超過百次規模。顯示該兩種期刊，其學院層級單位分布較為平均，且平均點擊下載量也較高，顯示該兩種期刊為多數學院層級單位皆有共同閱讀需求。如以排名第三之*Journal of Comparative Neurology*期刊為例，其點擊次數達15,537次，但點擊人數僅68人，其平均每人點擊下載次數雖高於前述*Hepatology*與*The International Journal of Cell Cloning*兩種期刊，但僅分布於5個學院層級單位，顯示該類期刊可能為特定學院或科系認定是重要期刊，參見表5。

表5 2013年期刊點擊下載次數前十名期刊

期刊標題	ISSN	點擊(含下載)各項次數統計			
		次數 (A)	人數 (B)	平均次數 (C=A/B)	學院/單位數 (D)
<i>Hepatology</i>	1527-3350	38,941	239	162.9	8
<i>The International Journal of Cell Cloning</i>	1549-4918	19,058	180	105.9	9
<i>Journal of Comparative Neurology</i>	1096-9861	15,537	68	228.5	5
<i>Journal of Comparative Neurology</i>	1550-7130	15,537	68	228.5	5
<i>Journal of Comparative Neurology and Psychology</i>	1550-7149	15,537	68	228.5	5
<i>International Journal of Cancer</i>	1097-0215	9,194	264	34.8	8
<i>Journal of Cellular and Comparative Physiology</i>	1553-0809	8,918	245	36.4	8
<i>Journal of Cellular Physiology</i>	1097-4652	8,918	245	36.4	8
<i>Journal of Cellular Biochemistry</i>	1097-4644	6,591	263	25.1	8
<i>Journal of Supramolecular Structure</i>	1547-9366	6,591	263	25.1	8
<i>Journal of Supramolecular Structure and Cellular Biochemistry</i>	1547-1748	6,591	263	25.1	8
<i>The Journal of Pathology and Bacteriology</i>	1555-2039	6,529	150	43.5	8
<i>The Journal of Pathology</i>	1096-9896	6,529	150	43.5	8
<i>EMBO Molecular Medicine</i>	1757-4684	6,205	65	95.5	6
<i>Journal of Biomedical Materials Research</i>	1097-4636	5,879	86	68.4	7
<i>Journal of Biomedical Materials Research Part A</i>	1552-4965	5,879	86	68.4	7
<i>Journal of Biomedical Materials Research Part B: Applied Biomaterials</i>	1552-4981	5,879	86	68.4	7
<i>Glia</i>	1098-1136	5,637	83	67.9	6

說明：部份期刊因名稱或ISSN改變，事實上仍是同一期刊，如*Journal of Comparative Neurology*、*Journal of Comparative Neurology and Psychology*為同一期刊。

資料來源：本研究自行整理

#### (四) 跨年度點擊(含下載)期刊之分析—以醫學院為例

以醫學院與期刊ISSN為區隔變數，從2010年8月至2014年4月為時間區間，共有295,161篇文獻下載紀錄，分散於561種期刊，初步以點擊下載次數排序之前十種期刊，參見表6。

表6 跨年度期刊點擊下載次數—以醫學院為例之前十名期刊

期刊標題	ISSN	點擊(含下載)次數
<i>Hepatology</i>	1527-3350	30,075
<i>The International Journal of Cell Cloning</i>	1549-4918	16,692
<i>Journal of Cellular and Comparative Physiology</i>	1553-0809	16,422
<i>Journal of Cellular Physiology</i>	1097-4652	16,422
<i>International Journal of Cancer</i>	1097-0215	8,469
<i>Glia</i>	1098-1136	8,419
<i>Journal of Cellular Biochemistry</i>	1097-4644	6,889
<i>Journal of Supramolecular Structure</i>	1547-9366	6,889
<i>Journal of Supramolecular Structure and Cellular Biochemistry</i>	1547-1748	6,889
<i>The Journal of Pathology and Bacteriology</i>	1555-2039	5,724
<i>The Journal of Pathology</i>	1096-9896	5,724
<i>Annals of Neurology</i>	1531-8249	4,721
<i>British Journal of Pharmacology</i>	1476-5381	4,708
<i>Journal of Neuroscience Research</i>	1097-4547	4,623

