

# 圖書館資料倉儲應用模式之研究 以客製化報表爲例

The Study of Data Warehouse Application Model in Libraries  
A Case Study of Customized Reports

楊依婷

**Yi-Ting Yang**

國立臺灣師範大學圖書資訊學研究所博士生

Ph.D. Student

Graduate Institute of Library and Information Studies

National Taiwan Normal University

謝建成\*

**Jiann-Cherng Shieh**

國立臺灣師範大學圖書資訊學研究所教授

Professor

Graduate Institute of Library and Information Studies

National Taiwan Normal University

## 【摘要 Abstract】

圖書館客製化報表是圖書館管理者亟欲解決的問題之一，圖書館自動化系統廠商對此至今依舊難有具成本效益之解決方案。近幾年，ExLibris 公司針對其自動化系統 Aleph 開發報表系統 ARC，提供客戶超過 100 種事先定義的報表。然此報表系統僅支

---

\*通訊作者：謝建成 [jcsieh@ntnu.edu.tw](mailto:jcsieh@ntnu.edu.tw)

投稿日期：2017 年 3 月 12 日；接受日期：2017 年 6 月 30 日

援 Aleph 系統，對國內多數非 Aleph 系統圖書館報表需求毫無助益，且無法滿足各館即時性多元報表需求。在基於成本效益及客製化需求，本研究提出以 SQL Server 之資料倉儲應用模式解決以上困境。本研究以實際案例驗證模式之可行性。

The development of library customized reports has always been one of the intractable problems that library managers urgently seek to solve. Similarly, library automation system manufacturers are bruised on this issue, not only the high cost of development services, but also the frequent changes in demand. The problem is still difficult to have by a cost-effective solution. In recent years, ExLibris has developed an add-on system Aleph Reporting Center (ARC) for its library automation system Aleph to provide Aleph customers more than 100 pre-defined statistical reports. However, such reporting systems only support specific automation systems, and cannot fully meet the customized needs of libraries'. Data warehouse technology has been quite mature since the 1970s. Most database management systems such as SQL Server also provide tools to support the establishment of data warehouses. Considering the cost-effectiveness and the demand elasticity of customization, this study proposes the data warehouse application model of libraries to solve the above dilemma. The application model proposed in this research, the system environment is constructed by Microsoft SQL Server, EXCEL; design and development are mainly accomplished by the librarian and system librarian. This study concludes with a practical case to verify the feasibility of the proposed model.

### 【關鍵字 Keywords】

資料倉儲；圖書館自動化；客製化報表

Data Warehouse; Library automation system; Customized reports

## 壹、前言

長久以來，圖書館作業相關報表的繁雜需求，一直是困擾圖書館自動化系統開發商沉痾已久的問題。以往的解決方式，多數是以客製

化另案開發方式解決。圖書館若採用國內廠商開發之自動化系統，其報表客製化需求，多數是由系統廠商以系統維護名義或是新增功能模式協助解決；而採用國外廠商開發之自動化系統者，其客製化報表需求解決之道，若非個案開發索價驚人，亦僅能反映原廠公司待可能增加新功能之改版系統。近幾年圖書館資訊能力逐漸強化，再加上圖書館自動化系統後端資料庫的開放，系統館員亦已能夠藉由結構化查詢語言 SQL (Carlos & Morris, 2016) 承擔部分報表產生，供管理階層解決燃眉之急。

Ex Libris 於 2005 年首先針對其圖書館自動化系統 Aleph 發展一以資料倉儲技術為基礎之報表產生系統 Aleph Reporting Center (簡稱 ARC)，提供其客戶選擇適於各館所需之預存報表，以協助解決圖書館報表需求之窘境。先不論各館應用 ARC 效益及功能如何，對於其他非 Aleph 各館之報表需求，該如何有效彈性解決？是圖書館自動化系統運用向來亟待解決的問題。

另外，台灣圖書館管理議題一直是少被重視的，尤其是管理決策層面的問題，或因牽涉太多環境政治因素影響，多數仍舊偏重於管理者經驗的決定，以實證為主 (Evidence-Based) 之圖書館管理決策，卻是少見。然在其他應用領域，資料倉儲技術卻是已經可提供環境運作事實結果供決策者參考，用以改善其作業效率，增進其服務品質。此部分之需求，當然會因各館之狀況不同而有所差異，也因此非事先定義之資料倉儲架構可以完整解決的，必須進行相關客製化以滿足圖書館個別之需求。

基於上述圖書館報表需求與管理決策參考需要，本研究提出一以資料倉儲為基礎之圖書館自動化系統客製化報表產生模式，並藉由實際案例證實，在現有稍具資料庫基礎之圖書館系統館員作業下，此模式之可行性及其效益性。

## 貳、文獻回顧

本研究於探討圖書館自動化報表客製化，所涉及之資訊技術概念，主要有：資料倉儲 (Data Warehouse)、資料超市 (Data Mart)、多維度資料模式 (Multi-Dimensional Data Model)、事實資料表 (Fact Table) 等。

## 一、資料倉儲

就現今資料庫資訊技術而言，資料倉儲（Inmon, 2008; Kimball & Ross, 2013; Massis, 2016; Nicholson, 2003; Nicholson, 2006; Shieh, 2010; Wah, Peng & Hok, 2007; 謝建成，2009）並不是甚麼特別新穎的技術，其應用範圍相當廣泛，主要是以支援提供管理階層所需之資訊。台灣圖書館於管理階層資訊之需求，多數是以固定週期之呈報報表為主，如呈報給上級單位之月報、季報、年報等。

資料倉儲為整合性的資訊儲存體（Repository），是由多個分散式、自主性、異質性的資料來源（data sources）彙整而成以供查詢及分析。運用資料倉儲主要目的，一是資料管理：由資料不同的角度從事資料分析或資料整合，以有效掌握現存資料狀態；二是資料展現：專注在建立新種類的報告，例如策略性的趨勢分析，以提供決策所需資訊。

Inmon（2008）定義資料倉儲：「是一俱主題導向（subject-oriented）、整合性（integrated）、時間變異性（time-variant）及不變動性（nonvolatile）的大量資料聚合，以支援管理之決策過程」。

### （一）主題導向

資料倉儲系統不同於日常操作異動系統，期主要是針對輔助決策者進行資料分析，因此資料倉儲必須依據某個主題或領域來建置，排除於決策過程中非參考性資料；以保險業為例，一般異動作業系統以壽險、健康險、意外險等產品為主題，而資料倉儲則以顧客資料、保險費成本分析、理賠資料分析等為主題。

### （二）整合性

資料倉儲系統之資料來源為不同異動作業層系統或異質資料庫，如關聯式資料庫、一般檔案與線上異動系統紀錄等，而資料型態與定義在其原有系統或資料庫中，可能有不一致的定義。因此資料轉置載入資料倉儲前，必須先經由轉換、篩選或淨化過程，以確保命名、結構與屬性的一致性，因此其資料具有整合性的特性；例如，系統 A 其男 / 女定義為 M/F、系統 B 其男 / 女定義為 1/0、而系統 C 其男 / 女定義為 X/Y，在將此三個系統資料彙整成資料倉儲時，必須重新定義為男 / 女為 M/F。

