

109 年度工作研究報告

題目：人工智慧(AI)共享基礎建設—
以生醫健康領域試行規劃

撰寫人：單位 前瞻及應用科技司

職稱 研究員

姓名 林滋梅

有意願參加本部獎勵科技行政研究發展評獎(有意願者請打勾)

單位主管評語	
推薦參加本部獎勵科技行政研究發展評獎	(請打勾)
單位主管簽章	

備註：

- 一、報告內容以10頁為原則。
- 二、本篇工作研究報告，如參加本部獎勵科技行政研究發展評獎，請依本部獎勵科技行政研究發展作業要點規定辦理。

(本頁空白)

目錄

壹、緣起.....	1
貳、研究方法與過程.....	2
參、國際上生醫健康 AI 資料與模型共享之政策和作法	2
(壹) 美國及歐洲政府相關政策.....	2
(貳) 國際組織相關推動.....	11
(參) 小結.....	17
肆、我國政府相關計畫或施政作為.....	18
(壹) 臺灣 AI 行動計畫(2018-2021 年).....	18
(貳) 衛生福利資料科學中心.....	18
(參) 健保資料人工智慧應用服務試辦計畫.....	19
(肆) 國家級人體生物資料庫整合平台.....	19
(伍) 醫療影像之巨量資料建立與應用研究專案計畫	20
(陸) 臨床資料庫與 AI 之跨域開發及加值應用計畫	21
(柒) 精準健康大數據永續平台計畫.....	21
(捌) 小結.....	22
伍、科技部 AI 專案共享基礎建設之規劃	22
(壹) 共享項目和參與成員.....	22
(貳) 共享平台.....	23
(參) 共享機制.....	24
陸、參考文獻.....	27

(本頁空白)

壹、緣起

近年來在科技部推動人工智慧(AI)與巨量資料相關專案的引導下，國內智慧醫療相關研究發展方興未艾，且有相當亮麗的成果。以科技部醫療影像 AI 專案為例，106 年 11 月至 109 年 5 月共 2.5 年期間開發了心、腦、肺、肝、脊椎、手腳關節等器官重要疾病的醫療影像 AI 模型共 28 項，其中臺北榮民總醫院的腦轉移瘤 AI 輔助診斷系統(DeepMets®)，讓醫師閱片時間由 10 分鐘縮短為 30 秒鐘，患者診療流程由 2 週縮短為 0.5 天；國立臺灣大學開發世界唯一自動分類及計算胸腔鈣化/脂肪定量的 AI 模型，分析一個病例平均只需要 0.4 秒(人工需 60 分鐘)；臺北醫學大學肺癌早期病灶偵測與分析的 AI 模型，10 秒鐘自動產生報告，醫師只須 5 分鐘複核(人工需 20-30 分鐘)。這些成果若能融入醫院診療流程，將可大幅減輕醫師診療負擔，也減少病人等待的時間與病情不確定所造成的焦慮感[1]。

然而，要將 AI 模型落地應用於醫療現場、實現科研成果的應用價值，取決於諸多因素，例如：(1) AI 模型的普適性：係指 AI 模型在不同儀器產出資料的判別上，都具有高準確度。AI 模型要具有普適性，需要足夠的跨機構資料甚至是全國性的資料(如：健保資料庫)來驗證；這點在影像資料尤其重要，因為各醫院不同廠牌、不同型號的儀器所產出的影像，在解析度和品質上都會有差異。(2)軟體確效及商品化：AI 模型在醫療現場應用，屬於醫療器材範疇，商品化前需要通過主管機關(如：TFDA、FDA)嚴謹的查驗過程；要擴散應用，需要能與其他類似或相關產品競合，並且要有良好的商業模式。(3)使用者端的接受度：需能符合臨床需求，且需分別針對不同醫院的臨床系統與工作流程，設計軟體界面進行介接機制與整合。

AI 模型的普適性是上述因素中最根本需要解決的問題，這個問題可以透過 2 種方式解決：一種是將不同醫院的資料集結以訓練及測試 AI 模型，借由增加訓練與測試資料集的廠牌歧異度，提高 AI 模型對不同廠牌資料的適應性，這種方式本報告稱之為集中式的資料共享；另一

種方式是將 AI 模型提供給不同醫院進行驗證，各醫院將驗證結果進行整合並調校 AI 模型，這種方式本報告稱之為分散式的資料共享，並需搭配模型共享。考量個資議題，有越來越多研究是利用第二種方式。

有鑑於此，科技部 AI 創新研究中心專案計畫（以下簡稱 AI 專案）已將 AI 研發所需的資料及模型共享，納入五大推動策略（國際級研究中心、核心技術研究、領域應用研究、**共享基礎建設**、人文法制環境）之一，預計於 2021 年初以此五大策略，徵求研究中心及核心技術攻堅等研究計畫，**獲補助的研究計畫將於 2021 年 8 月開始執行**。資料及模型的共享，讓 AI 開發人員不需要從事研發最底層的資料蒐集與清理、基礎 AI 模型建構等，而可以運用現有的標註資料與模型再加值，不僅降低 AI 開發的進入門檻，且具有介接和加乘 AI 開發者彼此能量的效果，將可集結眾人的能量加速深化 AI 研發成果，提升資源投入的效益和競爭力。

雖然生醫健康領域僅是 AI 的眾多應用主題之一，惟生醫健康領域所利用的資料多數涉及個人資料，在資料共享方面是最為複雜的領域，因此本報告將以生醫健康為先導領域，提出 **AI 共享基礎建設**的作法建議。

貳、研究方法與過程

- 一、蒐集國際上 AI 推動或生醫健康領域 AI 推動的資料及 AI 模型共享重要策略及作法，並歸納出共享基礎建設應考量的重要議題。
- 二、蒐集我國政府所推動的生醫健康領域資料建置和共享利用相關計畫。
- 三、從 AI 共享基礎建設的 5W1H：推動理由(why)、共享項目(what)、參與成員(who)、共享平台(where)、啟動時程(when)、共享機制(how)，規劃具體作法。

參、國際上生醫健康 AI 資料與模型共享之政策和作法

(壹)美國及歐洲政府相關政策

一、美國國家 AI 研發策略計畫(National AI Research and Development Strategic Plan)

由美國國家科技委員會(National Science and Technology Council)於 2016 年發布，並於 2019 年 6 月更新；含八大策略，其中策略五為「發展用於 AI 訓練及測試的共享資料集及環境」，其重點包括[2]：

- (一)發展各種不同的資料集，包括政府擁有的資料、政府資助產生的資料及民間資料，並透過資料的前處理(如:清理、整合、轉換等)與分析，確保資料完整性及可用性，以滿足各種不同領域 AI 利益與應用的需求。
- (二)增加資料集的取用，包括：建立共通資料格式、開發工具及提供運算資源，讓資料更易於識別、使用、操作、驗證來源及合規使用；以及建立資料共享的安全及隱私保護技術與機制。
- (三)使 AI 訓練和測試資源對應於商業和公共利益需求，由於目前資料擷取(capture)¹、管理(curation)²、分析和可視化等技術的發展，落後於資料大爆炸的速度，而且尚無法支援富語意的搜尋(semantically rich search)；將支持相關工具、模型、方法論的研發，以因應這些挑戰。
- (四)發展開源軟體庫和工具箱(open-source software libraries and toolkits)，將強化投入於開放 AI (open AI)的技術開發、支援及使用，或是開放政府所擁有的演算法或軟體，甚至透過政府本身導入開放 AI 資源來加速開放 AI 技術的利用，以降低 AI 創新者進入門檻。