資料來源：本研究自行整理

由表6，初步分析自2010年8月至2014年4月期間，醫學院各系所之點擊(含下載)次數合計以*Hepatology*期刊為最高，計有30,075次，其次為*The International Journal of Cell Cloning*，計有16,692次。其他依期刊ISSN為區隔，並按下載次數由高至低為：*Journal of Cellular and Comparative Physiology*(16,422)、*Journal of Cellular Physiology*(16,422)、*International Journal of Cancer*(8,469)、*Glia*(8,419)、*Journal of Cellular Biochemistry*(6,889)、*Journal of Supramolecular Structure*(6,889)、*Journal of Supramolecular Structure and Cellular*(6,889)、*The Journal of Pathology and Bacteriology*(5,724)、*The Journal of Pathology*(5,724)、*Annals of Neurology*(4,721)、*British Journal of Pharmacology*(4,708)、*Journal of Neuroscience Research*(4,623)。

進一步對於點擊(含下載)次數相同的期刊進行分析，部分點擊(含下載)次數相同的期刊，在進行書目資料比對後發現，如*Journal of Cellular Biochemistry*、*Journal of Supramolecular Structure*、*Journal of Supramolecular Structure and Cellular*等名稱與ISSN不同的期刊，在Wiley資料庫進行檢索驗證時，皆為同一份期刊，但在期刊清單內，若無特別註記，在資料處理時易視為各自獨立的期刊，但實際上為同一份期刊，在資料分析時亦容易造成處理的困難。

### (五) 單一年度期刊與學科領域科系之點擊(含下載)分析—以公共衛生相關科系為例

以公共衛生相關領域科系為區隔，該領域共有6個系所，以2013年為例，在2,331種期刊中，有117種期刊曾為公衛相關領域科系點擊(含下載)之數據，共點擊2,198次，可做為公衛領域相關期刊的初步區隔，而每一期刊平均點擊次數為18.78次。茲將117種期刊做進一步區隔，其分析步驟如下：

1. 以期刊平均點擊次數為基準，共有28種期刊高於平均點擊次數。
  2. 將28種期刊以曾點擊之系所個數為區隔變數，超過半數系所( $N \geq 3$ )曾點擊(含下載)該期刊，則視為重要程度較高之期刊群(A)，計有8種期刊。
  3. 由前一步驟，點擊(含下載)次數超過平均次數且點擊該期刊之系所未達總數系所之半( $N < 3$ )，則視為次重要之期刊群(B)，計有20種期刊。
  4. 低於期刊平均點擊次數則視為重要性較低期刊群(C)，計有89種期刊。
- 透過上述步驟，可區隔出以點擊(含下載)次數與系所為基準的3種期刊群。參見表7可知，2013年對於公共衛生領域科系，點擊(含下載)次數皆超過百次的三種期刊，分別為*Health Economics*(353)、*Arthritis & Rheumatism*(176)、*Statistics in Medicine*(111)三種期刊，且達一半的科系曾點擊(含下載)，其中*Health Economics*為點擊(含下載)次數最多，但*Statistics in Medicine*則為全部6類系所皆曾點擊(含下載)；其次低於百次但高於平均點擊(含下載)次數的期刊群共有5種期刊，分別為*Arthritis Care & Research*(88)、*International Journal of Cancer*(69)、*Pharmacoepidemiology and Drug Safety*(33)與*Arthritis & Rheumatology*(22)等五種，故這兩種分類之期刊群可做為公共衛生科系領域核心期刊之參考。

表7 較重要期刊群(A) —以公衛領域科系為例

期刊標題	ISSN	科系/點擊(含下載)次數						點擊 (含下載) 次數
		H1	H2	H3	H4	H5	H6	
<i>Health Economics</i>	1099-1050	110	188	54	1			353
<i>Arthritis &amp; Rheumatism</i>	1529-0131		92	54	30			176
<i>Statistics in Medicine</i>	1097-0258	66	21	11	5	4	4	111
<i>Arthritis Care &amp; Research</i>	2151-4658		46	27	15			88
<i>International Journal of Cancer</i>	1097-0215	29	33	4	3			69
<i>Addiction</i>	1360-0443	8			16		20	44
<i>Pharmacoepidemiology and Drug Safety</i>	1099-1557	1	7	23	2			33
<i>Arthritis &amp; Rheumatology</i>	2326-5205		12	9		1		22