### （三）資料不變動性

資料倉儲系統與日常操作交易系統是獨立的，一般操作交易系統

其資料庫皆提供資料新增、修改、刪除、查詢等功能，而資料倉儲系統之資料庫僅是資料載入與資料讀取分析功能，其新增、修改與刪除等異動功能，需在原有操作系統之資料庫處理；因此資料倉儲所需的操作只有資料初始載入與資料存取，其資料是具有不變動性的特質。

#### (四) 時間變異性

一般線上作業層異動系統 (online transaction processing, OLTP) 之資料處理，為即時與日常性每日異動資料，而資料倉儲系統處理資料為過去歷史與年度或季節性資料。並經由過去歷史性資料累積，以時間因素為角度來分析，提供長期性資料趨勢與特性，作為支援決策的依據。因此，一般資料倉儲皆有設計時間維度，以分析其資料長期性趨勢與特性。

綜整而言，資料倉儲是不同於日常營運異動的資料庫，是單位於分析特殊需求訂定主題後，萃取自日常異動的資料庫之相關資料所建構成，主要是為了協助單位執行管理上的決策。另外由於資料倉儲具備資料整合的特性，對於分散於不同系統間的資料而言，資料倉儲能提供穩固的平台，以整合性的架構進行系統整合，有效規劃歷史資料的管理。

##### 1. 資料倉儲架構

就系統架構層面探討，資料倉儲架構可分為三層 (Han, Kamber, & Pei, 2011)：資料倉儲伺服器 (data warehouse server)、線上分析處理伺服器 (OLAP server) 與前端工具 (front-end tools)，如圖 1 所示。資料倉儲的資料來源為現有作業層系統或其它外部資料，並經由篩選、轉換、擷取，轉至最底層的資料倉儲伺服器。最底層的資料倉儲伺服器其架構除資料倉儲外，還包含資料超市或定義資料結構的詮釋資料 (Metadata)。資料倉儲資料經由中間層次資料模式與維度定義、資料儲存格式與資料立方體預存處理；輸出層最前端為一般使用者報表、查詢、線上分析處理 或資料探勘等分析工具的利用。

##### 2. 資料超市

資料超市 (Data Mart) 為一小型的資料倉儲系統，邏輯上為資料倉儲的一個子集合，其規模較資料倉儲為小，可能涵蓋範圍為某一特定部門或區域的主題範圍，且僅能提供該部門或區域的資料，並給予特定人員或部門使用，以符合實際運作需求，例如圖書館採訪部門在資料倉儲的應用，訂定與其業務相關主題為館藏採購與圖書出版社間的關係，因此所需的資料即包括館藏資料、出版社資料及讀者推薦資

料等所建構之資料超市，而非整個圖書館所需之資料倉儲（讀者資料、流通資料等）；但其建構技術與整個資料倉儲無異。

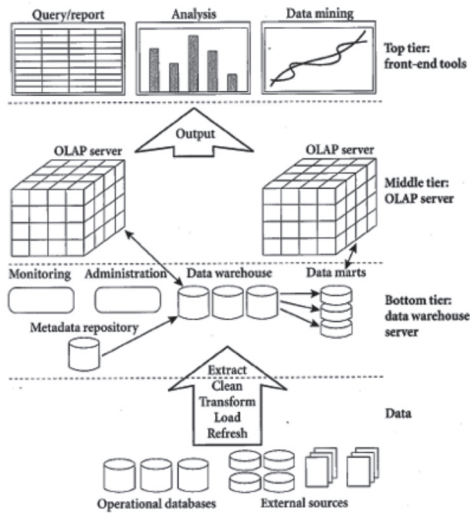


圖 1 資料倉儲架構 (Han, Kamber, & Pei, 2011)

### 3. 線上分析處理

線上分析處理 (Online Analytical Processing) 是針對以多維度資料模式 (Multi-dimensional Data Model, 或稱之為資料立方體, Cube) 所建立之資料倉儲資料進行分析。多維度資料模式是以主題事實資料表 (Fact Table) 為中心，加以組織彙整相關資料，此事實資料表包括事實資料表的名稱、計量 (measure) 項目，以及相關維度資料表 (dimension table) 之關鍵字 (key)，例如有關銷售紀錄之資料倉儲，定義名為銷售業績之事實資料表，可包含銷售款項、銷售數量等計量項目，以及其相關維度表之關鍵字：銷售員員工編號、產品編號、時間序號等，透過此事實表之組織整合相關資料，我們可以多維度分析探究：不同銷售員於各季銷售各項產品的數量、銷售總金額，不同產品於各月銷售數量、銷售總金額等有助於管理決策的資訊。

### 4. 維度資料模式

多維度資料模式是以所確定之主題為中心加以組織彙整相關資料來源，該主題以事實資料表表示。多維度資料模式可以是星狀綱要模式、雪花狀綱要模式或星座狀綱要模式的形式存在：

### (1) 星狀綱要 (Star Schema) 模式

此綱要之資料倉儲包含兩部分：一個包含大量資料的事實資料表，另有一組小的附屬表，稱之為維度資料表 (dimension table)，每個維度都有一個相對應的維度資料表。這種模式很像星星爆發，眾多的維度資料表環繞在中心表四周形輻射狀，如圖 2 所示。

### (2) 雪花狀綱要 (Snowflakes Schema) 模式

雪花狀綱要模式是星狀綱要模式的變形，其中的某些維度資料表經過正規化，為減少資料多餘情形，因而把維度資料表進一步分解成連結的兩個資料表，而資料綱要圖形就類似於雪花的形狀，如圖 3 所示。這樣的資料綱要表易於維護，並節省儲存空間，但是由於執行查詢需要較多的資料表結合 (table join) 動作，因此雪花狀綱要可能會降低瀏覽資料的效益。

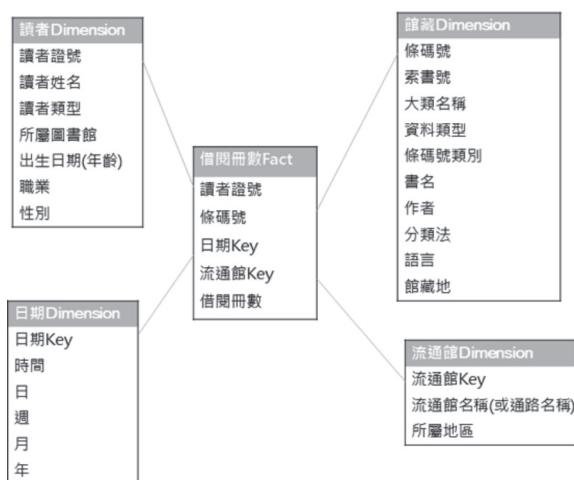


圖 2 星狀綱要模式

### (3) 星座狀綱要 (Constellation Schema) 模式

這類的綱要模式，可以視作星狀綱要模式的一群組合，因此又被稱為星系綱要模式 (Galaxy Schema)，此綱要存在許多事實資料表，這些事實資料表是共用維度資料表，如圖 4 所示。