目前本項策略的相關計畫例舉如下：

- (一)聯邦政府開放資料平台(data.gov)：匯集美國政府的開放資

¹ 資料擷取係指由外界擷取各種不同的資料，並將之處理使之能為各種計算機設備接受的資料的一種行為。

² 資料管理是指將不同來源資料加以組織及整合的行為。

料。

(二)**聯邦政府的開源軟體共享平台(code.gov)**：依據 2016 年 8 月發布的聯邦開源程式碼政策(Federal Source Code Policy)，美國聯邦政府各單位至少要開放 20% 新訂製軟體的程式碼，希望促進程式碼重複使用和開源，並教育和聯結開源社群來協助政府單位的合作夥伴和開發者，降低程式開發的花費及提升程式品質[3, 4]。

(三)**NIH STRIDES Initiative**：NIH 與 Google、Amazon 等雲端服務領先廠商建立夥伴關係，使研究人員能夠近用由 NIH 資助並儲存在雲端環境中資料[2, 5]。

二、維持美國在人工智慧領先地位行政命令(Executive Order on Maintaining American Leadership in Artificial Intelligence)

美國總統川普於 2019 年 2 月 11 日簽署此行政命令，共揭示六大目標，其第二項「**聯邦資料及模型的共享**」的目標，是促進高品質且可完全追蹤的(fully traceable)聯邦資料、模型和**運算資源的取用**，並同時在安全、保全、隱私、機密等保護上，維持與相關法規或政策一致。

2020 年 2 月美國白宮科技政策辦公室(OSTP³)針對上述行政命令發布的第 1 年報告，展現相關推動成果，其中有關「增加聯邦資料庫應用於 AI 研發及測試」部分，於 2019 年透過對大眾公開徵求意見(request for information)，綜整出聯邦資料及模型被研究社群近用(access)尚待解決的問題或需揭露的訊息，包括：資料及模型的可發現性(discoverability)、可近用性(accessibility)、可使用性(usability)、管理機制(如：來源、取用限制、隱私、安全、保全、智慧財產等)、資料的格式、品質和規模、模型訓練或驗證的方式等[6]。

三、美國國家衛生研究院(NIH⁴)之資料科學策略

³ 全名：White House Office of Science and Technology Policy

⁴ 全名：National Institutes of Health,

(一)研究資料共享聲明(NIH Statement on Sharing Research Data)：

NIH 於 2003 年 2 月發布，在隱私及資料所有權被保護的前提下，自 2003 年起向 NIH 申請的研究計畫如果每年研究經費達 50 萬美元以上，必須提出研究資料共享規劃，無法共享者須說明原因；共享的資料非指統計資料，而是指經整理用於產出研究結果的原始資料；在此之後，NIH 陸續發布了配套的實施指引[7]。目前由 NIH 的生醫資訊跨單位協調委員會(BMIC⁵)維護一個資料儲存庫清單，區分為 3 類：NIH 資助的特定領域開放資料儲存庫(共 67 個)、NIH 資助的特定領域的非開放資料儲存庫(共 31 個)、以及未限定資料類型、格式、內容或主題的通用儲存庫(共 7 個) [8]；研究人員可依資料性質及所屬學科將資料上傳到適合的儲存庫，僅有部分計畫要求上傳至特定儲存庫。另外，1988 年 NIH 設立的生物技術資訊國家中心(NCBI⁶)[9]，收錄生物醫學及基因等資料，也有一些基礎的、用於基因分析的軟體工具，提供給 NIH 資助的計畫及其他公民營機構(如：其他聯邦單位、公立健康實驗室、基因測試實驗室、生技或製藥公司等)使用，使用者也可以上傳資料到該中心的平台[10]。

(二)資料共享(Data Commons)：NIH 建立雲端平台，讓研究者可以儲存、分享及取用資料和軟體等數位資源，並與之互動，以加速生醫科技研發。2017-2018 年推動資料共享先導計畫(NIH Data Commons Pilot Phase)，以 TOPMed⁷計畫、GTEx⁸計畫和模式生物資料庫為對象，獲該計畫補助的產官學研單位必須組成聯盟，產出：(1)讓資料符合 FAIR⁹原則的指引

⁵ 全名：Trans-NIH BioMedical Informatics Coordinating Committee

⁶ 全名：National Center for Biotechnology Information

⁷ 全名：National Heart, Lung, and Blood Institute's Trans-Omics for Precision Medicine

⁸ 全名：NIH Common Fund's Genotype-Tissue Expression

⁹ FAIR 為可發現(Findable)、可近用(Accessible)、可互操作(Interoperable)及可再利用(Reusable)的縮寫。可搜尋：資料必須有唯一的識別符並於檢索資源中有效的標示出來；可近用：必須易於透過開放系統和有效且安全的身份驗證與授權程序進行檢索；可互操作：必須使用標準的語彙；可再利用：必須向使用者充分描述資料及使用授權，並可追溯資料來源[10]。

和評估指標、(2)資料全球單一識別符(Global Unique Identifiers, GUIDs)的方法、(3)以開放標準為基礎的 API、(4)明確的雲端平台架構、(5)尋找資料和相關工具並與之互動的工作空間、(6)研究倫理與隱私保全(含身份驗證和授權)作法、(7)索引和搜尋功能、(8)資料共享如何促進生醫研究的示例、(9)協調、訓練與外溢作法等[10, 11]。

(三)資料科學¹⁰策略計畫(Strategic Plan for Data Science)：NIH 2018 年發布，以 FAIR 為原則訂定 5 項目標，並針對每個目標提出其作法與成效評估指標；其中 3 項目標與資料和模型的共享相關[10]，分述如下：

目標一：支持高效能及有效的生醫研究資料基礎建設

1. 資料儲存及保全最適化：透過與民間雲端服務公司策略聯盟或採購，建立平台即服務(PaaS¹¹)環境，儲存 NIH 資助的多樣化且高價值資料，並在符合「健康保險可攜與可課責法案」(HIPAA¹²)的保全規範和國家標準與技術機構(NIST¹³)之健康資訊保全標準下，形成軟體即服務(SaaS¹⁴)範例[10]。
2. 串連 NIH 資料系統：以 NCBI 為中心，連結資料共享雲端平台及 NIH 被廣泛使用的資料庫(在適合的情況下也將連結 NIH 以外的資料庫)，促進跨資料庫的利用，避免資料孤島(data silos¹⁵)。目前國家轉譯科學促進中心(NCATS¹⁶)所推動生醫資料轉化器(biomedical data translator)計畫已於病理生理學方面有相關的推動[10, 12]。

¹⁰ NIH 將資料科學定義為獲取資訊的跨學科研究領域，該領域將發展並使用計量和分析方法、流程和系統，以利從越來越大和/或越來越複雜的資料獲取知識和見解[13]。

