

淺談「研究資料管理」(Research Data Management)

魏令芳與李明錦

「資料管理」(data management)與「研究資料管理」(research data management, 簡稱RDM)是當前圖書資訊領域相當熱門議題之一, 觀察歐美國家中的學術與研究型圖書館, 其利用指導課程或是學術工作坊, 也經常聚焦於資料管理主題, 特別是需要申請政府經費的研究機構與大學, 紛紛為研究者提供所需的資料管理知識與技能。雖然國內相關單位, 對於研究者申請研究計畫時, 尚未提出類似要求, 但對於圖書館而言, 「資料管理」與「研究資料管理」是一個重新設計利用指導及推廣數位素養的機會, 我們希望未來在圖書館的利用指導課程, 將不再侷限於資料庫的操作使用, 而轉為學術研究過程中的協助與支援, 期能達成開放資料後的資源共享目標。

OCLC於2017至2018年兩年間, 以RDM為題, 陸續發表了四份連續性報告, 分別探討RDM的背景脈絡、影響與研究型大學所需要建立與獲取的RDM能力。簡摘內容如下:

1. Part One: A Tour of the Research Data Management (RDM) Service Space

RDM的服務種類與架構分為教育(education)、專業(expertise)與度用(curation)三方面。教育架構指的是教導研究者察覺研究資料管理的重要性、鼓勵建立RDM技能與發掘管理的工具與資源; 專業架構方面指的是提供決策支援與客製化方案, 以具體解決研究者的資料管理問題; 度用架構方面, 則是瞭解支援資料管理的技術基礎架構, 期望於研究者的研究過程中善盡支援與服務。綜觀RDM服務, 的確啟發了一個大學圖書館可以提供多元新服務的面向。

2. Part Two: Scoping the University RDM Service Bundle

探討四間研究型大學在四個不同國家背景下獲得RDM的能力本質, 並強調如何塑造此一能力的13項關鍵因素, 啟動機構內的RDM服務。這四間大學分別是英國的University of Edinburgh、美國的University of Illinois at Urbana-Champaign、澳洲的Monash University與荷蘭的Wageningen University & Research, 報告中特別分析這四所大學在面對內外環境、制度要求與因地制宜選擇等複雜交錯因素下, 如何發展自己的RDM服務。

3. Part Three: Incentives for Building University RDM Services

作者根據上述研究案例中, 指出促使大學進行RDM服務的內部與外部動機, 分為四大項。首先是承諾(compliance), 來自贊助者與相關利益關係人的要求; 其次是持續演變的學術規範(evolving scholarly norms), 面對開放科學下持續發展的資料分享與資料再次利用; 第三項是機構策略(institutional strategy), 指

的是許多研究型大學將RDM視為重要的校務發展策略；第四項則是研究需求（research demand），除了大學由上至下的發展策略外，還有由下至上來自研究者研究過程中的需求。

4. Part Four: Sourcing and Scaling University RDM Services

最後一份報告重點在如何獲取RDM服務能力，以及獲得RDM服務能力的來源與選擇，例如相關的資料管理，如何在內部構建和維護，亦或是透過外部供應商獲得？如何支援這些服務，金錢、時間與人力都要考量，決策中重要的變數考量則包括了：效率/成本降低？可靠性/信任？這份報告聚焦在上述四所研究型大學如何支援建立RDM能力，其中包含如何使用外部典藏庫做資料分享，如University of Edinburgh將部分RDM外包給英國的Digital Curation Center。

對館員而言，最重要的是認識RDM，其次是獲取RDM的能力與技能，歐美國家之大學與圖書館所建立的資料，實務資源非常豐富，十分值得借鏡參考。藉由OCLC所出版的RDM研究報告引導下，就此來翻閱一些扼要的重點，複習一下熱門但又不甚熟悉的研究資料管理議題。

認識資料管理

有幾個不錯的自學資源可以參考：一是2016年MIT的開放課程「資料管理」（data management）課程，分由研究概念（study concept）、資料蒐集（data collection）、資料處理（data processing）、資料傳遞（data distribution）、資料探索（data discovery）與資料分析（data analysis）來建立研究資料的知識與技能，很值得研究者與館員一齊學習。另外，Purdue大學與Harvard大學於2015年舉辦的工作坊中也有一份詳細的PPT訓練資料，其中都介紹了加州大學數位圖書館度用中心的資料管理計畫工具-DMP Tools，並有一些計畫的實務樣版可以參考，哪些資料會被研究者產出？資料如何標準化、組織、標籤、儲存、典藏、分享、傳遞、保全等，即Data Management Plan如何計畫、描述資料、文件化、組織與儲存，進一步的數具存取、分享與再次利用問題及後續的典藏等不同面向，資料都非常詳實扼要。