資料來源：本研究自行整理

### (六) 特定期刊文獻之點擊(含下載)分析

以*Hepatology*期刊的「Comprehensive analysis of the independent effect of twist

and snail in promoting metastasis of hepatocellular carcinoma」文獻為例（註21），其DOI為「10.1002/hep.23221」，篩選出自2010-2014年的點擊（含下載）紀錄，共有2,144次。該文章在2011年與2012年之點擊次數接近，約450至480次，至2013年則達1,073次，跨年度之點擊讀者所屬學院/單位分布，以醫學院（5次）為最多，其次為生命科學院（4次）、生物醫學工程學院（3次）、牙醫學院（3次）、行政支援單位（2次），最少則是護理學院與藥學院各出現1次，顯示該文獻為醫學院與生命科學院之科系較具有點擊（含下載）需求，參見表8。

表8 特定文獻跨年度點擊次數與讀者所屬學院/單位分析

年份	點擊讀者所屬學院/單位分布	點擊(含下載)次數
2010	醫學院、生命科學院	64
2011	醫學院、生物醫學工程學院、牙醫學院、護理學院	457
2012	醫學院、生物醫學工程學院、生命科學院、牙醫學院、行政支援單位	486
2013	醫學院、生物醫學工程學院、生命科學院、牙醫學院、藥學院、行政支援單位	1,073
2014	醫學院、生命科學院、	64

資料來源：本研究自行整理

## 五、結論與建議

本研究提出一電子資源下載使用分析模式，以醫學電子資源做為研究個案進行該模式驗證，資料來源為某醫學大學圖書館之網際網路紀錄檔，綜整研究過程，除可分析該醫學大學圖書館電子資源之使用狀況，並依據圖書館之個別需求，進行資料統計與產生各項報表之研究結果數據。

### (一) 結論

本研究之結論，分為四點，包括1.提出圖書館電子資源下載使用分析之模式；2.使用數據分析顯示集中在少數期刊；3.重要期刊的選擇—透過學院、特定領域科系與期刊交叉分析；4.學院單位對特定文獻的需求分析，分述如下：

#### 1. 提出圖書館電子資源下載使用分析之模式

過去圖書館在進行電子資源使用分析，多仰賴廠商提供之數據，實務上呈現格式皆為採用COUNTER的電子資源統計報表，所得數據仍是以學術機構為單位的整體描述性統計，對於學術單位之子單位等細項分析卻無法得知，為解決分析上的困難，本研究依據現行的困境與實際需求，提出一電子資源下載使用分析模式，再以醫學電子資源做為研究個案進行模式驗證。該模式可依據不同需求，可做進階目的分析與使用報表，對於圖書館在期刊經費的編列與控管，更能提供其另一種觀點的參考。

#### 2. 使用數據分析顯示集中在少數期刊

本研究以2013年所訂購2,331期刊為基準，在橫跨多個年度的點擊（含下



載)次數分析,僅有833種期刊點擊(含下載)次數為至少1次,若以2013年單一年度作為時間區隔因素,僅有562種期刊曾為使用者點擊(含下載),顯示圖書館所訂購電子期刊,使用者點擊(含下載)數據集中在特定少數期刊,未來可協助圖書館在訂購政策的參考,同時亦可針對使用者的類型與文獻再進行深入探討。

### 3. 重要期刊的選擇—透過學院、特定領域科系與期刊交叉分析

本研究分別以醫學院與公共衛生相關科系為個案說明,顯示醫學院之重要期刊群為*Hepatology*、*The International Journal of Cell Cloning*、*Journal of Cellular and Comparative Physiology*,其跨年度下載次數皆超過萬次以上;而公共衛生相關科系之重要期刊群為*Health Economics*、*Arthritis & Rheumatism*、*Statistics in Medicine*、*Arthritis Care & Research*、*International Journal of Cancer*、*Addiction*、*Pharmacoepidemiology and Drug Safety*、*Arthritis & Rheumatology*等8種期刊。

### 4. 學院單位對特定文獻的需求分析

本研究以*Hepatology*期刊的「Comprehensive analysis of the independent effect of twist and snail in promoting metastasis of hepatocellular carcinoma」文獻為例,若以跨年度區隔分析,指出該文獻為醫學院與生命科學院之科系較具有點擊(含下載)需求,顯示日後若配合課程或研究需求,圖書館可提供相關建議給該學院或科系,以滿足在研究或課程準備之需求。