## 二、Aleph Reporting Center (ARC)

ARC (Heather, 2007; Heather & Liya, 2009; Hurlbert, 2008) 是 ExLibris 公司針對其圖書館自動化系統 Aleph 所開發的報表系統，目前最新的版本為 4.0。ARC 為一基於 Web 的報表、查詢及統計資料產生環境，其資料來源為 Aleph 自動化系統，讓使用者可立即存取有關圖書館各種特定統計資料。ARC 發展自 IBM Cognos ReportNet 軟體，並與其整合構成系統環境。ARC 內之作業層資料是蒐集自從各種 Aleph 資料資源，並以邏輯結構呈現以使其得以於系統多個區域間互相參照。使用者可以操作作業層資料產生報表以及每天工作流程所需之統計資料。ARC 結合特定、整合資料之蒐集，依據圖書館歷史深度（時間年度）與不同主題層面進行資料組織。ARC 利用資料倉儲方法實現快速有效的資訊檢索，而所建構之資料倉儲為一獨立之資料庫，以固定模式自現有作業層系統擷取資訊，並以適合使用者直接存取的特定架構設計。ARC 系統概念如下圖 5 所示：

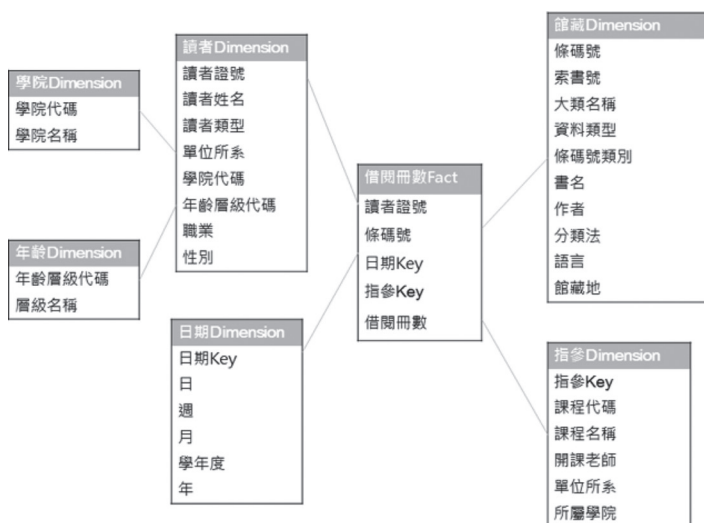


圖 5 ARC 系統概觀 (ExLibris, 2014)

ARC 是利用資料倉儲方法建構其資料庫架構，依據 ARC4.0 邏輯模式 (ARC 4.0 Logical Model) 資料顯示 (ExLibris, 2015)，其事實資料表設計多半是以星狀綱要模式為主，少數雪花狀綱要模式；資料倉儲內之資料直接來自於 Aleph 自動化系統資料表，均未經過任何整合

或前置處理，所提供之報表種類，對圖書館一般報表需求之支援有一定程度的助益。

然而，在今日需求多變的環境，圖書館亦因環境管理需求、管理文化變異等問題，對有助於提升服務品質或是增進作業效益之資訊需求亦更迭不斷，以至於圖書館相關管理資訊報表之需求亦隨之增加。圖書館員之資訊素養於台灣這幾年圖書資訊教育下大有進步，圖書館員均具備一定程度之電腦操作能力，與新資訊技術知識學習的能力。因此，對於圖書館報表需求因已成館員作業需求之一，是否提供一可操作之作業流程或是環境架構，以供圖書館館員自我產出所需之報表？此為本研究之主要議題與目標！本研究藉由實際演練過之歷程說明，提出解決上述議題之參考模式及流程步驟，以及基本知能要求，供國內圖書館館員建構並產生自我所需之報表。

## 參、研究方法

### 一、個案研究

本研究問題之起源，來自國內圖書館自動化系統廠商及大學圖書館對圖書館客製化報表之需求。依據該系統廠商過去維護與大學圖書館客製化報表需求彙整，共計有 93 種不同之報表類別，所涉及的相異資料欄位約有 50 項。針對此需求過去解決方式多數是由系統廠商技術支援產出，以維護費方式處理。但仍有以下幾個困難：

- (一) 客製化報表需求會因不同圖書館各組之業務或是管理方式不同而有所差異，且有經常變動之問題；
- (二) 系統廠商在成本壓力排擠下，客製化報表功能之完成時效性並不理想，經常性延遲交付；
- (三) 客製化報表功能在圖書館自動化系統資料之快速成長下，執行效率漸趨惡化，必須耗費系統更多資源；
- (四) 客製化報表功能執行期間與經常作業之系統功能間相互爭搶資源，致使圖書館自動化系統於上班時間難以正常運作（通常只能放棄即時報表產生）；
- (五) 圖書館經費拮据，較難建置如 ARC 報表產生系統，供圖書館館員運用。

另外，部分圖書館雖建置有類似於 ARC 之報表產生系統，但仍遇到以下的問題：

- (一) 類似 ARC 報表產生系統，對圖書館員並不容易學習，尤其是不具資料庫經驗的館員，非短時間就能駕輕就熟地開發客製化報表；
- (二) 此類報表系統多只適用於某廠系統預設報表產生，報表系統移植彈性差，投資成本效益不佳；
- (三) 此類報表系統仍無法完全滿足客製化報表需求。

藉由調查（以圖書資訊學研究所在職專班全體 25 位學生中，有 21 位來自全國大專校院圖書館及公共圖書館館員，以作業方式調查各館之報表需求）、並蒐集國內圖書館自動化系統廠商累積十餘年來自其客戶之報表需求及所面臨的主要困境，本研究試圖提出解決上述諸多問題的可行方案，期望能提供國內圖書館對客製化圖書館自動化系統報表需求一有效且經濟的解決之道。本研究提出之解決報表需求模式，為求其有效可行，透過館員及圖書資訊領域專家，針對所提解決方案模式進行實際演練，以驗證其效益。

## 二、系統架構

本研究探討議題之資訊架構，在設計上是以低成本高效益以及可建構在各校圖書館內為前提，提出以下之系統架構如圖 6 所示：

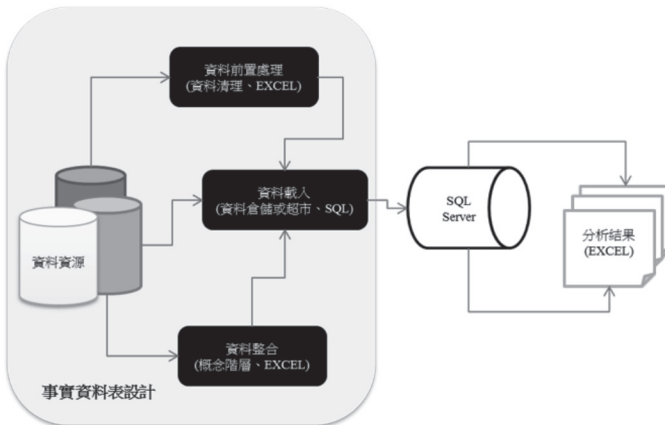


圖 6 本研究系統架構（本研究整理）

由以上系統架構圖可以看出，本研究環境以 SQL Server 及 EXCEL 為主，此兩軟體因微軟校園軟體方案故大多數圖書館並不需再額外花費成本建置。系統架構圖中，因資料倉儲是與圖書館自動化系統功能面切割，致使圖書館自動化系統未於圖中出現，僅以資料資源（data sources）呈現