其次是2017年Coursera開放課程「研究資料管理與分享」（Research Data Management and Sharing），由北卡羅來納大學教堂山分校與愛丁堡大學合作製作的，這門課共有五週進度，第一週主要是認識研究資料，說明其定義、生命週期、為何需要管理資料、相關利益關係人與其重要性。第二週進入計畫過程，介紹DMP、贊助者需求、DMP內容與工具。第三週說明研究過程的檔案管理，檔案如何命名、區分版本、組織資料、檔案格式、資料轉換、文件化、資料引用、儲存、備份、保全、加密、並談談好的檔案管理如何為研究加值。第四週談資料分享，分享資料的好處與挑戰、資料引用、保密、智慧財產權與擁有權。第五週是資料典藏，為何需要典藏、如何做到真實性與完整性、詮釋資料（metadata）、資料度用的標準與實例、指出典藏資料與數位保存的重要性。目前國內圖書資訊有關研究與課程，國立臺灣大學的鄭瑋

老師專長於此，並開授有「研究資料基礎架構」(Research Data Infrastructure) 課程，亦可經常追蹤更新。

「研究資料管理」的定義與範圍

資料與資料類型為何？根據美國國家科學基金會(National Science Foundation, 簡稱NSF) 指出，資料的界定由研究興趣社群經過同儕審核與計畫管理過程予以定之，不限於資料、出版品、樣本、實體館藏、軟體與模組；英國政府則指出是一些量化與質化的描述或是資料，因此無論是由實驗產出、統計而得、或是經由其他研究過程而來者，都是研究資料，所以資料對圖書館而言，已經延伸到研究者研究過程中的所有產出，而不再侷限於圖書館內的館藏。資料有結構化與非結構化資料、有文字與語意型的資料、或是天文學或化學主題方面的視覺化資料等，有簡單的也有複雜的。所以管理資料是長期支援資料的保存、存取與使用的活動與實踐，研究過程中持續進行的活動，因為要確保資料正確與完整，資料方能度用。

而圖書館界探討的資料管理服務，據《SPEC Kit 334：Research Data Management Services》中的說明：一種提供資訊、諮詢、培訓服務，或積極參與資料管理計畫、研究過程中的資料管理指導、研究文檔與後設資料的建制、對已完成的研究專案及出版資料的共用與度用等相關的服務(轉引自王梅玲翻譯)。由於研究過程中會產生各種資料，然這些資料絕不限於圖書館透過採購、交換與贈送途徑而得的館藏，研究過程中的產出，是圖書館試著延伸服務的地方。

管理資料

瞭解資料類型後，再來就是談談如何管理資料了。

1. 資料度用 (data curation)

強調的是數位資料的管理，根據《SPEC Kit 354：Data Curation》說明，資料度用是資料產生過程中對資料主動與持續的管理，對於學術與教育活動很重要，以學術圖書館的角色而言，會因此調整工作與改變服務流程與活動，或承擔工作更多資料管理責任。因此資料度用是學術圖書館發展的趨勢之一。這一期的SPEC Kit報告中，可以參考的資訊很多，羅列如下：

A. 機構典藏方面

British Columbia Research Libraries所提供的Data Services，其中有關Abacus Dataverse Network等的描述；

B. 基礎架構

如University of Notre 的CurateND Data Curation Infrastructure；

C. 工作流程

可參考Cornell University Library的eCommons Institutional Repository Curation Workflow等；

D. 資料模式與有關詮釋資料

可參考Virginia Tech University Libraries的EZID Quick Start Guide等；

- E. 資料存取政策
可參考University of Michigan Library的Deep Blue Data等；
 - F. 工作職責
可參考John Hopkins University Libraries的Data Services Manager等；
 - G. 服務方面
可參考University of California, Irvine Libraries的Digital Scholarship等。
2. 資料保存、典藏與儲存 (data preservation, archiving, and storage)
當前資料管理最大挑戰在於資料的保存與儲存上面,需要先解決此一問題,方能進行後續複雜的研究分析與資料的分享與重複利用。許多資料雖然都可以雲端方式存放起來,但資料格式不一,有些是需要使用端的軟體工具或是硬體設備才能使用,因此資料標準的建立亦顯重要。不同學科領域有不同的資料標準。
 3. 引用資料 (citing data)
要進行資料的描述,包含知道資料狀態 (status of data)、屬性 (attribution)、持久性 (persistence)、存取 (access)、探索或挖掘 (discovery)、出處 (provenance)、範圍或層次 (granularity)、可驗證性 (verifiability)、詮釋資料標準 (metadata standards)與資料使用上的彈性 (flexibility)。