## (二) 建議

至於本研究的建議,分為1.圖書館自行建置代理伺服器;2.自行建置讀者檔之備援資料庫;3.建立期刊資料之沿革歷史檔;4.透過雲端技術儲存與分析巨量點擊下載資訊;5.應用領域宜延伸至其他學科領域;6.配合相關專業領域,執行專家訪談等六點,分述如下:

### 1. 建議圖書館自行建置代理伺服器

在身份驗證層面,本研究提出建議有二:(1)圖書館自行建置代理伺服器;(2)限縮IP驗證範圍,僅將代理伺服器做為對外連至電子資料庫惟一合法IP。實務上,圖書館為考量讀者的便利性,在訂購電子資料庫與電子期刊時,會將屬於學校或機構內的IP範圍通知廠商,協助設定納入合法使用的網路IP範圍,亦即若讀者在校園或機構的網路連線至電子資料庫,僅驗證是否為合法網路IP範圍,無須透過身份驗證即可檢索與下載文獻。但從圖書館管理角度,讀者若未透過圖書館提供的代理伺服器進行身份驗證,圖書館則無法得知讀者的點擊下載狀況,故建議圖書館自行建置代理伺服器,無論讀者是否使用校園網路或在家連線,皆須連線至圖書館的代理伺服器,取得合理身份驗證後再連至電子資料庫,藉以取得合法讀者之點擊下載紀錄。

## 2. 自行建置讀者檔之備援資料庫

圖書館自動化系統內存的讀者檔，因讀者資料隨著時間、情境等因素進行更改讀者的狀態，如實務上大學圖書館在為學生辦理離校流程時，會將讀者身份類型由「學生」變更為「校友」類型，雖可明確辨識該學生為校友身份，且限制使用者權限，但在進行資料清理程序時，須多重回溯程序始能得知學生的原始科系，故本研究建議，除原本圖書館自動化系統的原始讀者檔，應另行建構讀者檔之備援資料庫，以利資料回溯。

## 3. 建立期刊資料之沿革歷史檔

本研究在進行資料比對過程中，因期刊會有改換名稱或ISSN的狀況，在電子資料庫廠商提供之年度期刊清單中，雖列出年度訂購之期刊資料，但無法確知期刊之間的關聯性，必須自行依據網址的樣式重新回溯與整理，故建議建立期刊資料之沿革歷史檔，以供期刊歷史資料之比對，

## 4. 透過雲端技術儲存與分析巨量點擊下載資訊

電子資源使用資訊，其點擊下載資料為即時產生，因資訊量之巨，無法純以人力進行資料處理，但資訊清理過程中會因大量讀寫而影響效能，亦無法以單一主機或伺服器處理，建議透過雲端主機進行儲存、更新、處理、分析等動作，一則減少實體主機之效能損耗，另可透過分散式架構進行資料處理，增加資料處理與檢索之效率。

## 5. 應用領域宜延伸至其他學科領域

本研究雖以醫學圖書館為研究個案，仍可適用於其他學科領域，如IEEE等，雖其困難點在於最初的原始資料收集、儲存、清理、整合等過程，與圖書館自動化系統與期刊書目等資訊之縱向與橫向整合，透過本研究建構之分析模式，可將巨量之原始使用資料，透過系統化的資料處理與分析方法，減少人工作業，依據圖書館之內部需求，增加分析類型與輸出所需資料，亦可針對其他學科領域進行期刊分類與分析，對於圖書館之管理層面，達到有效管理之目的，在學術研究方面，其輸出資料可應用在資訊計量學、書目探勘、資訊行為等圖書資訊學之研究領域。

## 6. 配合相關專業領域，執行專家訪談

本研究之研究結果為依據代理伺服器網際網路紀錄檔進行分析與統計，對於使用者本身在點擊與下載動機，或進一步的資訊尋求行為並無法直接得知，若能執行專家訪談，能更進一步得知使用者之相關資訊需求，但因個人資料保護法的法律規範，在執行層面上，必須事先取得原始使用者之同意，始能進行。

綜上所述，電子資源使用分析對於圖書館而言，無論實務層面的期刊續訂與控管、館藏政策的制定，或從資訊計量學、書目探勘、資訊行為等圖書資訊

學相關領域研究，皆提供有效數據以資分析或參考；圖書館或研究者亦可採用雲端運算進行大量數據分析，進而提供更多元觀點的實務分析與研究探討。

## 註 釋

註1 王梅玲，「電子計量應用在電子期刊評鑑」，*圖書與資訊學刊*，57期(2006年5月)：2。

註2 Rowlands Ian and David Nicholas, "The Missing Link: Journal Usage Metrics," *Aslib Proceedings* 59, no. 3 (2007): 223.

