

精準醫療年代的健康資料治理

許明暉 臺北醫學大學大數據科技及管理研究所教授兼所長

壹、前言

精準醫療是全球醫療發展趨勢，世界各國都投入大量資源推動相關發展。精準醫療著重個人的基因、所處環境和生活方式上的個體差異，據以提供個人化的照護。精準醫療使醫療人員面對不同病人時能夠更準確地選擇有效的疾病預防策略與治療方式。精準醫療一詞雖然相對新穎，但事實上此一概念在醫學中早已存在。例如輸血時須選擇正確血型、選擇抗生素需考慮細菌對抗生素的敏感性都是精準的概念。隨著生醫科技的快速進步，大量的健康資料為資料治理帶來新的挑戰。

貳、成功防疫仰賴有效的健康資料治理

我國健康資料治理有優良的傳統。全民健保在籌備時期就把資訊系統列為關鍵基礎設施，並逐步推動醫療領域電子化，包含健保 IC 卡的發行。25 年來，資訊科技是臺灣全民健保營運的重要支柱。此次臺灣新冠肺炎防疫成績舉世稱羨，優良的健康資料治理功

不可沒。我國新冠肺炎中央疫情指揮中心於 2020 年 1 月 20 日啟動，當時臺灣尚無新冠肺炎病例。中央疫情指揮中心啟動後即將健保相關資訊系統納入疫情控制機制。整合移民署提供的入出境名單資料，讓醫療院所於健保雲端系統可以查詢民眾旅遊史。除了查詢就醫病人的旅遊史，系統之後也開放給醫院查詢因陪病或探病需要進入醫院民眾的相關資料。廣受好評的實名制口罩預售也充分利用了健保相關資訊系統。

參、健康資料是精準醫療關鍵基礎

健康資料可以分成結構化與非結構化兩個類別（圖 1）。性別、年齡、身高、體重、心跳、血壓、體溫、檢驗數值與藥品處方等是結構化資料；心音、呼吸音、文字病歷是非結構化資料。有部分的健康資料雖非完全結構化，但已有一定的格式，如以 XML 定義的電子病歷單張、以 JSON 呈現的基因資料，可列為半結構化資料。隨著人工智慧的興起，如電腦視覺、自然語言處理等技術的引進，電腦處理非結構化或半結構化資料的能力大幅增加，以往完全仰賴人工閱讀的非結構化資料，目前也已經可以利用電腦進行演算。

