



從ISO42001出發 討論我國產業發展AI 應採取之資料信任證明作法

◎ 洪亮瑄／資策會科法所副法律研究員

全球產業數位化，人工智慧（Artificial Intelligence, AI）也越來越廣泛地應用於各領域產業。但依據IBM公司於2024年1月發布的資料顯示，包含美國在內20個國家的43%企業認為生成式AI的訓練資料

不透明，難以信任其資料來源。¹ 國際貨幣基金組織的2023年調查報告指出已開發國家應優先發展安全、負責任的AI監管框架。² 國際亦重點發展「負責任的AI」。³ 為平衡AI的創新與管理，證明AI資料可信

1 IBM公司於2024年1月發布《2023全球AI使用現況》調查報告，該報告調查範圍為美國、加拿大、法國、德國、英國、印度、意大利、中國大陸、日本、新加坡、韓國、澳洲等20國的8584名IT經理或更高職務人員。參考：IBM，IBM發表《2023年全球AI採用指數》：生成式AI最快產生影響的企業用例—IT 自動化、數碼勞動力、客服，https://hongkong.newsroom.ibm.com/2024-01-17-IBM-2023-AI-AI-IT#assets_all（最後瀏覽日：2024/03/18）。

2 Kristalina Georgieva, *AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity*, <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity> (last visited Mar. 18, 2024).

3 The White House (2023), *FACT SHEET: Biden–Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety*, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/> (last visited Mar. 06, 2024).

度，本文試以2023年12月18日公布的全球首個AI國際標準「ISO/IEC 42001：2023（下稱：ISO 42001）」⁴為出發，提供我國產業資料管理建議。

ISO 42001跟資料管理的關聯性：強調資料歷程管理

ISO 42001旨於幫助組織負責任地研發、維護、持續改善AI系統，為提供或研發AI系統的供應商、合作夥伴及第三方提供AI管理系統的標準化流程。ISO 42001的規範架構採取與其他管理系統標準相同的共通性結構，第1至3單元為適用範圍、版本標示、名詞定義之說明；第4至10單元為管理要項，其架構可參照下圖1整理的ISO 42001之計劃、執行、查核、行動（Plan, Do, Check, Action, PDCA）管理循環架構。

ISO 42001將「提供或研發AI系統的供應商、合作夥伴及第三方的資料（documented information，或稱為『文件化資訊』）」，定義為：「組織需控管及維護的資訊，以及該資訊的媒介（medium），比如AI管理系統的相關流程、為組織運作而創建的資訊、AI管理系統所取得成果的紀錄（證據）等」。ISO 42001適用範圍為全球「提供或使用AI系統產品或服務的組織」，不論該組織的規模、類型（企業、公部門、非營利組織等）及組織所提供或使用的AI系統產品或服務。