¹¹ Platform as a Service

¹² 全名：Health Insurance Portability and Accountability Act of 1996

¹³ 全名：National Institute of Standards and Technology

¹⁴ Software as a Service

¹⁵ 資料孤島是指無法與其他資料集互操作的資料集。

¹⁶ 全名：National Center for Advancing Translational Sciences

目標二：促進資料生態系現代化

1. **資料儲存庫現代化**：針對知識庫(knowledgebases)及資料庫(databases)¹⁷，分別以不同的補助策略、審查標準及成效評估機制來推動建置，並且動態衡量資料使用及修改情況；建立單一、有效且安全的授權使用作法，以確保隱私和資料保全 [10]。
2. **支持個人資料集的儲存和共享**：個人資料集是指個別研究計畫產出的資料集，通常這種資料集的資料量遠小於NIH專案資助建立的資料集，但資料集數目及多樣性則相對較高。NIH將建立友善的界面，讓個人資料集可以透過PubMed Central 和NCBI連結到其所發表的科學文獻，並以將NIH資料共享延伸至這些個人資料集為長期目標[10]。
3. **與衛生及公共服務部所屬的國家健康資訊技術協調辦公室(ONC¹⁸)合作，槓桿其他政府相關行動計畫(例如：All of US、癌症登月計畫、TOPMed、退伍軍人醫療資料等)**，將其蒐集的資料整合到生醫資料科學生態系中。

目標三：支持先進資料管理、分析及視覺化工具之發展及擴散，槓桿現有且活躍的工具共享系統，協助建立讓工具開發者和提供者更具競爭力的市集，並朝使用者僅需付出有限成本(甚至零成本)的方向推動，以利工具能更廣泛的被使用[10]。

1. **支持有用的、通用的、可近用的工具和工作流程開發**，推動方式例如：將與補助資料庫或知識庫建置的計畫分開補助，以利評估所開發出來工具的效用，減少計畫經

¹⁷ 資料庫(databases)主要是儲存經過品質管理的資料，而知識庫(knowledgebases)除了含有資料之外，會運用人的知識對資料進行闡釋後加入額外的資訊[9]。例如：集合醫學影像及其來源個體基本資料的是屬於資料庫，若這個資料庫的影像上有圈註病灶位置及標示病灶大小等，則屬於知識庫。

¹⁸ ONC全名為Office of the National Coordinator for Health Information Technology，是負責美國全國性的健康資料電子交換之協調與實施的主要聯邦單位。NIH將延用ONC所建立的實務作法，例如：臨床照護的標準化術語(Terminology)或用詞(Vocabularies)。

費在不同重點配置不同而造成計畫評估上的混淆；引導學界將其開發的原型和演算法透過系統整合商或工程師來優化，成為有效、具成本效益且可廣泛使用於生醫研究的軟體工具；以演算法競賽或公私協力等作法來促進資料科學和工具創新[10]。

2. **將為特定領域所開發的工具應用於其他領域，以擴大這些工具的效用、可使用性和可取用性：**鼓勵各領域採用或適應新興的方法、演算法、軟體和工作流程；促進數學、統計學、電腦科學、工程學和物理學等相關領域對生醫資料科學的創新貢獻；促進穿戴裝置和資料介面工具開發和應用；開發新方法，讓生醫研究者可以安全且合規取用電子病歷及其他臨床資料[10]。
3. **改善資料的編目，提高資料可查找性及再利用：**將槓桿 ONC 在 U.S. Data Core for Interoperability 的成果，促進社群在索引、引用及追蹤修訂(資料溯源)方面，發展並採用一致的標準[10]。

四、美國退伍軍人腫瘤精準醫學資料共享(Veterans Precision Oncology Data Commons, VPODC)

由退伍軍人事務部(VA)、芝加哥大學資料科學轉譯中心(the Center for Translational Data Science at the University of Chicago)及開放公共聯盟(Open Commons Consortium)合作，將 VA 的臨床、基因組和影像等資料整合為精準腫瘤學研究的國家資源[13]，提供給研究社群研究利用，並支援機器學習和 AI 工具的使用[2]，以促進精準治療、診斷及相關技術之研發。研究人員可在 VPODC 線上平台管理、分析及分享腫瘤資料[14]。VPODC 的資料保全是遵循美國「聯邦資訊安全現代化法案」(FISMA¹⁹)所衍生由 NIST 制定的「聯邦資訊系統及組織之保全及隱私控制規範²⁰」[2, 13]。

¹⁹ 全名：Federal Information Security Modernization Act,

²⁰ 全名：NIST Special Publication 800: Security and Privacy Controls for Federal Information Systems and

五、 歐盟人工智慧推動策略

強化公部門及私部門在 AI 的投資、培育充分使用 AI 和自動化的工作技能、以及確保適當的道德和法律框架，是歐盟推動 AI 的三大主軸。其中強化 AI 投資部分，已推動數位化歐洲計畫 (Digital Europe Programme)，促進大量資料集和演算法之儲存和利用，並在健康等領域強化既有 AI 測試和實驗設施；另已建立 AI4EU 平台，目標是創造歐盟 AI 生態系，匯聚專家、知識、演算法、工具和資源(含資料和運算資源)，提供以 AI 為基礎的服務、產品和解決方案，支援 AI 發展[15]。

六、 歐洲資料策略(A European Strategy for Data)

由歐盟執委會(European Commission)於 2020 年 2 月發布，包含四大主軸[16]：

- (一)建立跨政府部門的資料近用和使用架構。
- (二)透過在資料的投資，以及強化歐洲在資料管理、處理、使用和互操作性的能力和基礎建設，成為資料利用的促成者：將在 2021 年至 2027 年之間投入歐洲共同數據空間暨聯合雲端基礎建設高影響力計畫²¹。
- (三)增強個人和中小企業的資料利用能力。
- (四)建立策略經濟部門和具公共利益專業領域的歐洲共同數據空間(common European data space)：將參照歐洲開放科學雲(European Open Science Cloud)的建置經驗，建立 9 個歐洲共同數據空間：製造業(manufacturing)、綠色新政(green deal)、交通(mobility)、**健康(health)**、金融(financial)、能源(energy)、農業(agriculture)、公共行政(public administration)、技能(skill)。其中，**健康數據共同空間**：
 1. 一方面可強化資料再利用並支援實證基礎的健康決策，以利醫療照護體系更容易有效且永續近用；另一方面也

Organizations

²¹ High Impact Project on European data spaces and federated cloud infrastructures

可促進民眾依據 GDPR 規定，行使電子病歷近用權、可攜權和同意處理權等。

2. 歐盟執委會將針對健康數據共同空間，佈建數據基礎建設、工具和運算能力，以加強支援各國電子病歷發展，也透過電子病歷交換格式的建立，強化健康資料的互操作性；另外，也將擴大跨國界健康資料交換，在符合 GDPR 規範下聯結並使用各類型的健康資料，如電子病歷、基因(2025 年將至少有 1,000 萬人)、數位健康影像等。

七、 歐盟資料治理規則草案(Data Governance Act²²)

由歐盟執委會於 2020 年 11 月 25 日提出，對以下 4 種情況建立法制規範[17]：

- (一)公部門所保有且受到他人權利約束的資料(例如：個資、智財、營業秘密或其他商業資訊等)之再利用。
- (二)企業間的資料共享。
- (三)透過「個人資料共享中介單位」(personal data-sharing intermediary)，協助個資當事人實踐 GDPR 的權利並促進個人資料的再利用。
- (四)建立資料利他組織的登記制度和相關規範，並且建立歐洲共通的資料利他同意書表(consent form)，以利個人和企業資料在利他的基礎上被利用。