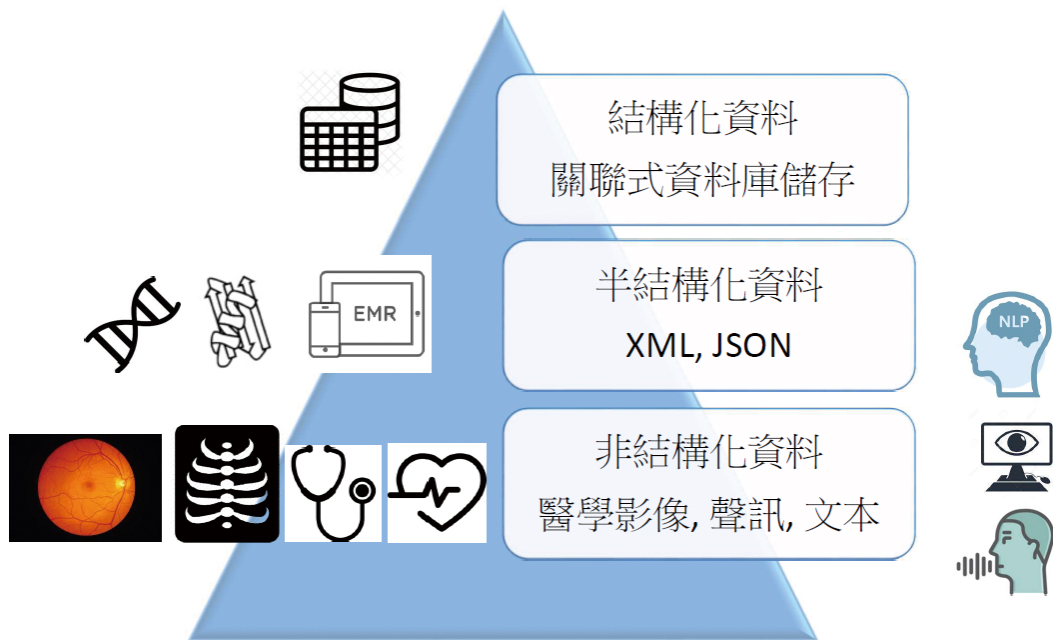


圖 1 健康資料類別

資料來源：作者繪製

肆、健康資料分享模式

應用多機構健康資料進行醫學研究，是精準醫療的重要趨勢。如何兼顧倫理、隱私等議題，分享健康資料是各國面對的共同挑戰。健康資料分享的模式可分為三類：

一、第一類模式

研究通過 IRB (Institutional Review Board, 人體試驗審查委員會) 審查後，研究者向擁有資料的機構申請使用資料，經機構確認後，研究人員可將資料直接攜出，並約定一定時間後歸還資料 (圖 2)。資料攜出後實為使用者所有，無法有效管控用途。這個模式如早期國衛院提供之健保資料庫服務。

二、第二類模式：如衛福部之「衛生福利資料科學中心」

衛服部參考先進國家資訊隱私保護作法，訂定「衛生福利資料統計應用中心作業須知」。所有提供應用之資料均經嚴密去識別化處理，研究人員使用資料須遵守下列原則：

- 1、研究人員須於「衛生福利資料科學中心」之獨立作業區域內進行資料應用分析。獨立作業區域為設有門禁管制並裝設全天候監視設備之獨立空間、出入需配戴識別證、為單一出入口、僅可於指定電腦作業、無對外網路，不可攜入實體參考文件、手機、攝(錄)影機、筆記型電腦、隨身碟及各類可攜式儲存設備。獨立作業區內電

腦安裝側錄系統，研究人員資料處理使用之軌跡永久保存（圖 3）。

2、申請者於獨立作業區進行統計分析，僅允許攜出不具識別性之統計結果，個案資料不得攜出。「衛生福利資料科學中心」之獨立作業區域提供了嚴密的資訊安全架構，但因席次有限，無法滿足使用者需求。且因位於臺北市，對其他縣市使用者需舟車勞頓，非常不便。為了提升資料使用的

可近性，目前在全臺灣設置 10 個研究分中心，提供與「衛生福利資料科學中心」相同的服務。研究分中心設置的獨立作業區，規範與「衛生福利資料科學中心」之獨立作業區域相同。第二類模式中研究者申請於獨立作業區使用資料，登入封閉網路的伺服器或主機上操作，無法將資料攜出。最後產出的統計分析結果須交由審查員，審查通過後檢送研究者。此模式嚴格管控資料安全，但研究時間地點受限。

