



從內政大數據經驗探討建立國家級數據中心的必要性

「資料」是數位時代最重要的戰略性資產，大量資料的分析及應用不可或缺。本文從推動內政大數據經驗，歸納探討建立「國家級數據中心」之必要性，並提出建議。

饒志堅（內政部統計處處長）

壹、前言

內政部 107 年正式啓動「內政大數據連結應用專案計畫」，整合跨域資料，進行議題分析及政策應用，經過 4 年（107-110 年）的密集推動，已形塑成爲機關文化，從決策者到基層執行者都非常重視大數據的連結應用。110 年內政部更草擬「內政大數據發展藍圖」，以數據範圍國家級、數據保障國家級、數據應用國家級之 3 大願景爲基礎，提出未來 4 年（111-114 年）「內政大

數據 2.0 計畫」。此與前聯合國統計處張保羅處長，同時也是新加坡大學教授，於 109 年 12 月發表於世界銀行的報告內容頗有相似之處。張教授指出，傳統政府統計工作已面臨強烈挑戰，國際政府資料治理趨勢，大家希望運用更多的「數據」、而非「統計」來解決問題，以滿足現代政府決策細緻化、精準化的需求。

貳、國際益趨重視政府數據治理

張教授在報告中，引用荷

蘭「數據政府（Data Agenda Government）」所言：爲了解決社會問題，必須使連結「不同政府單位資料」這件事成爲可能（It must be possible to combine data from different government organizations.），並且必須讓分析這些資料成爲政府工作中的一環（Analysis of these data must be carried out within the chains in which government tasks are organized.）。同樣 109 年 12 月，英國數字文化媒體體育部（Department for Digital,

Culture, Media & Sport, DCMS) 發布《國家數據戰略》(National Data Strategy), 表示將強化公共部門數據科學的培訓, 設立政府首席數據官(Chief Data Officer), 改變政府間資料蒐集、管理、使用和共享數據的方式, 並建立向上集中和可互為使用的數據基礎環境(transform the way data is collected, managed, used and shared across government ... and create joined-up and interoperable data infrastructure)。此皆顯示國際間益趨重視政府數據連結、共享、應用之重要性, 透過資料驅動方式, 找到解決問題的方法。

參、我國跨域數據治理猶待提升

近年我國政府統計在各種公務資料大數據的運用如雨後春筍般地出現, 從主計總處每年辦理的創新精進競賽即可看出這股趨勢, 也從其中看到許多非常棒的作品, 如財政部「運用大數據方法把關海關巨量資

料品質」、法務部「創建假釋受刑人再犯預測模型」、新竹市政府「建物稅籍清查模組提升稅籍清查效率」等。而這幾篇的共同點是, 利用單一機關資料源, 即可將原始數據解析與資料治理發揮得淋漓盡致, 令人佩服。

惟從另一方面來看, 大數據應用若要提升資料治理效益, 那麼原始資料多元化, 即跨域連結是不可避免的趨勢。跨域連結的物件可以是開放資料(可透過相同空間統計單元—如最小統計區連結), 或原始資料(可透過 ID 或地址連結), 下頁表 1 為筆者近年接觸或參與之大數據應用案例, 從是否發揮資料治理之角度出發, 思考精進之對策。

舉例而言, 是否每個機關一定要做跨域資料連結? 當然不是。我們不是閒閒沒事, 整天想要去做跨域連結, 而應視有無實際之需要。從下頁表 1 發現, 「1 房屋價格指數、港口運量管理」、「2 國道塞車預測、假釋再犯預測」以單一資料源即可做得很好, 可馬上

應用; 只是假釋再犯預測模型若能將假釋犯之親屬關係(內政部戶政資料)納入, 相信更能有助提升預測準確度。另外某些應用, 做跨域連結就有必要, 如「8 新住民弱勢家庭精準輔導」、「9 青年無殼之就業身分判定」。同樣地, 是否有 Open Data 即足夠做資料治理? 對某些應用來說是足夠的, 如「1 房屋價格指數、港口運量管理」、「3 地創人生計畫、藍海市場探索»; 但某些應用一定得需要原始資料, 如「2 國道塞車預測、假釋再犯預測」、「4 銀髮安居計畫、警力資源配置」。