八、 歐盟 ELIXIR

ELIXIR 是歐盟跨政府的生物資料資源協調組織，目前共有 23 國參與，其目標在於協調、整合及支持跨國生物資訊學資源，形成單一的基礎建設，包括：資料庫、軟體工具、訓練材料、雲端儲存與超級電腦等，也提供分析工具使用之教育訓練課程，以利研究人員共享並快速取得研究所需資料及分析工具。目前建置

²² 全名：Proposal for a Regulation of the European Parliament and of the Council on European Data Governance

了 5 個平台：資料平台、工具平台、運算平台、互操作(含標準)平台、教育訓練平台，並對產、學界提供服務。該組織專注於提供化學生物學、酶與其交互作用與路徑、演化與系統發育、基因與基因體、文獻、分子與細胞結構、蛋白質與蛋白質組等科學領域的服務[18]。

ELIXIR 以樞紐(hub)及節點(nodes)相互合作方式運作：樞紐類似 ELIXIR 的總部，負責跨國協調整合；節點則是每個會員國內數個研究機構所形成的網絡，其中 1 個機構負責監督轄下所有機構的工作。每個節點聚焦於其強項，提供會員國服務[18]。

(貳)國際組織相關推動

一、ITU-T/WHO 之 FG-AI4H 資料儲存庫及 AI 模型標竿平台

生醫健康 AI 模型的正確性及可信賴性對醫師是否願意使用是關鍵因素，而訓練資料和測試資料的正確性以及 AI 模型的產生方式與可解釋性，則會決定 AI 模型的正確性及可信賴性。目前已有如強化式學習(reinforcement learning)等 AI 技術發展出來，完全不需要具有領域專業(domain knowledge)的人員參與即可訓練出 AI 模型。因應類似技術的發展及應用，AI 模型的安全性需要加以考量。

為了確保 AI 模型的安全性，國際電信聯盟電信標準化部門 (ITU-T²³)結合世界衛生組織(WHO²⁴)在醫療健康領域的專業，於 2018 年 7 月共組「健康人工智慧焦點小組」(FG-AI4H²⁵)，廣邀臨床醫師、AI 研發人員、企業、實際場域專家、潛在標竿資料提供者、監管機構、政策制定者等利害關係人參與，將建構健康 AI 模型的標準評估架構及標竿程序(benchmarking process) [19, 20]。配合 AI 模型的標竿評估需求，將創建公共資料儲存庫(public data repositories)並建置 AI 模型標竿測試平台。目前已在 20 個健康主

²³ 全名：International Telecommunication Union- Telecommunication Standardization Sector

²⁴ 全名：World Health Organization

²⁵ 全名：Focus Group on Artificial Intelligence for Health

題組成主題小組，公開徵求各界參與。

FG-AI4H 初步規劃的標竿程序包括 6 個步驟如下[19]：

- (一)由 FG-AI4H 創建公共資料儲存庫，存放高品質(準確、可靠、可驗證)的 AI 訓練用資料集，以利想要參與標竿測試程序人員形成多樣化生態系。
- (二)AI 模型標竿的參與測試者以公共的或非公開的資料集來訓練 AI 模型。
- (三)將 AI 模型傳送到標竿測試平台，由該平台對模型的基本合格性進行檢查。
- (四)合格的模型以 FG-AI4H 管理的非公開測試資料集進行評估。FG-AI4H 管理的非公開測試資料集將做為標竿評比 AI 模型的金標準資料集(gold standard)。
- (五)將 AI 模型的標竿評估結果提供給參與測試者，做為參與測試者後續模型開發或精進的參考。
- (六)中央排行榜(central leaderboard)可進行模型性能比較，或產生通過／不通過的評分，另外也可由 FG-AI4H 的亞組(Subgroup)定義評估的基準。

上述程序要能執行，FG-AI4H 需要先針對每個特定健康領域(Domain)，界定 AI 開發的需求(如：一般診斷、專科診斷、健康自然語言處理、一般臨床資料提取和編碼、實驗室編碼等)，並且取得測試資料集、擇定資料集金標準的測試成功等級(例如：某個測試資料集的專業分數)、設定 AI 系統的等級(例如：可用於醫師決策支援、可用於自動操作)、定義可接受的故障模式(例如：若低於給定的可信度臨界值則警告操作人員)等[19]。因應這個需求，FG-AI4H 已針對醫療健康主題建立主題小組(Topic Group, TG)，另外也針對各領域的共通課題建立工作小組(Working Group, WG)，廣邀利害關係人參與。截至 2020 年 9 月共有 **20 個主題小組** [21]：

1. 心血管疾病管理(TG-Cardio，其下另有風險預測次主

題)；

2. 皮膚科(TG-Derma)；
3. 避免老人跌倒 (TG-Falls)；
4. 組織病理學(TG-Histo)；
5. 神經系統疾病 (TG-Neuro)；
6. 感染性疾病的爆發偵測(TG-Outbreaks)；
7. 眼科(TG-Ophthalmology)；
8. 精神病學(TG-Psy)；
9. 蛇咬及其蛇種鑑別(TG-Snake)；
10. 症狀評估(TG-Symptom)；
11. 結核病(TG-TB)；
12. Volumetric 胸部電腦斷層於肺癌的診斷 (TG-DiagnosticCT)；
13. 細菌感染的診斷與抗生素抗藥性 (AMR) (TG-Bacteria)；
14. 牙科診斷和數位牙科 (TG-Dental)；
15. 第 1 型與第 2 型糖尿病預測 (TG-Diabetes)；
16. 內視鏡(TG-Endoscopy)；
17. 偽造藥物檢測(TG-FakeMed)；
18. 瘧疾檢測(TG-Malaria)；
19. 婦幼健康 (TG-MCH)；
20. 放射學 (TG-Radiology)。

6 個工作小組及其任務包括[22]：

1. 資料及 AI 解決方案評估方法工作小組(WG-DAISAM)：界定及定義資料及 AI 解決方案的評估方法，並於方法實際應用時，與 WG-DASH 協調；也會就評估方法和標準化評估架構的實施技術面，分別與主題小組和其他工作小組密切合作。
2. 資料及 AI 解決方案處理工作小組(WG-DASH)：聚焦於

標竿平台上的資料(公共的及私人的)及 AI 解決方案(演算法及模型)在其生命週期之處理操作層面;此工作小組將定義最佳操作實務、建立操作程序與政策及創建相關文件。與評估方法工作小組(WG-DAISAM)的差異在於本工作小組著重於操作層面而不是工具。