註3 郭麗芳，「網路電子期刊評估之研究」，*大學圖書館* 1卷，3期(1997年7月)：58-59，檢索於2014年6月13日，[http://www.lib.ntu.edu.tw/Publication/univj/uj1-3/uj3\\_5.html](http://www.lib.ntu.edu.tw/Publication/univj/uj1-3/uj3_5.html)

註4 Linda S. Mercer, "Measuring the Use and Value of Electronic Journals and Books," *Science and Technology Librarianship*, no. 25 (Winter 2000), accessed June 15, 2014, <http://www.istl.org/00-winter/article1.html>.

註5 王梅玲，「電子期刊網路服務與評鑑模式之研究」，*圖書資訊學研究* 3卷，1期(2008年12月)：79，檢索於2014年6月13日，[http://lac3.glis.ntnu.edu.tw/vj-attachment/2009/10/attach33\\_1.pdf](http://lac3.glis.ntnu.edu.tw/vj-attachment/2009/10/attach33_1.pdf)

註6 同上註，99。

註7 詹麗萍，*電子資源與圖書館館藏發展*(台北市：麥田，2005)，175。

註8 同上註，170-173。

註9 同上註，192-193。

註10 Andrew White and Eric Djiva Kamal, *E-Metrics for Library and Information Professionals* (New York: Neal-Schuman, 2006), 5-6.

註11 Peter T Shepherd, "Counter: Usage Statistics for Performance Measurement," *Performance Measurement and Metrics* 7, no. 3 (2006): 142-143.

註12 葉諶敬，「電子資源統計問題看COUNTER計畫之發展」，*教育資料與圖書館學* 43卷，4期(2006年6月)：453，檢索於2014年6月2日，<http://joemls.dils.tku.edu.tw/fulltext/43/43-4/453-469.pdf>

註13 COUNTER, "COUNTER: Counting Online Usage of NeTworked Electronic Resources," COUNTER, March 1, 2002, accessed May 28, 2014, <http://www.projectcounter.org/index.html>.

註14 Association of Research Libraries, "Statistics & Assessment," Association of Research Libraries, n.d., accessed June 2, 2014, <http://www.arl.org/focus-areas/statistics-assessment>.

註15 COUNTER, "The COUNTER Code of Practice for e-Resources: Release 4", COUNTER, April 2012, accessed May 28, 2014, <http://www.projectcounter.org/r4/COPR4.pdf>.

註16 COUNTER, "The COUNTER Code of Practice for Articles: Release 1", COUNTER, March 2014, accessed May 28, 2014, [http://www.projectcounter.org/documents/counterart\\_cop\\_MAR2014.pdf](http://www.projectcounter.org/documents/counterart_cop_MAR2014.pdf).

註17 NISO, "Standardized Usage Statistics Harvesting Initiative(SUSHI)—National Information Standards Organization," NISO, 2013, accessed May 28, 2014, <http://www.niso.org/>

workrooms/sushi.

註18 Marshall Breeding, "Offering Remote Access to Restricted Resources," *Information Today* 18, no.4 (May 2001): 52-53, accessed September 27, 2014, <http://www.librarytechnology.org/ltg-displaytext.pl?RC=8991>, 轉引自詹麗萍, 電子資源與圖書館館藏發展, 180。

註19 Jiawei Han and Micheline Kamber, 資料探勘：概念與方法, 二版, 王派洲譯 (台中市：滄海書局, 2008), 50-52。

註20 SARG, "Squid Analysis Report Generator," Sourceforge.net, 2002, accessed March 8, 2013, <http://sarg.sourceforge.net>.

註21 Muh-Hwa Yang et al., "Comprehensive Analysis of the Independent Effect of Twist and Snail in Promoting Metastasis of Hepatocellular Carcinoma," *Hepatology* 50, no.5 (November 2009), accessed May 14, 2014, <http://onlinelibrary.wiley.com/doi/10.1002/hep.23221/abstract>.