經由圖書館自動化系統所儲存之資料或是外部其他資料來源。

### (一) 資料資源

資料資源為資料倉儲（或超市）建構所需之資料來源，本研究中因只考量圖書館自動化系統之報表，故資料資源以圖書館自動化系統所儲存之資料為主。但因圖書館自動化系統資料繁雜，不易於館員理解，故建議應先以各作業單位所需之資料為目標，建構各組作業所需之資料超市。

### (二) 事實資料表設計

事實資料表設計為資料倉儲核心，本研究提出四步驟完成：（1）選定主題，簡單說，所謂的主題就是你想產生何種報表？如借閱報表、館藏報表等；（2）單位量，也就是計量的最小單位，若以讀者借閱為例，本研究所選定的最小單位為借出之一筆館藏；（3）選擇維度（Dimension），這是件重要工作，所謂維度就是產生出來的報表資料以何種角度呈現，可以有多個，稱之為多維度，如讀者借閱報表，借閱館藏筆數呈現跟時間有關係，如日、週、月、季、學期、學年、日曆年等，此時就必須有一時間維度；（4）確認事實資料表（Fact Table），確定事實資料表中計算的量度（Measure），在讀者借閱的例子中，本研究在第（2）步驟中已選定最小單位為借出之一筆館藏，所以此例子中的量度皆為 1。下圖 7 為一館藏借閱報表之事實資料表（用以建構一資料超市）：



圖 7 借閱冊數事實資料表設計

此事實資料表之設計，可以產出以讀者、日期、館藏等不同角度借閱冊數之統計報表（其中借閱冊數為事實資料表之量度）；如下圖 8：不同年不同館藏資料類型之借閱數量情形。

Amount	欄標籤							
列標籤	中文書區	西文書區	特藏區	參考書區	期刊區	視聽資料區	圖書館	總計
87	13789	2	23	31				13845
88	19769	1	9	45				19824
89	11087	4	3	23				11117
90	1817	2	1	1				1821
91	11580	32	11	16		45		11684
92	14145	27	36	1		417	3	14629
93	14427	17	31	1		242		14718
94	14079	58	56	3		15		14211
95	23369	175	143	3		10		23700
96	23658	157	188			922		24925
97	26769	243	211			1959	3	29185
98	21954	470	198	1		1896	2	24521
99	15337	97	177			1588	1	17200
100	11530	194	110			1263	4	13101
101	13039	75	152			758	90	14114
102	3323	31	42			113	55	3564
103	98	1	1		1	2	1	104
<b>總計</b>	<b>239770</b>	<b>1586</b>	<b>1392</b>	<b>125</b>	<b>1</b>	<b>9230</b>	<b>159</b>	<b>252263</b>

圖 8 不同年不同館藏資料類型之借閱數量

### (三) 資料整合

架構中之資料整合是目前 ARC 系統不足的地方，因為它會牽涉各館不同報表需求而不同，ARC 無法完全產生客製化報表。如上圖 7 事實資料表日期維度中之「學年度」或是「週」，圖書館自動化系統借閱資料並未有學年度或是週的欄位，只有借閱時間（於流通櫃台出納的時間），因此無法產生各學年度或是各週借閱統計報表（當然，如果運用 SQL 查詢條件是可以解決此問題）。因此對於此需求，必須先將原有資料經過處理整合，以適合於從學年度或是週之維度呈現借閱資料。以「週」或是「學年度」之需求為例，可以透過建立時間概念階層（Concept Hierarchy）協助解決，時間概念階層即是時間階層化，如流通櫃台出納時間，可以記錄（彙整轉換）成：時間、日、週、月、季、年；亦或是：時間、日、月、學期、學年等時間階層；如此便可從不同學年度或是週檢視館藏借閱數量。

### (四) 分析結果 (EXCEL)

EXCEL 是一操作性極佳的線上分析處理 (OLAP) 工具，其樞紐分析功能與資料倉儲多維度概念一致，處理分層資料之解析彙整相當

便利；同時，EXCEL 資料繪圖工具亦是簡單方便，對於統計報表資料視覺化之需求，足以勝任有餘。EXCEL 2013 以後之版本，其最大列數（資料筆數）為 1,048,576 列，欄位最多為 16,384 欄，對多數圖書館以時間（月）為基準的報表需求，資料承載量應可以處理。因此在本研究中被視為資料倉儲結果輸出第一選擇之工具。

## 肆、研究結果

本研究在基於上述系統架構，首先提出一可行且成本效益考量之客製化圖書館報表作業之參考流程，並以實際個案說明；最後藉由圖書館員（專業館員及系統館員）及圖書資訊學研究生，針對圖書館實際需求，以本研究所提出之客製化報表模式實際建構完成所需之 EXCEL 報表，以分析驗證本研究之可行性與確實效益。

### 一、圖書館自動化系統客製化報表建構模式

以資料倉儲技術客製化圖書館報表，其基本仍是以資料倉儲為主體，故本研究所提出之參考流程部分參考現有資料探勘，並作必要調整以適用於圖書館運用，圖書館自動化系統客製化報表建構模式（資料倉儲應用模式）如下（圖 9）：

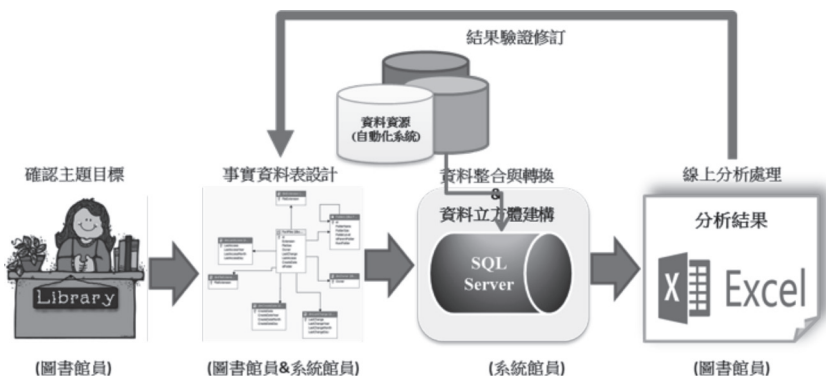


圖 9 圖書館自動化系統製化報表建構模式（資料倉儲應用模式）

#### （一）確認主題目標

整個作業流程第一件事情就是要確認主題目標，簡單的說就是要產生甚麼樣的報表。例如，典閱組需要流通統計資料（做甚麼用？要



想清楚），目的是「為了進一步分析學校各單位不同身分借閱圖書大類之情形，以作為圖書採購或是推薦之參考」。這個主題目標與目的很明確，圖書館員對此問題陳述要能夠越精準越有利於後續設計。