3. 健康領域 AI 的倫理考量工作小組(WG-ETHICS)：產生用於全球健康 AI 設計及實施的倫理指引；解決 AI 技術的道德發展和使用問題，包括中低收入國家如何從 AI 發展中受益。關注課題包括：**公平近用、隱私、適當的使用與使用者、責任、偏見、包容性**。
4. 健康領域 AI 的監管考量工作小組(WG-RC)：協助 FG-AI4H 瞭解監管的全貌及關鍵注意事項；促進 FG-AI4H 與監管機構的聯繫、資訊交流和合作；協助 FG-AI4H 定義出健康 AI 的成功標竿方法。
5. 健康領域 AI 的臨床評估工作小組(WG-CE)：將與其他工作小組及利害關係人建立臨床評估合作社群，並產出可供研究人員、臨床醫師、病人、開發者與政策制定者使用的評估指引。關注的課題包括：評估的階段、效能及相對效能、安全性、通用性/偏見與包容性、評估自適性/學習性模型(evaluating adaptive/learning models)、遵循「促進健康研究品質與透明網絡」(Equator Network²⁶)的指引研議評估報告、臨床上有意義的終點(包含用於主題內評估的終點)、部署後監管(post-deployment surveillance)、針對中低所得國家的特殊考量。
6. 操作工作小組(WG-O)：處理 FG-AI4H 內部程序運作相關工作，例如：後勤、會議、議程規劃、報告綜整等。

二、 IEEE 聯合機器學習(FML²⁷)框架標準

²⁶ Enhancing the QUALity and Transparency Of health Research (<https://www.equator-network.org/>)

²⁷ 全名：Federated Machine Learning

國際電機電子工程師學會(IEEE²⁸)於 2020 年 7 月發布聯合機器學習架構與應用指引草案²⁹，以促進聯合學習模型的建立，該指引的內容包括：(1)聯合機器學習(以下簡稱聯合學習)的描述及定義、(2)類型與應用情境說明、(3)使用者之主要活動界定、(4)系統架構、(5)一般關切議題(隱私保護和資料保全、效能、經濟可行性)、(6)成效評估機制[23]。

(一)聯合學習的定義與優點

聯合學習係指以分散儲存於不同機構/單位的資料集共同訓練 AI 模型，不同機構不需要相互揭露各自的原始資料，仍可以如同將多個資料集集中在一起一樣，得到比以單一資料集訓練有更好表現的 AI 模型[23, 24]。由於資料不需集中存放，可降低個資和隱私洩漏風險，有利於資料治理和解決機敏資料無法離開原始蒐集單位而造成的資料孤島問題。

(二)聯合學習的類型與應用情境

聯合學習依資料集與資料集之間個體重疊程度和資料欄位重疊程度的多寡，區分為以下 3 種類型：

1. **橫向聯合學習(Horizontal FML)**：資料集間的個體重疊程度明顯低於欄位重疊程度者[23]。例如：醫院都有肺 CT 影像，但其病人的重疊程度低，是以橫向聯合學習來訓練模型。透過橫向聯合學習，以多個個體數少的資料集訓練模型，可達到將所有資料集集中訓練一樣的效果。
2. **縱向聯合學習(Vertical FML)**：資料集之間的個體重疊程度明顯高於欄位重疊程度者[23]。例如：運用醫療資料與電信資料、經濟資料、勞動資料、日常生活資料等不同機構所蒐集的同一群體的資料，是以橫向聯合學習來訓練模型。

²⁸ 全名：The Institute of Electrical and Electronics Engineers

²⁹ 全名：Draft Guide for Architectural Framework and Application of Federated Machine Learning

3. **遷移式聯合學習(Federated Transfer Learning)**：資料集之間的個體重疊程度和欄位重疊程度都低者[23]。例如：用甲醫院所訓練好的肺癌影像辨識模型，再用乙醫院的乳癌影像資料訓練模型，則為遷移式聯合學習。

(三)使用者與其主要活動

使用者是指資料擁有者、協調者、模型使用者、稽核者等4類與聯合學習系統相關的成員；聯合學習若在商業需求下進行，各成員間可能有付費行為。[23]

1. **資料擁有者(Data Owner)**：主要活動為蒐集並準備訓練資料、參與模型訓練、在模型推論階段與其他成員交流中間運算結果。在整個參與過程必須確保資料隱私保護。
2. **協調者(Coordinator)**：主要活動為聯合學習系統功能(含演算法、基礎設備和服務等)之開發、運算活動(含模型訓練與測試、資料保全與隱私保護等)、模型管理(含模型訓練、測試與推論等)、行政管理(含提供服務、服務效能公布等)、資料管理(含資料集的說明資料之管理與釋出等)、經濟誘因設計(含計算資料擁有者和模型使用者的費用等)。
3. **模型使用者(Model User)**：主要活動是提出服務需求及使用服務，與協調者有商業往來關係。
4. **稽核者(Auditors)**：負責檢驗聯合學習程序的正確性及符合系統規範限制，包括：驗證資料來源的合法性、監控並記錄模型建立過程、要求協調者解釋資料管理、模型管理及經濟誘因策略、監督模型使用和要求使用者確認模型使用需求、確認聯合學習系統具有資料保全和隱私保護的流程和規範並確實遵循。

(四)系統架構

系統架構依聯合學習主要活動可分為5層，每層各有數項

功能模組(functional module)，摘要如下[23]：

1. **服務層(Service Layer)**：包含使用者服務模組、資料服務模組、參與者協調模組、工作管理服務模組。
2. **演算法層(Algorithm Layer)**：包含樣本比對模組(用於縱向聯合學習)、特徵比對模組(用於橫向聯合學習)、聯合特徵工程模組(可用於橫向、縱向和遷移式聯合學習)、聯合學習演算法模組、演算法評估模組(含表現、效能、隱私保護和資料保全)、個別資料擁有者的貢獻度評估模組、經濟誘因計算模組。
3. **操作層(Operator Layer)**：提供聯合學習演算法開發所需的基本操作、隱私保護和資料保全等工具模組。
4. **基礎建設層(Infrastructure Layer)**：將含加密資料單元所需的運算、儲存和通訊等元件以及各元件間或與操作層間的介面等，提供給操作層。
5. **跨層模組(Cross-layer Module)**：包含系統保全功能模組、法規與稽核模組等，以支援以上各層之間的互動。

(參)小結

美國、歐盟及國際組織因應 AI 發展，紛紛訂定策略或計畫，並提出實務作法；資料及 AI 模型共享為歐美政府的重要 AI 策略，民間組織亦有相關作法及平台形成。

歸納各國策略或作法，在訂定資料和 AI 模型共享作法時，應考量的重要原則或作法有：

- 一、資源(含資料及模型)共享必須以隱私受到保護、合法合理使用及安全無虞為前提。
- 二、為提高共享資源的可發現性、可近用性、互操作性、可再利用性(亦即 FAIR 原則)，必須在格式、詮釋資訊(metadata)、儲存庫用詞、模型訓練和驗證等方面，訂定一致化或標準化的規範，並且需要支持這些面向的技術研發，以促進資料科學和相關工

具創新。

- 三、需要建立完善的管理機制及規範，例如：來源追蹤、取用限制、隱私保護、資源保全、授權方式、智慧財產權分配等。
- 四、不同領域或疾病在檢查、檢驗或臨床等的需求會有差異，因此在資源共享的需求或作法也會有所不同，需要由相關專業的人員共組聯盟或工作小組的方式，在各個面向形成共識，以利建立有效的資源共享機制及配套。
- 五、共享的資料必須具規模、高品質且不偏頗(fairness)，共享的 AI 模型必須進行效能評估或驗證，才能讓使用者運用正確和無偏差的資料以及安全的 AI 模型進行增值利用，以促進後續發展出來的 AI 的可信賴度(trustability)。
- 六、整個 AI 資源共享參與成員／角色相當多，宜導入生態系概念，以利完善其規劃與運作。