# The Study of Analytical Model of Library Electronic Resources Usage: A Case of Medical Electronic Resources

Chung-Yen Yu<sup>a</sup> Jiann-Cherng Shieh<sup>b\*</sup>

## Abstract

*With the advents of internet, the importance of electronic resources is growing. Due to the increasing expensiveness of electronic resources, university libraries normally received budgets from parent institutions annually. They necessarily applied effective and systematic methods for decision making in electronic resources purchase or re-subscription. However, there are some difficulties in practices: First of all, libraries are unable to receive user records; second, the COUNTER statistics does not include details about users and their affiliation. As a result, one cannot conduct advanced user analysis based on the usage of users, institutions, and departments. To overcome the difficulties, this study presents a feasible model to analyze electronic resource usage effectively and flexibly. We set up a proxy server to collect actual usage raw data. By analyzing items in internet browsing records, associated with original library automatic system, this study aims at exploring how to use effective ways to analyze big data of website log data. We also propose the process of how original data to be transformed, cleared, integrated, and demonstrated. This study adopted a medical university library and its subscription of medical electronic resources as a case. Our data analysis includes (1) year of subscription, (2) title of journal, (3) affiliation, (4) subjects, and (5) specific journal requirements, etc. The findings of the study are contributed to obtain further understanding in policy making and user behavior analysis. The integrated data provides multiple applications in informatics research, information behavior, bibliomining, presenting diverse views and extended issues for further discussion.*

**Keywords:** *Electronic resource usage analysis; Electronic journal; Informatics; Information behavior; Bibliomining*

## SUMMARY

### Introduction

With the nature of going beyond traditional carriers, having no limits or boundaries in time and space, and being transmitted rapidly, electronic resources

<sup>a</sup> Ph.D. Candidate, Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taipei, Taiwan

<sup>b</sup> Professor, Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taipei, Taiwan

\* To whom all correspondence should be addressed. E-mail: jcsieh@ntnu.edu.tw

thrive in this era. With the emergence of both computers and the Internet, it forces most libraries to update all the sources in an electronic format and changes the way how we access articles. Electronic resources have several advantages compared to the traditional print sources, which include: 1) there are no boundaries in time and space for resource retrieval; 2) it has faster retrieval time; and 3) the cost of resource retrieval is inexpensive. By the reasons of rapid internet development and library budget reduction, recently academic libraries would mainly order electronic journals and related resources instead of paper journals. Because of the internet's rapid development and library budget cuts, in recent years when academic libraries are purchasing journals, the first priority would be ordering electronic journals and relevant resources, instead of journals in paper formats. This situation has dramatically changed the ways of publishing, organizing and preserving academic resources, and thus affected academic library collection policies and related analysis approaches to the purposes for effective management of library collections.

From evaluating library collections viewpoints, libraries tend to focus on investigating electronic resources collections usage and the degree of meeting patrons' needs that were approved by analyzing the statistic data of collections usage. However, with electronic resources' nature of high variability and ever-increasing quantities, the traditional approaches of evaluating paper journals usages cannot reveal the real usage status of electronic resources. It is also difficult for libraries to obtain raw data on usage statistic data from vendors of electronic resources. Furthermore, it is still another technical difficulty when librarians consider how to conduct usage analysis on databases and electronic journals. Therefore, in such varied, multiple and complex environments, it is important for libraries to develop effective analysis approaches to evaluate electronic resources, and help them to cope with multiple obstacles mentioned above.

Based on reviewing the difficulties of analyzing electronic resources that libraries were faced, this study was aimed to solve such practical problems by proposing a systematic analysis model for electronic resources usage. In order to verify the model available, this research conducted medical electronic resources as study case, to illustrate the processes of model construction and how the raw data was applied to analyze to output bibliographical data and detail records of databases usages. The analyzed results can be used to be the reference of electronic resources usage investigations in libraries.

### **Research Goals and Research Questions**

In order to help librarians investigate and analyze the usage behaviors of libraries' electronic resources, this research tried to propose a practical and

systematic model from collecting, cleaning, clustering, and analyzing usage data. Thus, the research main questions of this study include (1) How to build a usage analysis model for electronic medical resources? (2) How to analyze raw data on internet records? (3) How to conduct statistics analysis for individual colleges and departments with different natures? For searching the answers to the research questions mentioned above, this study is designed for conducting case studies on the electronic medical resources provided by a medical university in Taiwan. For verification purpose of this case, the researchers of this study have first built a proxy server for the special purpose of this study to collect raw data about users' electronic full-text clicking and downloading operations. Two research limitations need to be specified here. First, for the privacy consideration of personal data protection, cases would be discussed only at the levels of departments and colleges; Second, all discussed cases include only full text PDF files with clicks and downloads, not including files in HTML formats.