開放分享

許多資料使用雲端硬碟分享,如Dropbox、OneDrive、Google Drive與Box。許多研究組織與大型單位為進行專案管理,訂購如BaseCamp軟體;另外,科學線上合作平台如colwiz或是Zoho,也是很熱門的管理資料與研究檔案合作平台。進行資料管理與資料分享時,會因學科與專業不同而有其文化差異,有些個案會以單一的中央資料庫作管理與分享,有些個案則於資料出版後方進行分享。

根據2015年Baykoucheva的論著,書中指出:主題為生命科學者,約有66%會將資料分享;主題為社會與人文科學者,則約有36%會將資料分享。不同學門領域對於開放資料分享的比率不同,而對研究者而言,在探討資料分享此議題上,需要深入探討的議題是法律問題、資料的誤用與資料相容的問題。目前在美國許多資料使用沒有版權問題,而對於資料使用上的版權問題,各個國家的狀況也不盡相同,但這些都是研究者在出版資料前必須瞭解的有關權利。

資料典藏庫與相關組織

許多政府、專業組織或是商業公司會贊助許多資料典藏活動,特別是歐美國家有關組織或單位日益增多;其次,一些專業的機構典藏都會為其學門設計適用的詮釋資料,以進行資料的描述。Baykoucheva的書中提到有20個組織(羅列如下),可為研究者提供許多相關標準與國際性的資料典藏中心,有助於深入瞭解資料典藏相關議題與提高研究資料的能見度。

1. **The Board on Research Data and Information**
(<http://sites.nationalacademies.org/PGA/brdi/index.htm>)
目標是提升科學與社會數位資料的使用。
2. **The Coalition for Networked Information** (<http://www.cni.org>)
出版有關eScience與資料管理的文章。
3. **The Committee on Data for Science and Technology** (<http://www.codata.org>)
致力於提升科學與技術資料的品質、管理與取得。
4. **DataCite** (<http://datacite.org>)
由加州數位圖書館、Purdue大學圖書館能源部的科學技術辦公室共同合作，提供資料登記與識別DOI。
5. **Databib** (<http://www.re3data.org/>)
研究資料典藏的目錄、註冊與指引目錄。有助於指出與識別研究資料典藏機構。
6. **Digital Curation Center** (<http://www.dcc.ac.uk>)
位於英國境內國際最著名的資料管理中心。
7. **The Data Observation Network for Earth** (<http://www.dataone.org>)
NSF支援，提供全世界環境生態方面的資料典藏。
8. **The Data Net Project** (US NSF，**查無URL**)
NSF的跨領域研究的聚集性資源。
9. **The Dataverse Project** (<http://thedata.org>)
提供開放資源軟體，可讓擁有Dataverse典藏的機構或組織下載、裝機與客置化使用。
10. **The Digital Public Library of America** (<http://dp.la>)
由許多組織共同使用的免費機構典藏，涵蓋的機構如美國國會圖書館、HathiTrust與the Internet Archive。在此平台中，可檢索與取得許多較大的數位圖書、影像、歷史文件視聽資料館藏。
11. **Distributed Data Curation Center** (<http://d2c2.lib.purdue.edu>)
Purdue大學的資料管理研究中心。
12. **The Dryad Digital Repository** (<http://datadryad.org>)
屬非營利性質、且是免費取得國際間科學與醫學文獻的典藏地，可存放許多不同的資料類型，並提高資料的能見度。
13. **DuraSpace** (<http://www.duraspace.org>)
2009年由Fedora Commons組織與DSpace Foundation合作成立的開放資料存取組織，並提供開放資源典藏軟體。
14. **Figshare** (<http://figshare.com>)
商業性的數位機構典藏，研究者可以保存其圖片、資料集、影像、視頻資料。可免費上載內容與存取，基本原則是開放資料。
15. **The Harvard Dataverse Network** (<http://thedata.harvard.edu/dvn>)
開放給全世界的研究者出版其研究資料，並可讓各學門的研究者長期典藏資料之處。