肆、我國政府相關計畫或施政作為

(壹)臺灣 AI 行動計畫(2018-2021 年)

行政院 107 年 6 月 29 日核定，包括 AI 人才衝刺、AI 領航推動、建構國際 AI 創新樞紐、場域與法規開放、產業 AI 化等五大主軸行動計畫。其中**場域與法規開放行動計畫**將透過實證場域建立資料集，並將**資料流通、資料開放與介接平台**等納為推動重點；惟從其細部計畫來看，主要重點包括：自駕車、智慧綠能、無人機、多元車種混合車流、民生公共物聯網、智慧城鄉等，**尚無生醫健康領域之相關推動作法**[25]。

(貳)衛生福利資料科學中心

由衛生福利部統計處建置，在「保障個人健康隱私，促進健康資訊共享，減少資源重複投入」的核心價值下，收錄公共衛生(如：出生通報、死因統計、癌症登記、三高調查、出生世代調查、家庭與生育率研究調查、青少年健康行為調查、罕見疾病通報、人工生殖、癌症篩檢)、醫療保健資料(如：全民健保之明細、醫令、承保、重大傷病

等)、與健康相關之社會、經濟與地理資訊(如：國民健康訪問調查、中老年調查、吸煙調查、少年身心狀況調查、兒童及少年保護通報、婦女生活狀況調查、老人調查等)，期望透過**健康資料共享**，讓個別**健康資料**增值以產生具應用價值之**集體資訊**，以促進公共衛生決策品質、相關學術研究及醫療保健服務業等相關產業研發創新，並增進全民福祉[26]。

(參)健保資料人工智慧應用服務試辦計畫

衛生福利部中央健康保險署(簡稱健保署)為減少民眾就醫重複檢查，自 2018 年起鼓勵醫院將醫療檢查影像資料上傳到「健保醫療資訊雲端查詢系統」，截至 2020 年 8 月 31 日收錄以電腦斷層(CT)和磁共振造影(MR)為主的影像共約 23 億筆。為了促進健保資訊共享及社會創新服務，健保署自 2019 年 6 月至 2020 年 6 月試辦「全民健康保險資料人工智慧應用服務試辦計畫」，將醫療影像和相關就醫明細、檢驗檢查報告等申報資料提供給公務機關、學術研究機構(含產業)及專業機構申請，經「全民健康保險資料人工智慧應用管理審議會」審查通過的研究團隊，可至**實體隔離的「全民健康保險資料人工智慧應用服務中心」**，利用核准的資料進行 AI 應用研究[27, 28]。

(肆)國家級人體生物資料庫整合平台

我國於 2010 年立法通過「**人體生物資料庫管理條例**」，以規範人體生物資料庫之設置、管理及運用，保障生物資料庫參與者之權益，促進醫學發展，增進人民健康福祉(條例第 1 條)。該條例所定義的生物資料庫是指：為生物醫學研究目的，以人口群或特定群體為基礎，內容包括參與者之生物檢體、自然人資料及其他有關之資料、資訊；且其生物檢體、衍生物或相關資料、資訊為後續運用之需要，以非去連結方式保存之資料庫(第 3 條第 4 款)；該條例對於參與者的個資管理和商業運用之利益回饋，有嚴謹的規定。2020 年 5 月衛福部許可的人體生物資料庫共 33 個，分別由 29 家醫院、2 個研究機構(中央研究院、國家衛生研究院)、1 個政府單位(臺南市政府)和 1 家基金會(財團

法人演譯基金會)所建置[29]。

2019年10月衛福部委託國家衛生研究院建立的國家級人體生物資料庫整合平台正式成立，目標是：**建立標準檢體處理流程、建立一致性的檢體出庫品質管理流程、建立一致化生物資料庫數據、建立申請運用平台**。截至2020年11月，已有25個人體生物資料庫參與該平台，收案內容包括：肺炎、癌症(肝癌、消化道癌症、乳癌、前列腺癌、口腔癌等)、心血管疾病、糖尿病、慢性泌尿道疾病等[30]。

(伍)醫療影像之巨量資料建立與應用研究專案計畫

由科技部前瞻及應用科技司於2017年11月至2020年12月間推動(以下簡稱科技部醫療影像專案計畫)，補助國立臺灣大學、臺北榮民總醫院及臺北醫學大學分別結合醫療、AI及法律專長研究人員，進行醫療影像自動判讀AI演算法的開發，並將研發過程所建立的高品質醫療影像標註資料庫，提供學術研究機構再利用來進行醫藥衛生研究目的之AI模型研發。

為了共享資料庫，3組研究團隊也組成跨團隊法制工作小組，在符合個人資料保護法及人體研究法規範下，建立了資料共享之申請、審查及成果追蹤規範及配套文件，對資料申請案的審查項目包括：

1. 研究內容的學理根據、研究人員對資料讀能力等與研究可行性的相關項目。
2. 資料使用的合理性與安全性：評估資料使用與研究目的之間的關聯性、必要性和是否符合比例原則，並考量資料處理作業過程之安全性。
3. 互惠性：申請案對該專案醫療影像資料庫及其應用價值的附加貢獻。
4. 社會影響：評估對社會發展的正面與負面影響。若資料利用對於資料當事人之權利、與倫理道可能造成的衝擊，得要求申請者採取必要預防措施以降低衝擊。

截至2019年9月已建立腦、心、肺、骨骼、眼等疾病共25個資

料集合 520,874 個案例，其中 16 個資料集共 7,850 個案例自 2020 年起共享再利用[31]。資料集自 2020 年 4 月開始受理申請利用，截至 2020 年 11 月共受理 6 梯次、總計 16 件申請案，其中 11 件已審查通過，每件申請案需要與資料建置團隊簽署資料使用合約，自完成簽約起 1 年內可利用資料。

(陸)臨床資料庫與 AI 之跨域開發及增值應用計畫

由科技部生命科學研究發展司自 2020 年 7 月開始推動的 4 年期計畫，補助學界或醫院運用具備完善架構之各式臨床資料庫(如：高品質標註的醫療影像資料、標準化的醫療病歷等)，與具 AI 技術的科技公司合作，發展可商品化之臨床決策系統或分析輔助醫療應用平台[32]。該計畫原未規劃資料共享，**配合科技部整體 AI 推動策略，將參考科技部前瞻及應用科技司醫療影像專案計畫所建立的資料共享機制與規範。**

(柒)精準健康大數據永續平台計畫

由衛生福利部、科技部(生命科學研究發展司)及經濟部自 2021 年起推動的 4 年期計畫，以**跨部會分散式資料共享架構**，並將資料格式**統一化及規格化**，來建構精準健康大數據(分項一)，並利用此大數據進行轉譯研究及產業應用(分項二)。分項一所規劃資料項目包括[33]：