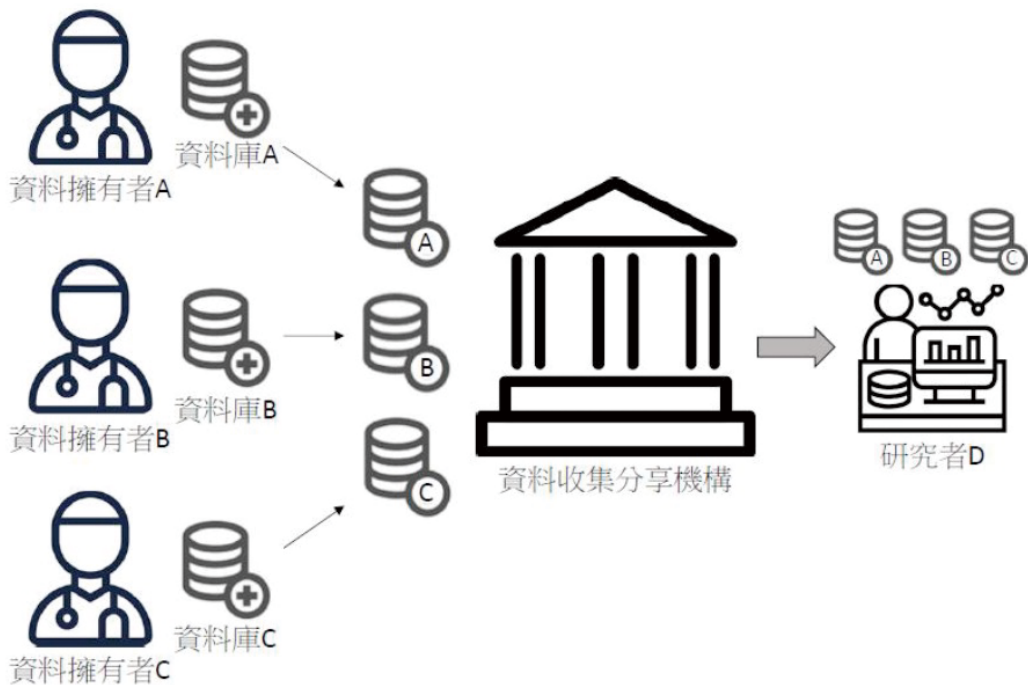


圖 2 第一類模式示意圖

資料來源：作者繪製

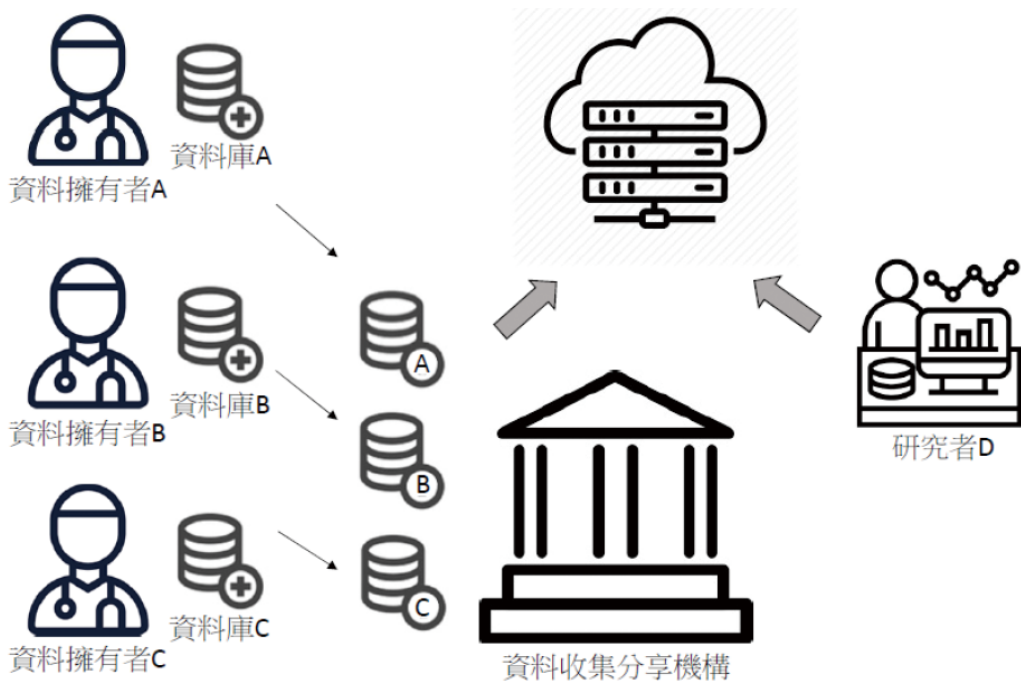


圖 3 第二類模式示意圖

資料來源：作者繪製

三、第三類模式

個別研究者保有自己的資料，只交換運算邏輯，不交換資料，在自己的機構運算完成後，再將統計結果進行整合分析。各機構研究者共同研究一主題，首先選定主題，選定疾病、確定疾病碼、確認統計方法等，各自分析資料，再整合多機構結果。此模式不需交換健康資料，只分享運算結果，研究者於個別機構進行分析，可增強合作之效率，又能確保資料安全（圖 4）。這是觀察性健康醫療資料科學與資訊學聯盟（The Observational Health Data Sciences and Informatics,

OHDSI）所推動的模式。觀察性醫療結果研究聯盟（The Observational Medical Outcomes Partnership, OMOP）成立於 2008 年，是由美國食品藥物管理局（FDA）主導，美國國家衛生研究院（National Institutes of Health）管理，國際藥廠資助的公私協力夥伴關係，該組織與研究人員和健康數據夥伴共同合作，其目的在透過分析觀察性健康數據主動監測醫療產品的安全。為克服在集中式環境和分散式研究網絡中跨不同觀察性數據庫進行研究的技術挑戰，該團隊設計了 OMOP 通用數據模型（Common Data Model, CDM），該機制可標準化觀察性數據的結構，內容和語