為證明AI系統的適用性、完備性及有效性（suitability, adequacy and effectiveness），避免重複地發生高風險、

ISO 42001 之「PDCA」管理循環

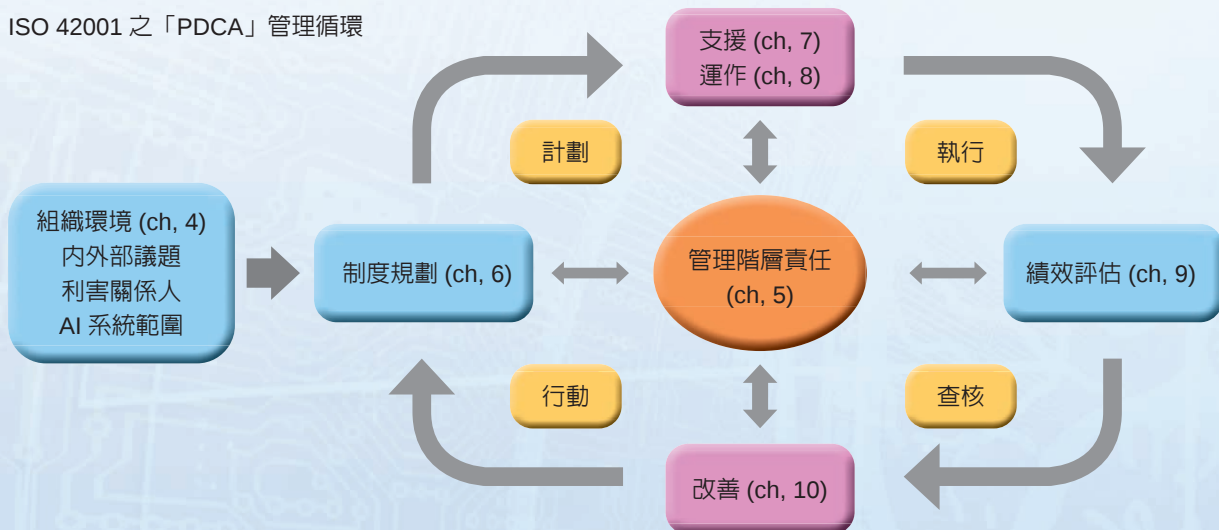
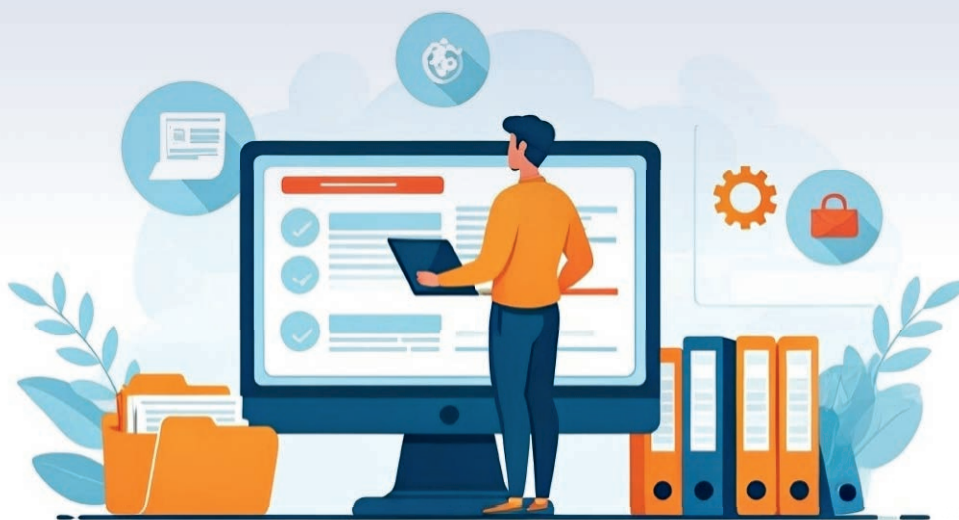


圖 1 ISO 42001 之「PDCA」管理循環架構。資料來源：本文整理

4 ISO/IEC 42001:2023, <https://www.iso.org/standard/81230.html> (last visited Mar. 06, 2024).



不合格的項目，ISO 42001國際標準強調透過「管理可信的AI系統（trustworthiness of AI system）生命週期所涉及的資料歷程」，⁵可以提升資料的可溯性（Traceability）、透明度及可信度，展現組織負責任地使用AI，確保AI資料品質。

為解決AI不受信任的課題，本文由ISO 42001國際標準所涉及的數位資料管理內容出發，得出ISO 42001與國內財團法人資訊工業策進會科技法律研究所創意智財中心於2021年7月發布的「重要數位資料治理暨管理制度規範（Essential Data Governance and Management System，下稱：EDGS）」採取相同的PDCA管理循環架構，且EDGS的數位資料管理範圍已涵蓋「AI資料」，並嘗試從AI軟體公司的角度，舉例說明資料管理循環的重點。

我國AI軟體公司除可以參考ISO 42001，亦可以先參考EDGS，該規範涵蓋了「AI生命週期資料的管理機制」，且採取相同的PDCA循環規範架構。

以下嘗試從AI軟體公司的角度舉例，帶入EDGS規範，說明AI軟體公司需要留意的管理作法。

AI軟體公司採取的資料管理循環

由於軟體相關研發及應用日新月異，應用於不同產業別的AI軟體所涉及的具體細項資料亦有不同。本文假設AI軟體公司透過AI分析無人機拍攝的影像，使AI產出「大型基礎設施建案的工程現場報告」，⁶並提供施工進度追蹤及資料品質管理平臺。

5 組織可考量以下因素，調整其需要記錄、管理的資料範圍：組織的規模及其活動、流程、產品及服務的類型；AI管理系統流程的複雜性；員工能力。

6 所謂大型基礎建設建案，比如高速公路、太陽能發電廠、機場和鐵路軌道、天然氣管道等。

由於AI的訓練資料不透明以及AI系統的管理作法不同於一般產業所應用的系統，使AI軟體公司難以證明其AI生命週期的資料可信度，因此本文將AI軟體公司帶入EDGS所採取的可信任AI的管理作法，並依PDCA循環分為四點：

一、P（計劃）：對應EDGS第四~六單元

（一）EDGS第四單元「組織環境」

「組織環境」代表公司需要構思與數位資料相關的內外部議題、利害關係人，才能夠擇定需要管理的數位資料。比如AI軟體公司需要權衡市場應用需求、公司規模及資源等，以評估研發可行的AI功能。