- 一、衛生福利部統計處資料科學中心整合健保署的全民健保資料庫和國民健康署的癌症登記、癌症篩檢及死亡檔等；
- 二、健保署的實驗室數據檔、CT/MRI 等影像與報告資料檔、癌症免疫療法臨床療效評估檔等；
- 三、財團法人國家衛生研究院的癌症、新興傳染性疾病、生物檢體等主題式資料庫；
- 四、衛生福利部建立癌症醫療次世代基因定序臨床資料，以新診斷晚期或術後復發癌症患者為收案標的。
- 五、科技部建置轉譯導向生醫巨量資料，聚焦癌症、心血管等疾病，以基因資料為核心，加上放射影像、病理影像、病歷等資料庫，

透過醫學中心前瞻式收案方式，每年新收案 2,000 例，上傳於財團法人國家實驗研究院國家高速網路與計算中心(簡稱國網中心)所建置的平台。

該計畫將以國內外公私合作聯盟的方式，建立大數據及標準化工作小組；並將由國網中心建置國家級生醫資料分享平台，收納各部會或醫院願意釋出的資料，結合資料、分析工具和計算資源形成友善的一站式服務網站，透過 VPN 和資料不落地機制，提供研究利用[33]。

(捌)小結

我國已有一些具規模的生醫資料庫及利用機制，但尚無具整體性的資料和模型共享策略及作法。

伍、科技部 AI 專案共享基礎建設之規劃

整個共享機制建立及未來的執行過程，須考量本報告第參(參)節所歸納出的重要原則或作法，包括：隱私保護與合理安全使用、符合 FAIR 原則、相關專業人員共組聯盟或工作小組、共享資源的可信賴度、導入生態系概念等。

參考上述原則和作法，本報告就 AI 共享基礎建設的共享項目、參與成員、共享機制、共享平台等，提出作法建議如下。

(壹)共享項目和參與成員

共享項目宜聚焦重點疾病診療與預防所需資料集和 AI 模型，由政府、學界及醫院共同建立。資料宜採集中式共享與分散式共享並存方式，並同步建立跨機構(各資料集的管理機構之間)AI 模型共享與聯合學習機制，以利分散式資料聯合建模產出具普適性的 AI 模型。如此建議主要是考量國內現況：目前我國個人資料保護法的相關規範及配套仍有模糊地帶，又無明確指引可供非法律專業人員利用[31]，因此對個人資料的利用，都以資料不離開原始蒐集機關(構)為原則，集中式資料共享較為困難。

一、共享資料的來源

(一)已建立且具規模的資料庫，例如：健保資料、電子病歷、檢

查與檢驗資料、人體生物資料庫等。這部分需要與資料管理機關協調，預期這些資料的共享將以分散式共享的可能性較高，需要同步建立跨機構聯合學習的機制。

(二)科技部重大生醫健康專案研究計畫：可鎖定於 2020 年以後啟動專案計畫，例如：轉型後的 AI 專案、精準健康大數據永續平台計畫等，自 2021 年起要求前瞻式收案資料、回溯性資料以及 AI 模型需釋出共享。

(三)搭配 AI 專案之領域應用研究，聚焦重要研究領域或疾病所需資料：掃描國際已有的重要資料庫或 AI 資源共享平台，參考已有成熟架構的平台(例如：美國 NIH 的 NCBI、美國的 All of US 計畫、ITU-T/WHO 的 FG-AI4H 等)所收錄的資料，甚至與之合作，來建立台灣自有資料集，此作法將可槓桿國際 AI 共享資源。

(四)針對重要研究領域或疾病，直接向學研醫界徵求可共享的資料：因為領域(或疾病)間具有歧異度，考量資源有限，因此必須聚焦重要領域或疾病，以極大化資源投入效益。共享資料可為集中式或分散式，由科技部補助必要經費，以利共享相關工作的執行，包括：資料和模型上架之標準化整備、個人資料利用之告知、共組聯盟或工作小組等。

二、共享 AI 模型的來源

(一)科技部 AI 相關專案所開發之 AI 模型共享：自 2021 年起要求相關專案(例如：AI 專案)將 AI 模型共享。

(二)AI 模型徵集：以計畫補助或競賽的方式，向產學研界徵集 AI 模型，於共享平台提供給其他醫療機構使用者試用，或供 AI 開發者進行介接與增值，提升 AI 模型可用性[34]。

(三)開源(open source code)模型蒐集：蒐集具指標性或重要性之開源模型[34]。

(貳)共享平台

由國網中心建立與維運資料和工具共享平台，並將聯合學習架構中所共同需要的基礎軟硬體環境，也納入平台範疇。

將資料、工具和運算資源串連於同一平台，提供一站式 AI 服務 [34]，是國際趨勢。國網中心在 2017 年至 2020 年間配合科技部醫療影像專案計畫的推動，建立醫療影像標註資料集共享平台及資料不落地利用機制 [31]；國網中心在前瞻基礎建設預算的挹注下，也建立了 TWCC 臺灣 AI 雲運算服務平台。未來的 AI 共享基礎建設平台宜在國網中心既有基礎上加以擴充，收錄集中式共享資料集和所有的共享模型，並串聯 TWCC 運算資源，以發揮資源投入乘數效果。

受限於個資隱私與資料保全課題，未來多數資料將以分散式方式透過聯合學習技術來訓練 AI；惟聯合學習系統的建構，需要考量應用情境而選擇適合的聯合學習類型，並確認各類成員的任務分配及架構完善的聯合學習系統。因此，國網中心平台可將不同領域在聯合學習架構中所共同需要的基礎軟硬體環境，也納入平台建置範疇，將有助於未來各項聯合學習專案之協調者們 (coordinators) 加快聯合學習系統的建置。這部分除了可參考 IEEE 所出版的聯合學習架構指引之外，也可借重臺灣人工智慧實驗室在影像判讀和基因分析的聯合學習平台建置以及利用多家醫院資料進行 AI 模型訓練的經驗。

(參) 共享機制

一、視資料和模型的來源和性質，定義共享模式類別，建立分級共享作法

資料共享可依資料機敏性，區分為：開放資料、機敏資料(需申請但免費使用、需申請且收費)、直接購買等類型。另外，集中式共享資料集，可參照科技部醫療影像專案計畫所建立的資料共享作法，由資料蒐集機構以委託共享資料平台管理資料的方式集中 [31]；無法集中於共享平台的資料集，則分散式儲存於原蒐集機構。機敏資料集中式共享和分散式共享的機制差異比較綜整如表 1。

表 1. 機敏資料自行使用、集中式共享和分散式共享的機制差異比較

考量項目		自行使用	資料共享(供他人使用)		
			分散式	集中式	
資料面： 資料治理	1. 隱私保護	(1) 去識別化	非必要	一階段 ^註	二階段 ^註
				非必要	建議納入
	(3) 使用情形回饋				
	2. 資料品質要求	V (品質可能有盲點)	V (品質有強化的可能性)	V (品質可強化)	
	3. 可發現性	X	X	V	
	4. 可近用性	X	X	V	
	5. 互操作性	X	V	V	
6. 可利用性	X	X	V		
7. 釋出前審核	X	X	V		
8. 釋出後追蹤	X	X	V		
技術面：聯合學習		X	V	X	
資訊安全：資料及 AI 模型參數的保全		V (強度較低)	V (強度較高，尤其是模型參數部分)	V (強度較高，尤其是資料部分)	