換句話說, 某些應用要做好, 非得做原始資料的跨域連結不可, 但若要做跨「多個機關」的資料連結, 迄今仍有一個難解的問題。多年前, 內政部即曾運用新住民資料探討其就業、教育、健康等狀況, 並得到許多有價值的資訊。但當時的數據連結是單源數據, 也就是內政部(X 新住民)分別與勞動部(A 就業)、教育部(B 就學)、衛福部(C 身心

論述》統計 · 調查



障礙、中低收)等資料連結，得到 $X+A$ 、 $X+B$ 、 $X+C$ 的統計結果，但如果我們想知道「無業」之「中低收」「新住民」有多少？抱歉，不知道，因為各機關嚴守個資去識別化原則，連結後傳回內政部的資料都不帶 ID，不能進一步串聯得到 $X+A+B+C$ 的結果，無法提供更進一步的精準照顧服務，令人扼腕、殊為可惜。

肆、資料中心重工問題逐漸浮現

各機關為解決 $X+A+B+C$

的問題，紛紛成立實體隔離的資料實驗室 (Data Lab，各機關名稱不一，較常稱「資料科學中心」，以下簡稱資料中心)，進行大數據資料處理。目前至少有財政部、衛福部、主計總處、勞動部、內政部及教育部共 6 個部會成立，確實發揮了數據連結應用的效益。例如主計總處國勢普查處，近年連結政府各單位有關個人「薪資」之公務檔案資料，包括財政部綜合所得稅檔、勞動部勞保被保險人檔及勞退提繳工資檔、衛生福利部全民健康

保險檔，以及公務人員保險檔，於 109 年辦理完成「結合大數據編編縣市別薪資統計」，掌握縣市別個人薪資變動詳情，供擬定相對策，深獲好評。

各機關成立資料中心固有其好處，也有其缺點，最主要的就是重工問題。我們都知道，大數據分析最費工、最長時間、花錢的地方在於資料的連結清整。前述國勢普查處做過的，將戶籍資料與勞保被保險人檔、勞退檔，以及公保檔的連結，內政部現在為無殼青

表 1 大數據應用案例－按資料治理功效、資料領域及資料開放性分

功效	領域性	開放性	應用案例	特性或困難	精進之對策
已發揮功效	單一資料	開放資料	1 房屋價格指數、港口運量管理	不含個人資料之應用分析	-
		原始資料	2 國道塞車預測、假釋再犯預測	機關內部作業改善精進	跨域效果更佳
	跨域資料	開放資料	3 地創人生計畫、藍海市場探索	電信信令 + 電子發票……	-
		原始資料	4 銀髮安居計畫、警力資源配置	人 + 房 + 地；電信 + 巡邏箱	-
功效可再提升	單一資料	開放資料	5 十公里內無醫院診所之人數	何不 15 公里？ 50 公里？	客製化 + 收費
		原始資料	6 社會住宅包租代管屋主來源	租屋雷達計畫落地之困難	鬆綁應用限制
	跨域資料	開放資料	7 少子化：已婚 + 就業 + 有殼	$A \& B \& C \neq A+B+C$	鬆綁連結限制
		原始資料	8 新住民弱勢家庭精準輔導	$(X+A) \& (X+B) \neq (X+A+B)$	鬆綁連結限制
			9 青年無殼之就業身分判定	不知保工會者之身分	鬆綁連結限制

資料來源：作者自行整理。

年研究案，也在做相同的事。更不用說，一些基本項目的清整比對，例如不同檔案（戶籍檔、建物檔、房屋稅籍檔、勞保檔）的地址比對，過去主計總處、財政部、內政部等許多機關都做過，只是清整結果及經驗無法有效移轉，殊為可惜。未來隨著大數據連結應用範圍擴大，資料中心將愈來愈多，重工現象及影響（浪費公帑、資料不一致）也將愈來愈大，值得重視。