16. **The National Information Standards Organization (<http://www.niso.org>)**
非營利組織，支援出版內容的探索、檢索、管理與典藏。將圖書館、出版商與資訊系統廠商做連結，作為教育發展並提供資訊交換的技術。
17. **OpenDOAR (<http://www.opendoar.org>)**
開放存取學術機構典藏名錄。
18. **The Registry of Research Data Repositories (<http://www.re3data.org>)**
為研究者、贊助組織、圖書館與出版商提供現有的機構典藏名錄，主由 Humboldt-Universität 之 Berlin School of Library and Information 建立。並於2015年末，將所有來自 DataCite 的資料全部整合於此機構典藏內。
19. **VIVO (<http://www.vivoweb.org>)**
由 Cornell University 開始，是一個跨學科的網路，提供合作的研究者檢索相關的個人資料、部門、課程、獎助資訊與出版物。此為一個開放資源語意網路應用，參與機構可掌控所有的內容與檢索及瀏覽能力。
20. **Zenodo (<http://zenodo.org>)**
CERN 管理的機構典藏，為促進歐盟跨學科研究資料之分享。

資料管理計畫 (Data Management Plan)

NSF 與其他政府機構對於資料管理有新的政策，在美國申請 NSF 研究計畫時，必須要提出「資料管理計畫」(data management plan, 簡稱 DMP)，DMP 特別強調於申請計畫時即要規劃研究成果的傳播與分享，以符合 NSF 的要求。DMP 要求撰寫的內容包含如下：

1. 資料類型、樣本、實體館藏軟體、課程資料、與其他於此計畫產出之資料。
2. 資料、詮釋資料與內容使用的標準。
3. 存取與分享的策略。
4. 重複使用、傳遞、與產生的策略。
5. 典藏資料、樣本與其他研究的規劃。

參考資料：

1. 王梅玲 (2015 年) 。研究數據管理服務 (Research Data Management Services) 。檢索日期：2018 年 6 月 22 日 。網址： <http://techserviceslibrary.blogspot.tw/2015/01/research-data-management-services.html> 。
2. 鄭瑋 (2018 年) 。研究資料基礎架構。檢索日期：2018 年 6 月 22 日 。網址：https://nol.ntu.edu.tw/nol/coursesearch/print_table.php?course_id=126%20U1520&class=&dpt_code=1260&ser_no=32048&semester=106-2&lang=CH 。
3. Baykoucheva, S. (2015). Coping with “Big Data”: eScience. Managing Scientific Information and Research Data. Amsterdam, Chandos Press.

4. Bryant, Rebecca, Lavoie, Brian, & Malpas, Constance. (2017). The Realities of Research Data Management: Part One: A Tour of the Research Data Management (RDM) Service Space. Retrieved from <https://www.oclc.org/content/dam/research/publications/2017/oclcresearch-research-data-management-service-space-tour-2017.pdf>
5. Bryant, Rebecca, Lavoie, Brian, & Malpas, Constance. (2017). The Realities of Research Data Management: Part Two: Scoping the University RDM Service Bundle. Retrieved from <https://www.oclc.org/content/dam/research/publications/2017/oclcresearch-rdm-part-two-scoping-service-model-2017.pdf>
6. Bryant, Rebecca, Lavoie, Brian, & Malpas, Constance. (2018). The Realities of Research Data Management: Part Three: Incentives for Building University RDM Services. Retrieved from <https://www.oclc.org/content/dam/research/publications/2018/oclcresearch-rdm-part-three-incentives.pdf>
7. Bryant, Rebecca, Lavoie, Brian, & Malpas, Constance. (2018C4). The Realities of Research Data Management: Part Four: Sourcing and Scaling University RDM Services. Retrieved from <https://www.oclc.org/research/publications/2017/oclcresearch-rdm-part-four-sourcing-scaling.html>
8. Coursera (2017). "Research Data Management and Sharing. RE-06-13-0052-13" from <https://www.coursera.org/learn/data-management>
9. MIT OpenCourseWare. (2016). "MIT Libraries Data Management. RES.STR-002 Data Management." Massachusetts Institute of Technology. License: Creative Commons BY-NC-SA. from <https://ocw.mit.edu/resources/res-str-002-data-management-spring-2016/>
10. Purdue Universities, & Harvard Libraries. (2015). Research Data Management workshop. Retrieved from https://library.harvard.edu/hpdms_workshop
11. University of California Curation, Center of the California Digital Library (2018). "DMP Tools - Build your Data Management Plan." from <https://dmptool.org/>.

聯絡資訊：

魏令芳
東吳大學圖書館讀者服務組編纂
weilf@scu.edu.tw

李明錦
國立臺灣大學醫學圖書館主任
tracy@ntu.edu.tw