註：一階段去識別化係指資料蒐集機關將直接識別個資以刪除、假名化等方式處理；二階段去識別化除了含一階段去識別化之外，在資料集中後，依據個資再識別風險的評估結果，進行去識別化處理。

表 2. AI 模型共享機制初步規劃。

AI 模型類別	應用情境	近用作法	共享機制
基礎工具	模型開發	1. 開放原始碼 2. API	免費使用
應用模型	使用者試用、開發者交流	1. API 2. 聯合學習	1. 互惠共享 2. 相互評比後回饋意見
確效軟體	臨床使用(通過 TFDA 後可商業化)	1. 於雲端平台使用 2. 下載使用	1. 付費刊登於共享平台 2. 對使用者收費

資料來源：醫療影像 AI 平台規劃(簡報)[34]。

同樣的，AI 模型的共享也不必然是免費的，可由模型開發者視模型的性質界定為：免費使用(例如：影像分割或分類的基礎模型、提供使用者試用者)、有條件免費使用(例如：希望與其他開發者交流精進者)或收費使用(例如：已達臨床需求解決方案等級者)，初步規劃綜整如表 2。

二、資料和模型共享的參與者共組聯盟或工作小組，共同定義特定領域或疾病的標準化作法，並產出配套文件

資料面需建立資料標註、儲存格式、用詞、詮釋資料、資料品質檢核等之一致性或標準化作法；模型面需研議模型效能測試、驗證和評比方式、上架、功能說明文件等之標準化作法，以提高共享資源的可發現性、可近用性、可使用性、互操作性。ITU-T/WHO 的 FG-AI4H 的作法是很好的參考對象。

三、科技部設立 AI 共享資源管理組織，並制定各類共享模式之管理規範和配套文件

管理組織設立的主要目的是：確保受他人權利約束的資料及 AI 模型能在符合授權的情形下合理利用並保障安全。為達到這個目的，管理組織需要有 AI 技術、應用領域、法律、共享平台維運等領域的專家共同參與，以利建立適當的規範。

目前規劃管理組織分為 4 個工作小組，任務為[34, 35]：

1. 技術顧問小組：對資料和 AI 模型如何介接至共享平台和釋出、釋出風險管理、資料使用說明、模型功能說明、確效說明、資料和模型價值評估、品質管理等技術面共通規範和文件提供諮詢。
2. 法制顧問小組：對資料和 AI 模型共享有關的個資／隱私保護、釋出風險管理、資料和模型來源授權管理、共享成果歸屬和擴散等法制議題及共通規範和文件等提供諮詢。
3. 平台維運小組：共享平台系統建立、管理、調整、維護；平台使用者意見徵詢與回應。

4. 審查小組：特殊 AI 資源(如機敏資料集)提供使用之審議、特殊 AI 資源申請使用情況與成果追蹤。將以科技部醫療影像專案之資料管理審查委員會之運作機制為基礎，配合 AI 專案未來推動應用領域的多樣性，調整審查小組的委員組成及作業作法。

陸、參考文獻

- [1] 科技部。2020。科技部記者會新聞資料：智慧醫療人機協作新時代臨床輔助診斷精準且快速。4頁
- [2] NSTC. 2019. *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. The White House OSTP. 50pp.
- [3] <https://code.gov/>，2020年9月查閱。
- [4] OMB. 2016. *Federal Source Code Policy: Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software*. Office of Management and Budget, Executive Office of the President, USA.
- [5] NIH. 2018. *The NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability Initiative*.
- [6] OSTP. 2020. *American Artificial Intelligence Initiative: Year One Annual Report*. The White House OSTP. 39pp.
- [7] Do, N., R. Grossman, T. Feldman, N. Fillmore, D. Elbers, D. Tuck, R. Dhond e, L. Selva, F. Meng, M. Fitzsimons, S. Ajjarapu, S. Ayandeh, R. Hall, S.Do, M.Brophy. 2019. *Veterans Precision Oncology Data Commons: Transforming VA data into a national resource for research in precision oncology*. *Seminars in Oncology*, 46: 314-320.
- [8] <https://vpodc.org/>，2020年10月查詢。
- [9] <https://www.ncbi.nlm.nih.gov/>，2020年10月查詢。
- [10] NIH. 2018. *NIH Strategic Plan for Data Science*. National Institutes of Health, U.S. Department of Health & Human Services. 31pp.
- [11] <https://commonfund.nih.gov/commons>，2020年10月查詢。
- [12] <https://ncats.nih.gov/translator/about>，2020年10月查詢。
- [13] https://grants.nih.gov/grants/policy/data_sharing/，2020年10月查詢。
- [14] https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html，

- 2020年10月查詢。
- [15] EU. 2019. *Trustworthy AI: Joining efforts for strategic leadership and societal prosperity (brochure)*.
- [16] European Commission. 2020. *A European Strategy for Data*. 34pp.
- [17] European Commission. 2020. *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act)*. 42pp.
- [18] <https://elixir-europe.org/about-us> , 2020年10月查詢。
- [19] Salathé, M., T. Wiegand, M. Wenzel and R. Krishnamurthy. 2018. *White Paper of Focus Group on Artificial Intelligence for Health*. ITU-T. 7pp.
- [20] Wiegand T., R. Krishnamurthy, M. Kuglitsch, N. Lee, S. Pujari, M. Salathé, M. Wenzel and S. Xu. 2019. *WHO and ITU establish benchmarking process for artificial intelligence in health*. *The Lancet*, 394 (10192): 9-11.
- [21] <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/tg.aspx> , 2020年9月查詢。
- [22] <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/wg.aspx> , 2020年9月查詢。
- [23] IEEE. 2020. *Draft Guide for Architectural Framework and Application of Federated Machine Learning*. The Institute of Electrical and Electronics Engineers, Inc.70pp.
- [24]微眾銀行人工智能部、電子商務與電子支付國家工程實驗室、鵬城實驗室、平案科技、勝訊研究院、雲計算與大數據研究所、招商金融科技。2020。聯合學習白皮書v2.0。37頁。
- [25]行政院。2018。臺灣AI行動計畫(2018-2021年)核定本。86頁。
- [26] <https://dep.mohw.gov.tw/dos/np-2497-113.html> , 2020年11月查詢。
- [27]衛生福利部中央健康保險署。2020。健保25週年·健保資料人工智慧應用研討會手冊。36頁。
- [28]衛生福利部中央健康保險署。2019。全民健康保險資料人工智慧應用服務試辦要點。6頁。
- [29] <https://dep.mohw.gov.tw/doma/cp-3133-12824-106.html>
- [30]<https://nbct.nhri.org.tw/zhtw/index.php> , 2020年11月查詢。
- [31]林滋梅。2019。醫療影像專案推動成果及後續推動方向規劃。科技部聘用人員工作研究報告。36頁。

- [32]科技部生命科學研究發展司。2020。109年度臨床資料庫與AI之跨域開發及加值應用計畫徵求公告。科技部。4頁。
- [33]衛福部、科技部、經濟部。2020。精準健康大數據永續平台(簡報)。110年度科技預算Top-down計畫檢視及協商會議。28頁。
- [34]科技部。2020。醫療影像AI平台規劃(簡報)。
- [35]財團法人工業技術研究院。2020。醫療影像智慧應用推動辦公室第1次工作會議簡報。