## （二）設計事實資料表

事實資料表之設計主要是依據前述之目的，首先確定計量為何？再者確定檢視的維度；爾後依照事實資料表之設計，由作業階層資料庫擷取所需之與計量相關之維度資料，以建構特定目的之資料超市。延續以上流通統計資料案例，計量為借閱筆數；而維度包括學期時間、圖書館藏（大類）、讀者身分（教職員工，研究生或是大學部等）等，共有三個維度。其中系所單位可以進一步探討，以教學單位為例，其所屬學院。計量事實資料表及其維度資料表關聯構成如下圖 10 之雪花狀綱要模式。

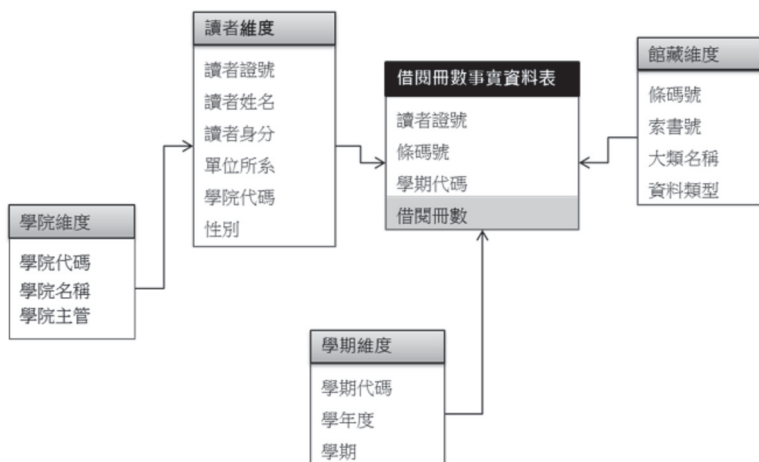


圖 10 流通統計事實資料表雪花狀綱要模式

## （三）資料整合與轉換

資料整合著重於領域知識的運用，為客製化重要一環，亦為如 ARC 類之報表系統較難支援者。其中概念階層（concept hierarchy）之建立為資料整合之關鍵知識。如上圖 10 學期維度之學年度與學期，圖書館自動化系統所登錄之借閱時間，是以流通櫃台之出納時間為準，並非以學年、學期、季等值記錄，同時出納時間在應用上亦須經過處理才能分出年、月、週（日曆週）、天、上午、下午、晚上等值。若



時間維度有此檢視資料之需求，就必須在資料庫資料擷取之前先建構時間概念階層：時段（上午、下午、晚上）、天、星期、日曆週或月、季、年；或是時段（上午、下午、晚上）、天、星期、日曆週、學期、學年等，端視檢視資料之需求設計。因各館各校學期或是每年之日曆週定義不同，故於資料庫轉出處理上就較為繁瑣複雜，這也是必須客製化的原因之一。上述流通統計學期維度檢視資料以學年及學期為主，故在資料庫轉出時必須將櫃台出納時間依據學校學年學期時間，另外改以新增學年、學期欄位登錄資料，如此才能以不同學年學期彙整分析流通借閱資料。

依前所完成之事實資料表、維度資料表、整合轉換新增資料欄位等建立所對應之資料表；然後以資料擷取轉換載入工具（Extraction, transformation and loading, ETL tools）（Kimball & Ross, 2013）由作業層資料庫將所需之資料截取轉換載入至所建構之相對應資料表；抑或是直接以 SQL 結構化查詢語言處理資料載入相關工作。此步驟為未具資料庫知能之圖書館員較多障礙之步驟，初期建議可由圖書館系統館員協助，館員可於初期過程學習資料庫觀念與 SQL 語法，以為後續自我建構備用。ETL 工具通常需有一定的學習曲線及額外軟體費用，且多半只能針對所指定之資料庫管理系統，因此建議以 SQL 語言是最直接經濟有效的方法。

#### （四）資料立方體建構

於所選定之資料庫管理系統（SQL Server）提供之工具（Microsoft Visual Studio）或是支援資料倉儲（資料探勘）建構工具，根據所設計之上述資料表（事實資料表與維度資料表）間之關聯關係，宣告建構資料立方體，也就是所謂之資料倉儲。

#### （五）線上分析處理 (online analytical processing, OLAP)

線上分析處理即為主要之使用者介面工具，能夠支援多維度檢視、分析、彙整目的資料，然其與 ETL 有類似之上述問題，因此在考量不增加成本前提下，本研究以 EXCEL 為推薦之 OLAP 工具，其樞紐分析（pivot）功能正與多維度資料檢視分析概念契合且操作容易，是相當適合於支援 OLAP 之使用者介面工具。當館員或是管理者於線上分析處理過程中，若發現結果並不符合需求，須修正原有設計，則回至（二）進行新增、修改或刪除原事實資料表設計內容後重新執行。

## 二、客製化報表模式可行性及效益驗證實驗

以下本研究即以實際案例之操作演練實驗，作為驗證上述所提步驟於圖書館應用之可行性及效益之驗證。實驗之進行包括圖書館典閱組組長、系統館員以及圖書資訊學研究所研究生各一員，共計三人；其中系統館員具資料庫概念、SQL 語言撰寫能力（曾修習過資料庫相關課程）以及 SQL Server 實務操作能力；而圖書資訊學研究生為提供技術諮詢，除修習過資料庫與 SQL 語言課程外，亦曾選修實作資料倉儲案例者。

### （一）確認主題目標

典閱組組長為充分了解讀者借閱相關訊息，期望能依據需求客製化借閱資料報表，以訪談引導方式瞭解需求：

1. 主要產生何種資訊報表？統計人次或館藏量？（主題為何）
2. 以何種角度檢視報表資訊？有何？（維度分析）
  - a. 以讀者角度檢視報表資訊，讀者需要分類？有何？（維度屬性）
  - b. 以時間角度檢視報表資訊，有哪些不同時間？（維度屬性）
  - c. 以館藏角度檢視報表資訊，有哪些館藏相關因素？（維度屬性）

紀錄訪談結果整理後，再與系統館員與圖資所研究生討論確認主題，列出如下所述維度需求：（約經 65 分鐘討論）

1. 依讀者之性別、系所、身份分析借閱資料；
2. 依據館藏類號、館藏地、館藏類型分析借閱資料；
3. 依據學年度、月、季、年分析借閱資料；

### （二）設計事實資料表

隨後由系統館員及圖資所研究生依據上述需求繪製初步事實資料表與維度資料表如下圖 11 之左所示；再與典閱組組長討論其需求並確定設計是否符合。討論過程中，對於以館藏館藏地分析資料之需求有所質疑，館藏地因該館可以跨館借閱且跨館館藏，館藏地為一變動項目且並非分析館藏借閱之重點；反倒是分析了解讀者之借閱於何處借閱的較為重要，因此修改原設計事實資料表，新增流通地維度以解決此需求之改變，結果如圖 11 之右所示。（約討論 75 分鐘後確定）