另外，成立資料中心不代表就一定能解決「跨機關」X+A+B+C的問題。例如內政部為進行上頁表1之「9青年無殼之就業身分判定」，因不知投保工會者之身分，向資料（健保身分）來源機關申請

時，對方即堅持因為法規，一定必須去識別化後才能提供，致無法進行X+A+B+C，將影響部分資料確度。這部分若要突破，必須鬆綁連結、應用限制。

伍、國家級數據中心為終極目標

綜上所述，大數據連結應用的最大痛點就是，沒辦法完全做到X+A+B+C，即便建立完善的資料中心還是沒辦法，甚至過多的資料中心可能會產生重工、效益不彰的情形。筆者認為最好的方法是建立「國家級數據中心」，可提升跨域資料連結、清整效率，省時、省錢、省麻煩，並在公平、公正、公開監督下，發揮資料

最大應用效益。而「國家級數據中心」不必將所有機關的資料都傳輸拷貝過來，其最重要的工作是建立資料交換管理機制，也就是輔導符合使用條件的機關，進行資料清整並「回饋」清整後資料，避免重工浪費，完善資料管理。當然在這些落實前，最重要的是鬆綁資料連結、應用限制，建議立專法，在具公信力委員會監督下，可帶出資料進行X+A+B+C，發揮資料最大應用效益，福國利民。

考量實際環境，前述建議可分短中長期三階段進行，各期目標與方法整理如表2，期待藉此提升政府資料治理功效。

表2 提升大數據連結應用效益期程

期程	目標	方法
短期	鬆綁連結、應用限制	立專法，在具公信力委員會監督下，可帶出資料進行X+A+B+C。
中期	建立資料交換管理機制	回饋資料清整結果，避免重工浪費，完善資料管理機制。
長期	建立國家級數據中心	無須每個機關都成立各自的資料科學中心、重複拷貝資料，可提升跨領域資料連結、清整效率，省時、省錢、省麻煩。在公平、公正、公開監督下，發揮資料最大應用效益，達到數據範圍（跨域整合）、數據保障（資安保護）、數據應用（循證決策）三大國家級目標。

資料來源：作者自行整理。



陸、結語

世界銀行報告指出，許多國家已將政府機關資料連結應用列為重要工作並落實推動。例如馬來西亞於 109 年 9 月正式成立國家數據委員會（National Data Council），職司不同來源資料的連結、管理與應用；愛爾蘭發布「公共服務數據策略（Public service data strategy 2019-2023）」五年計畫；新加坡則是要建立政府資料辦公室（Government Data Office，GDO）來執行資料策略發展，「使數據成為數位政府的心臟」；更遑論加拿大、英國、荷蘭等高度重視數據管理的西方國家，早將數據政府視為必然的發展。由此可見，政府不同部門間的跨域數據蒐集、連結與應用並落實於工作中，已經成為國際趨勢，也是未來各國政府服務良窳的關鍵。期待我國能早日建立國家級數據中心，引領我國資料治理更上一層樓。

參考文獻

1. 饒志堅、黃毓怡，創辦內政大數據連結應用專案計畫 提升統計支援決策效能，主計月刊，753 期，6-13 頁。
2. Reimagining National Statistical Systems (Professor Paul Cheung, National University of Singapore, 16 Dec 2020)
<http://pubdocs.worldbank.org/en/547491608138891001/WB-16-12-2020-Reimagining-Dec-16.pdf>。
3. Data Agenda Government (Nederland)，
<https://www.nldigitalgovernment.nl/overview/new-technologies-data-and-ethics/data-agenda-government/problem-solving-with-a-data-driven-approach/>。
4. Public service data strategy (Ireland)，
<https://www.gov.ie/en/publication/1d6bc7-public-service-data-strategy-2019-2023/>。
5. Bringing Data into the Heart of Digital Government (Singapore)，
<https://www.csc.gov.sg/articles/bring-data-in-the-heart-of-digital-government>。

6. 英國 2020《國家數據戰略》與世界各國對比解析，
<https://read01.com/az8PBj7.html>
<https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>。