義，同時只需編寫一次統計分析代碼即可在每個機構重複使用。OMOP 建立了一個公共數據模型和標準化詞彙，可以針對不同的觀察性數據庫進行高效率的分析。從一開始，OMOP 就採用開放科學的方法，將其所有工作成果（包括研究設計，數據標準，分析代碼和結果）公開於公共領域（<https://github.com/OHDSI/>），以提高透明度。OMOP 最初的重點是藥品安全，但 OMOP CDM 不斷發展，目前已擴展至可應用於醫療處置和衛生政策的研究。在 OMOP 的基礎上，形成了 OHDSI。OHDSI 進行方法學研究，以建立科學的最佳實踐，適當地使用觀察健康數據，

開發開源分析軟體，統整編碼為一致，透明，可再現的解決方案，並將這些工具和實踐應用於臨床問題以產生指引醫療保健政策和病人健康照護的證據。臺北醫學大學目前代表臺灣參與此一國際健康數據聯盟。

ODHSI 著重在使用統計分析的方法學，人工智慧的發展也有類似的概念，稱為聯邦式學習。聯邦式學習是一種分散式機器學習框架，無需建立單一數據湖（Data Lake），即可訓練人工智慧演算法，可以強化隱私保護。數據不需要離開設備端各自在自己的設備訓練模型，並且通過特定的加密機制在雲

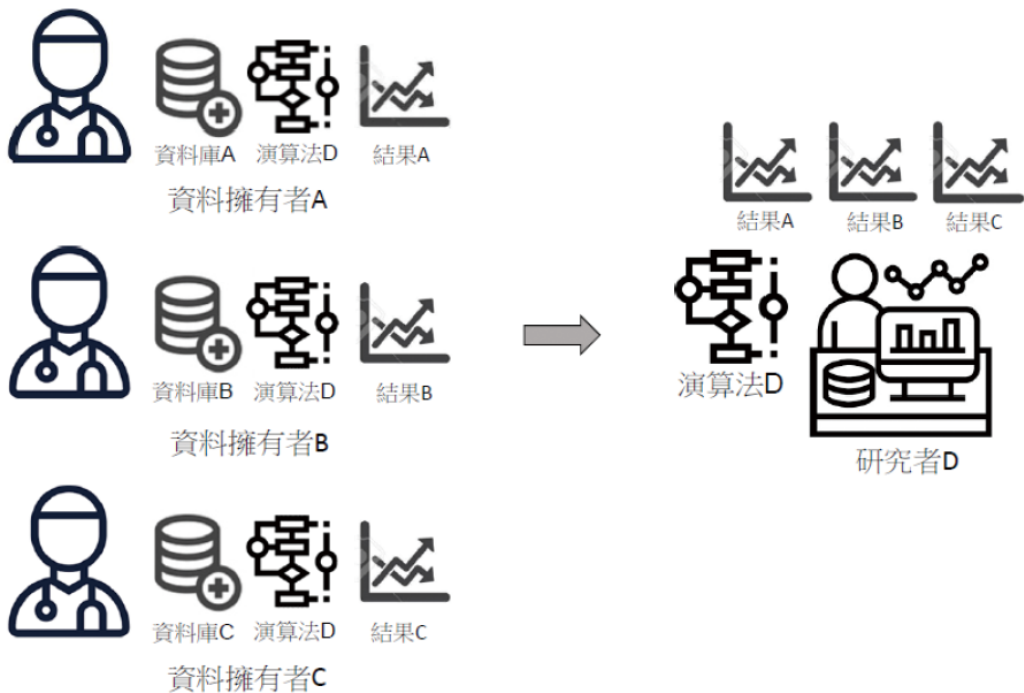


圖 4 第三類模式示意圖

資料來源：作者繪製

端建立一個共有的模型與進行模型的更新，用以改善產品設備的品質，透過聯邦式學習的概念所有的訓練數據都仍然保留在原本的設備上。橫向聯邦式學習適用於特徵重疊性高且樣本重疊少時的情境，比如不同地區的醫院，他們的服務內容相似（特徵相似），但病人不同（樣本不同）。

伍、結語

因為全民健保的實施，我國累積了巨量優質的健康資料。這是發展精準醫療最有價值的資產。如何透過健康資料治理將這些資產轉換成改善國民健康，發展新興產業的利器，是我們責無旁貸的任務。



Public Governance Quarterly