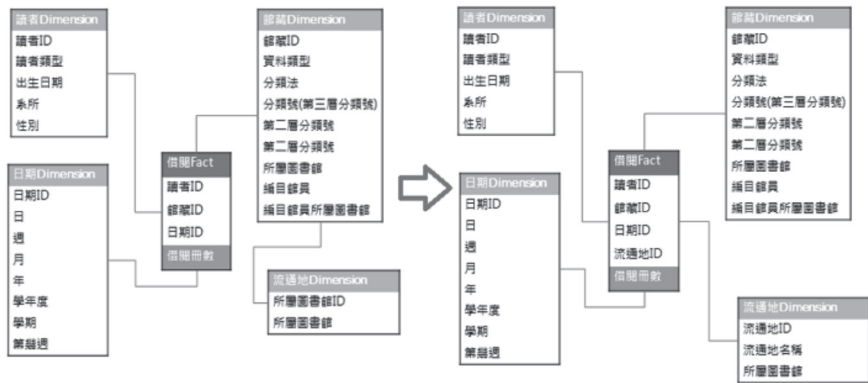


圖 11 案例事實資料表設計

### (三) 資料整合與轉換

對於借閱資料期望以時間軸檢視分析資料，可以學期或月或季等單位彙整資料，因此對時間維度必須作此轉換：將出納時間對應制適當的時間軸單位，自系統開始運作時間至蒐集資料之最後時間；為求資料處理之效益，由典閱組組長依其對時間軸單位需求，以 EXCEL 建立每年每日對應之，如下 EXCEL 所示：

date_ID	date	week	month	academic_year	semester	year	week_number
1	1998-10-30	Friday	10	87	1	1998	44
2	1998-10-31	Saturday	10	87	1	1998	44
3	1998-11-01	Sunday	11	87	1	1998	45
4	1998-11-02	Monday	11	87	1	1998	45
5	1998-11-03	Tuesday	11	87	1	1998	45
6	1998-11-04	Wednesday	11	87	1	1998	45
7	1998-11-05	Thursday	11	87	1	1998	45
8	1998-11-06	Friday	11	87	1	1998	45
9	1998-11-07	Saturday	11	87	1	1998	45
10	1998-11-08	Sunday	11	87	1	1998	46
11	1998-11-09	Monday	11	87	1	1998	46
12	1998-11-10	Tuesday	11	87	1	1998	46
13	1998-11-11	Wednesday	11	87	1	1998	46
14	1998-11-12	Thursday	11	87	1	1998	46
15	1998-11-13	Friday	11	87	1	1998	46
16	1998-11-14	Saturday	11	87	1	1998	46
17	1998-11-15	Sunday	11	87	1	1998	47
18	1998-11-16	Monday	11	87	1	1998	47
19	1998-11-17	Tuesday	11	87	1	1998	47
20	1998-11-18	Wednesday	11	87	1	1998	47

圖 12 EXCEL 處理日期時間單位

以 EXCEL 產生對應之時間軸單位後，再轉入隨後於 SQL Server 所建立之時間維度資料表中。接著由系統館員（或圖資所研究生）依據上述確定之事實資料表與維度資料表，於 SQL Server 建置相對應之事

實資料表與維度資料表（新增一資料庫後），及資料表間之關聯。然後利用 SQL 語言將自動化系統中相關欄位資料匯入各資料表中；此步驟除須具備資料庫技術能力外，另一重要關鍵在於資料正確性之確定，即匯入所建構資料表之資料必須謹慎確定為正確所需之資料，此項工作必須由系統館員與自動化廠商系統工程師共同確認無誤。

#### （四）資料立方體（資料倉儲）建構

本研究之資料倉儲是藉由 Microsoft Visual Studio 2013 所建構，在新增一 Analysis Service 多維度和資料採礦專案之後，與 SQL Server 資料庫連線，並選擇先前所建立的資料庫做為資料來源（如下圖 13 所示）：圖右方的「方案總管」區塊，可設定資料來源與欲分析的資料庫與資料表連線，並新增 Cube 與其維度，設定的內容會顯示在圖中間和左方的區塊中。

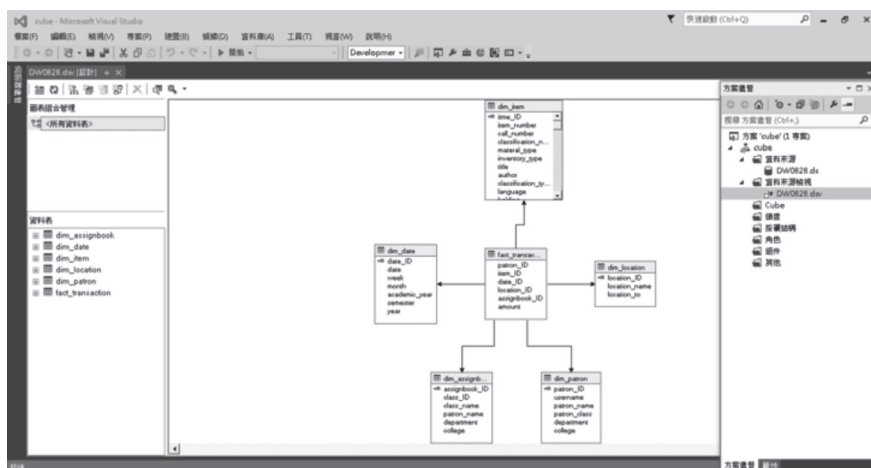


圖 13 資料立方體（資料倉儲）建構

接著新增建立一資料立方體（Cube）並依序設定相關之各維度與其屬性，結果如下圖 14 所示。最後執行部署即完成資料立方體（資料倉儲）之建構：圖 14 為 Cube 維度的設定與檢視，圖右方「資料來源檢視」的區塊中，為原始 SQL Server 資料庫中資料表的欄位，圖左方的「屬性」區塊為已設定加入 Cube 的欄位，若要新增 Cube 的欄位，將欲分析的欄位拖曳至左方屬性的區塊中即可。