## **Methods**

For solving problems regarding usage analysis of electronic resources, the researchers of this study propose a usage analysis model for electronic resources, and focus on four concepts, including physical data, data preprocessing, data representation, and applied fields. The point of these four concepts lies in how to use a systematic analysis approach to clean and process usage data, export the cleaned data for further advanced uses, and then create usage records and relevant bibliographies for different applied fields. The analysis outcomes are expected to provide new data sources for relevant researches in library and information studies.

For data collection of this study, 4,181,850 records from Wiley databases are selected for the case study university library, with a total file size of 1.04 GB. The recorded period is from August 17, 2009, to April 21, 2014. After excluding unrecognizable or incomplete records, these records are selected with a SAS approach, based on the 2013 journal list. Before a further analysis of literature bibliographies, the number of clicks and downloads of PDF full-text is 1,216,156.

## **Findings**

A usage analysis model for electronic resources is proposed in this study, and tested in an electronic medical database for case studies. Six major findings include: (1) the total amount of clicks and downloads of the electronic database within 5 years; (2) a usage analysis of the electronic database; (3) the annual analysis of amounts of clicks and downloads of the electronic database; (4) a

cross-year analysis of clicks and downloads of journals—taking the Medical School for example; (5) a single-year analysis of clicks and downloads of journals among different departments and academic fields—taking the Public Health departments for example; and (6) an analysis of clicks on specific journals and literature. The first two statistics can be obtained from Counting Online Usage of NeTworked Electronic Resources (COUNTER) statistics, whereas the last four statistics could not be obtained from the COUNTER statistics.

## Conclusions

In this study, we propose a practical and systematic analysis model for library electronic resources usage, and validate this model based on the case of medical electronic database. Through experiments and analysis, the conclusions will be explored in four facets of findings, including (1) the usage analysis model for downloading and using of library electronic resources is proposed; (2) the data analysis that the majority journal revealed retrievals are only fallen on a few specific journals based on the number of user clicks and journal downloads; (3) a cross analysis is conducted among the variables of colleges, specified academic fields and journals, to reveal the choice for important journals; (4) a need analysis is conducted to reveal the special needs of different colleges or departments for journals or literatures.

In addition, the researchers of this study expect that not only that the proposed model can be utilized in the case of medical electronic resources, but also can be applied to other science fields. In other words, our model can help provide more incorporated and integrated applications for library and information related researches and studies, such as bibliography statistics, bibliomining, and user behaviors modeling in electronic resources, etc.

### **ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT**

註1 王梅玲 [Mei-Ling Wang], 「電子計量應用在電子期刊評鑑」[“The Application of E-Metrics to Electronic Journal Evaluation,”], 圖書與資訊學刊, 57期(2006年5月): 2 [Bulletin of Library and Information Science, no. 57 (May 2006): 2].

註2 Rowlands Ian and David Nicholas, “The Missing Link: Journal Usage Metrics,” *Aslib Proceedings* 59, no. 3 (2007): 223.

註3 郭麗芳 [Li-fong Guo], 「網路電子期刊評估之研究」[“A Research on the Evaluation of Electronic Journals on the Network,”], 大學圖書館 1卷, 3期(1997年7月): 58-59 [University Library Quarterly 1, no.3 (July 1997): 58-59], 檢索於2014年6月13日 [accessed June 13, 2014], [http://www.lib.ntu.edu.tw/Publication/univj/uj1-3/uj3\\_5.html](http://www.lib.ntu.edu.tw/Publication/univj/uj1-3/uj3_5.html)

註4 Linda S. Mercer, “Measuring the Use and Value of Electronic Journals and Books,” *Science and Technology Librarianship*, no. 25 (Winter 2000), accessed June 15, 2014, <http://www.istl.org/00-winter/article1.html>.