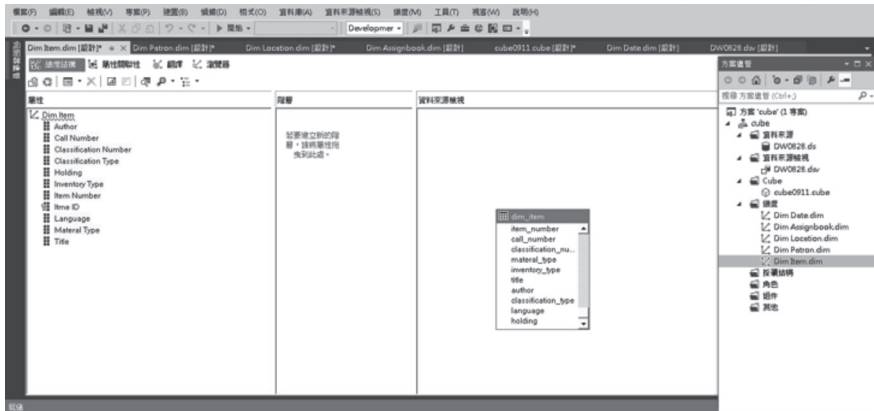


圖 14 案例部屬完成

### (五) 線上分析處理

資料線上分析處理以 EXCEL 由圖書館館員操作，開啟已部署好之 Cube 的瀏覽器功能，匯出成 EXCEL 檔，要以樞紐分析表呈現，則將欲分析之資料欄位拖曳至檢視位置，得到 EXCEL 樞紐分析結果，如下圖 15 所示是以各學年度檢視不同類型館藏之借閱數量：圖右方可選擇欲分析或產生報表的欄位值，拖曳至右下方的「欄」或「列」區塊中，並將欲分析的數值欄位拖曳至「值」的區塊中，將欄位拖曳至「篩選」區塊，可限縮資料的範圍，結果呈現於圖左方 EXCEL 樞紐分析中。

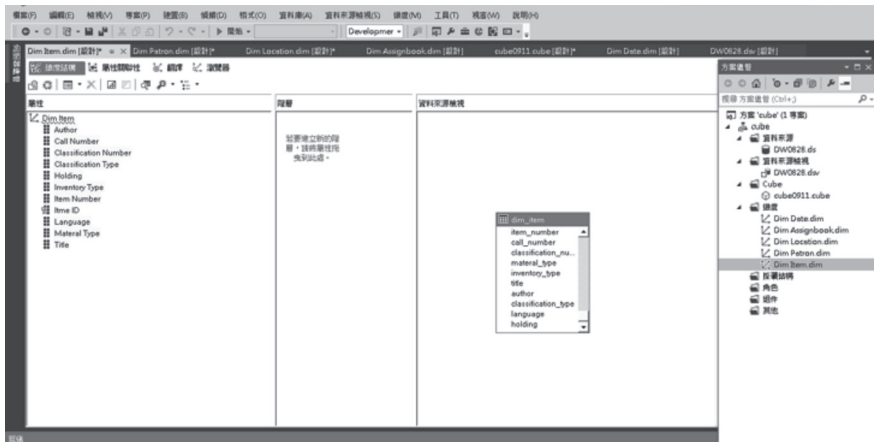


圖 15 EXCEL 樞紐分析（線上分析處理）

後續相關所需之客製化報表，則以同樣方式處理產生 EXCEL 檔

案，輸出統計分析數字或是繪製圖表等圖書館作業管理客製化報表。

### 三、模式驗證實驗討論

以資料倉儲技術實踐圖書館彈性客製化報表產生是本研究主要目的，在考量目前圖書館人力資源侷限及經費捉襟見肘的困境下，提出此客製化報表模式。本研究於其可行性及效益驗證實驗，以一 3 人參與之圖書館案例說明，亦是在此考量下進行，以最少人力評估其可行性。

資料倉儲之應用，向來不是資訊技術問題，因其主題導向特性，向來是專業領域需求問題。以同樣之模式，亦進行其他 2 類需求報表之開發，如下表 1 之資料整理。其中訪談引導對象除本文所列之典閱組組長外，另有圖書館系統館員及系統公司 PM（專案經理），可發現系統公司專案經理在專業領域（圖書館、自動化系統、資料庫）之知能較優於圖書館員，故其訪談引導事實資料表需求時間較短；然所列三個案例之報表開發時間，以同一組系統館員及圖書資訊學研究所研究生進行，報表開發時間均優於過往圖書館尋求自動化系統廠商開發所耗費之時間。

表 1  
需求報表開發比較

需求主題	訪談引導對象	圖書館知能	資料庫知能	訪談引導(分鐘)	維度資料表建構(分鐘)	事實資料表建構(分鐘)	資料倉儲建構(分鐘)
讀者借閱圖書	典閱組組長	圖書資訊碩士	無	65	300+480+360+60=1200	90	30
讀者預約圖書	圖書館系統館員	資訊工程碩士	資料庫設計經驗	67	維度同上無需再建構	60	30
讀者續借圖書	系統公司 PM	圖書資訊碩士	資料庫設計經驗	50	維度同上無需再建構	60	30

開發需求報表花費時間最多的是在於維度資料表的建構，此個案是以國內自動化系統廠商所提供之某大專校院圖書館 1998 年至 2014 年之資料為主，共計 252200 筆借閱資料，館藏資料 178605 筆，讀者資料 21923 人，在 CPU 為 i5，記憶體 8G 的個人電腦下，花費約 20 小時（1200 分鐘）完成所需之維度資料表建構（共有四個維度資料表，表格中之數字 300、480、360、60 為其各別建構所需之時間）。各資料表建構所需之時間，與所涉及之資料量有絕對關係，而開發之電腦



系統環境效能的影響亦不可忽略。

圖書館員對資料倉儲或是資料探勘技術之應用，經常擔心其高門檻之技術問題，然事實並非如此，本研究所提出之模式即是要破除此迷思。圖書館館員只要專注於需求報表的正確完整陳述（此項工作並不容易，亦為資料倉儲應用於圖書館困難之所在），且所必要之資料確實存在自動化系統資料庫中，後續相關資料庫、資料倉儲建構等，均可委由具資料庫開發、資料倉儲建構、SQL Server 應用經驗之系統館員接手處理。

## 伍、結論與建議

圖書館客製化報表需求一直以來是圖書館亟欲解決的問題，縱然今日有如 ARC 報表系統可供圖書館運用，但因其專屬性為特定圖書館系統獨用、利用資料倉儲技術建構預設可能報表缺乏彈性、館員學習不易且若無資料庫知能實作開發仍需系統館員協助、圖書館需額外付費此系統功能其成本效益不彰等因素，圖書館客製化報表需求仍舊為今日圖書館至為棘手之困境。本研究以國內圖書館自動化系統廠商累積多年來自用戶之報表需求為基礎，提出以 SQL Server 與 EXCEL 為環境之具成本效益解決模式，藉由資料倉儲技術協助圖書館員客製化需求報表之發展；並以實際圖書館案例實證其可行性。

於圖書館客製化報表案例實證過程發現，對於參與實驗之圖書館專業館員，所具備之資料庫素養，包括概念、設計、SQL 等仍嫌不足，尚需圖書館系統館員多所指導協助，此為圖書館爾後可以加強館員訓練與強化之內容。同時，因本研究所提出之建構模式，除資料擷取轉換需有程度以上之 SQL 查詢語言能力外，其餘皆為例行性操作動作且差異不大（僅資料欄位名稱或是資料表名稱不同），故圖書館自動化系統管理者，可以考慮此模式建立一標準作業流程規範及模板或模組，供圖書館專業館員更有效率之運用，以發展個別客製化需求之資料超市或報表。