註5 王梅玲 [Mei-Ling Wang], 「電子期刊網路服務與評鑑模式之研究」[“A Study of Electronic Journals’ Network Service and Evaluation Models,”] 圖書資訊學研究 3 卷, 1 期 (2008 年 12 月): 79 [Journal of Library and Information Science Research 3, no. 1 (December 2008): 79], 檢索於 2014 年 6 月 13 日 [accessed June 13, 2014], [http://lac3.glis.ntnu.edu.tw/vj-attachment/2009/10/attach33\\_1.pdf](http://lac3.glis.ntnu.edu.tw/vj-attachment/2009/10/attach33_1.pdf)

註6 同上註 [Ibid.], 99。

註7 詹麗萍 [Li-Ping Chen], 電子資源與圖書館館藏發展 [Electronic Resources and Library Collection Development] (台北市: 麥田, 2005) [(Taipei: Rye Field, 2005)], 175。

註8 同上註 [Ibid.], 170-173。

註9 同上註 [Ibid.], 192-193。

註10 Andrew White and Eric Djiva Kamal, *E-Metrics for Library and Information Professionals* (New York: Neal-Schuman, 2006), 5-6.

註11 Peter T Shepherd, “Counter: Usage Statistics for Performance Measurement,” *Performance Measurement and Metrics* 7, no. 3 (2006): 142-143.

註12 葉謙啟 [Hui-Hsin Yeh], 「電子資源統計問題看 COUNTER 計畫之發展」[“Statistical Problems of Electronic Resources and the Development and Application of COUNTER Project,”] 教育資料與圖書館學 43 卷, 4 期 (2006 年 6 月): 453 [Journal of Educational Media & Library Sciences 43, no. 4 (July 2006): 453], 檢索於 2014 年 6 月 2 日 [accessed June 2, 2014], <http://joemls.dils.tku.edu.tw/fulltext/43/43-4/453-469.pdf>

註13 COUNTER, “COUNTER: Counting Online Usage of Networked Electronic Resources,” COUNTER, March 1, 2002, accessed May 28, 2014, <http://www.projectcounter.org/index.html>.

註14 Association of Research Libraries, “Statistics & Assessment,” Association of Research Libraries, n.d., accessed June 2, 2014, <http://www.arl.org/focus-areas/statistics-assessment>.

註15 COUNTER, “The COUNTER Code of Practice for e-Resources: Release 4,” COUNTER, April 2012, accessed May 28, 2014, <http://www.projectcounter.org/r4/COPR4.pdf>.

註16 COUNTER, “The COUNTER Code of Practice for Articles: Release 1,” COUNTER, March 2014, accessed May 28, 2014, [http://www.projectcounter.org/documents/counterart\\_cop\\_MAR2014.pdf](http://www.projectcounter.org/documents/counterart_cop_MAR2014.pdf).

註17 NISO, “Standardized Usage Statistics Harvesting Initiative (SUSHI)—National Information Standards Organization,” NISO, 2013, accessed May 28, 2014, <http://www.niso.org/workrooms/sushi>.

註18 Marshall Breeding, “Offering Remote Access to Restricted Resources,” *Information Today* 18, no.4 (May 2001): 52-53, accessed September 27, 2014, <http://www.librarytechnology.org/lgtg-displaytext.pl?RC=8991>, 轉引自 [quoted in] 詹麗萍 [Li-Ping Chen], 電子資源與圖書館館藏發展 [Electronic Resources and Library Collection Development], 180。

註19 Jiawei Han and Micheline Kamber, 資料探勘: 概念與方法, 二版, 王派洲譯 [Data Mining: Concepts and Techniques, 2nd ed., trans., Pai-Chou Wang] (台中市: 滄海書局, 2008) [(Taichung, Taiwan: Tsang Hai, 2008)], 50-52。

註20 SARG, “Squid Analysis Report Generator,” Sourceforge.net, 2002, accessed March 8, 2013, <http://sarg.sourceforge.net>.

註21 Muh-Hwa Yang et al., “Comprehensive Analysis of the Independent Effect of Twist and Snail in Promoting Metastasis of Hepatocellular Carcinoma,” *Hepatology* 50, no.5 (November 2009), accessed May 14, 2014, <http://onlinelibrary.wiley.com/doi/10.1002/hep.23221/abstract>.



Chung-Yen Yu ORCID 0000-0001-6531-2206

Jiann-Cherng Shieh ORCID 0000-0003-0353-0106