圖書館客製化報表需求可以資料超市建構模式解決圖書館館內各組需求，若進一步整合全館各組所建構之資料超市，逐步形塑圖書館整體客製化需求報表之資料倉儲；後續可以進一步考量以此資料倉儲為基礎，發展建構圖書館之管理資訊系統（Management Information System, MIS）或高階主管資訊系統（Executive Information System, EIS），以協助提供圖書館管理者決策支援所需之訊息。



## 參考文獻

- 謝建成 (2009)。書目探勘。台北市：文華圖書資訊。
- 【Shieh, J. C. (2009). *Bibliomining*. Taipei: Mandarin Library & Information Service.】
- Carlos, C. & Morris, S. (2016). *Database Systems: Design, Implementation, & Management*(12th ed.). Course Technology.
- ExLibris. (2014). *ARC 4.0 User's Guide*. Available at [https://knowledge.exlibrisgroup.com/Aleph/Product\\_Documentation/Aleph\\_Reporting\\_Center/Version\\_4.0](https://knowledge.exlibrisgroup.com/Aleph/Product_Documentation/Aleph_Reporting_Center/Version_4.0).
- ExLibris. (2015). *ARC 4.0 Logic Model*. Available at [https://knowledge.exlibrisgroup.com/Aleph/Product\\_Documentation/Aleph\\_Reporting\\_Center/Version\\_4.0](https://knowledge.exlibrisgroup.com/Aleph/Product_Documentation/Aleph_Reporting_Center/Version_4.0).
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*(3rd ed.). Morgan Kaufmann.
- Heather, H. & Liya, A. (2009). *ARC 2.0: What's new in the Aleph Reporting Center*. Paper presented at ELUNA 2009, 6-8 May, Richmond, VA.
- Heather, H. (2007). *The Aleph Reporting Center: Hype or Help?* Paper presented at ELUNA 2007, 5-8 June, Spearfish, SD.
- Hurlbert, J. M. (2008). *Defining Relevancy: Managing the New Academic Library*, Libraries Unlimited.
- Inmon, W. H. (2008). *Building the Data Warehouse*(3rd ed.). Wiley.
- Kimball, R. & Ross, M. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*(3rd ed.). Wiley.
- Massis, B. (2016). Data-driven decision-making in the library. *New Library World*, 117(1/2), 131-134.
- Nicholson, S. (2003). The bibliomining process: Data warehousing and data mining for library decision making. *Information Technology and Libraries*, 22(4), 146-151.
- Nicholson, S. (2006). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing and Management*, 42, 785-804.
- Shieh, J. C. (2010). The integration system for librarians' bibliomining. *The Electronic Library*, 28(5), 709-721.

Wah, T. Y., Peng, N. H. & Hok, C. S. (2007). Building Data Warehouse. *Proceedings of the 24th South East Asia Regional Computer Conference*, Bangkok, Thailand.

# The Study of Data Warehouse Application Model in Libraries A Case Study of Customized Reports

**Yi-Ting Yang**

Ph.D. Student

Graduate Institute of Library and Information Studies  
National Taiwan Normal University

**Jiann-Cherng Shieh**

Professor

Graduate Institute of Library and Information Studies  
National Taiwan Normal University

The development of library customized reports has always been one of the intractable problems that library managers urgently seek to solve. To date, library automation system manufacturers are unable to provide cost-effective solutions. Most of the previous solutions to library customized reports were based on case-by-case projects. For Taiwan libraries using domestic library automation system, their demands of library customized reports are usually fulfilled by system manufacturers through maintaining system or adding new function module. For those libraries deploying foreign library automation system, the solutions for the library customized report demands are either by developing specific project with extremely expensive cost or by requesting the system manufacturers to add new features in the following version updates. In recent years, ExLibris has developed Aleph Reporting Center (ARC), an add-on system for its library automation system Aleph, to provide more than 100 pre-defined statistical reports for those Aleph customers. However, due to this reporting system only for Aleph, ARC is useless for most Taiwan libraries whose library automation systems are not Aleph. Furthermore, pre-defined reports system can not fulfill the real-time and diverse demands of different libraries. Additionally, in such an uncertain environment today, libraries' information needs for improving service quality or operating efficiency change with the needs of environment management and the diversities of management

culture. Therefore, the demands of library management reports increase. Due to the consideration of cost-effectiveness and the need of customization, this study proposes a customized report generating mode of library automation system based on data warehouse. Under the premise that system librarians have reasonable background knowledge about database, a practical case study supports that this module is feasible and efficacy.

Data warehouse is different from operational database that keeps changing. In order to help organization to make decisions, data warehouse is composited of those data extracted from the database according to the results that organization analyzes its specific demands. Because one characteristic of data warehouse is data integration, for those data from various systems, data warehouse is a robust platform to integrate system in an integrative infrastructure and manage long term data effectively. After collecting and investigating the demands received by domestic library automation system manufactures in more than ten years, this study proposes a system architecture base on SQL Server and EXCEL to design the fact table, integrate data, and analyze with EXCEL in online environment. The design principles of this system architecture are under the premise of low cost, high efficacy, and feasibility in different school libraries.

The customized report generating mode of library automation system proposed by this study is as follows:

- (1) Confirm themes and objective. Make sure what kind of reports should be created.
- (2) Design fact table. Determine measures and dimensions based on themes and objectives. For example, measure is the number of borrowings and dimensions include semester, library collections, and readership etc. Then, according to the measures and dimensions, extract data from operational database to build data marts for specific purpose.
- (3) Integrate and transform data. Emphasize the application of domain knowledge, e.g. the formation of concept hierarchy. This will be a critical step for customization.
- (4) Create data cube. Based on the fact table and dimension table, create data cube, viz. data warehouse.
- (5) Analyze data online. Recommend Excel' s PivotTable as the tool to view, analysis, and integrate data.

This study uses a practical case as an actual operating experiment. Three subjects are a library collections director, a system librarian, and a graduate student of library and information science. System librarian, had studied courses about database, has concepts about database, ability to use SQL language, and capability to operate SQL server. A graduate student who had studied courses about database and SQL language as well as has experience in creating data warehouse provide technical assistance. The practical case provided by domestic library automation system manufacture is the data of a university library from 1998 through 2014, including 252,200 borrowings and 21,923 readers. The experimental results are as follows. It took 65 minutes to confirm themes and objective. Designing fact table required 90 minutes to assemble. It took about 20 hours to complete four dimension tables when using personal computer with i-5 CPU and 8G RAM. Creating the data cube based on demand required 30 minutes. The result supports the feasibility and efficacy of the library data warehouse application mode proposed by this study.

For library customized report demands from different department in the library, creating data marts mode can be an available solution. One step further, by integrating all data marts created by every department in the library, a data warehouse for fulfilling the library-wide customized report demands will be created step by step. In the future, based on this data warehouse, the development of management information system or senior administrator information systems for library can be took into consideration in order to provide library administrators with information for decision making support.