（二）EDGS第五單元「數位治理暨管理階層責任」

擇定需要管理的數位資料後，公司需要有具備統籌權利的高層人員訂定數位資料管理的政策目標及相應的業務執行規範，以確保執行成效。比如AI軟體公司的最高管理階層應制訂AI政策，於規劃、制定AI相關管理規範時，需要考量滿足組織整體政策目標、適用要求、承諾持續改善AI系統。為了讓導入EDGS的組織，可將EDGS與組織既有業務的管理流程、其他管理系統標準等進行整合，EDGS沒有限制組織必須採取何種具體的管理作法，最高管理階層須思考AI的管理作法能否與組織既有的管理系統標準整合，如ISO 9001（品質）、ISO 27001（資訊安全）等。



（三）EDGS第六單元「制度規劃」、「風險與機會因應」

1. 制度規劃

制度規範需要包含數位資料的標的範圍、業務流程等內容。為了讓AI提供可信任的「精確地監督施工進度」的功能，AI軟體公司選擇保存的數位資料為「使用AI軟體時，所涉及的建築工地上所有活動的完整資料證據鏈」，包含：個別工程的環境資料、建築資訊模型、人力資源資料、材料及設備清單、財務資料、安全檢查資料、工程預算資料、由前述資料整合而製作的線上視覺化互動圖表、AI生成覆蓋率

100% 施工區域的工程現場報告的PDF檔案以及AI的系統日誌（log）等。

2. 風險與機會因應

數位資料會受到可預期、非預期的風險影響，所以規劃組織制度時，組織需要提前思考可以處理的風險，才能因應其所造成的衝擊。比如AI軟體公司為減輕AI風險，公司最高管理階層須制訂並持續修訂AI風險標準，規範「可接受與不可接受的風險等級」、風險評估、風險處理、有效性評估以及AI系統影響評估流程，並保留前述歷程資料為證。

二、D（執行）：對應EDGS第七、八單元

（一）EDGS第七單元「支援」

管理流程的執行，需要透過人員教育訓練來宣導規範內容，以及設定人員接觸資料的權限。比如AI軟體的參數操作、功能增修，需經培訓確認AI軟體的經手人員具有職能。關於限制對數位資料的接觸權限，比如人員僅能透過AI軟體公司的資料品質管理平臺，並登入個別帳號取得資料接觸權限。

（二）EDGS第八單元「重要數位資料治理暨管理制度實踐流程」中的「數位資料之生成、保護與維護」

1. 數位資料之生成

為方便公司管理數位資料，需要能夠從檔案標題、檔案內文格式、儲存檔案的資料夾名稱等相關資訊，對應到特定的數位資料檔案或檔案類型。比如AI軟體公司的AI製作的線上視覺化互動圖表，具有固定的檔案內文呈現框架，其可以依其提供之建案進度追蹤服務，將圖表標題分為「建案概覽」、「地層狀況」、「基礎設施」、「樁（piles）的施工風險」、「施工時程表」等資料。⁷

2. 數位資料之保護與維護

由於難以發現數位資料的增刪修改痕跡，EDGS說明可以透過留存資料不同版本、設定資料的接觸權限等方式保護重要數位資料。

關於檔案版本留存，比如AI軟體公司的AI所製作的工程現場報告，不同時間點的施工進度，會產出不同內容的工程現場版本，透過於檔案註記追蹤施工的日期，可以標示不同版本。

三、C（查核）：對應EDGS第九單元「績效評估」

AI軟體特點為風險難以預見，為確保AI軟體可信任，公司應持續評估、處理AI

⁷ 為了避免土壤無法承壓，而導致建築傾斜或倒塌的風險，於工程施工階段，施工團隊會將混凝土、鋼等材料製成的「樁」打入地中，來補強土地的承載能力。施工風險因不同材質、不同結構的「樁」以及不同地質結構的土壤需要打入樁的深度等因素而不同。參考：臺灣大學，科學Online，基樁－建築穩固的基礎，<https://highscope.ch.ntu.edu.tw/wordpress/?p=70441>（最後瀏覽日：2024/03/11）。



AI軟體公司須持續評估、處理AI風險。
Photo Credit: <https://www.shutterstock.com>

風險。比如AI軟體公司需評估AI軟體訓練資料缺乏透明度、系統故障、對社會環境的影響等潛在風險，且須留存風險評估、風險處理相關資料，作為其執行的證明。

四、A（行動）：對應EDGS第十單元「改善」

AI的特點為持續學習、更新，當公司依據AI規範而使用AI、進行風險影響評估、風險處理及內稽後，將會持續透過機器學習等方式優化AI功能及調整原先的AI管理作法，並留存審查紀錄為證。

比如AI軟體公司為符合使用者需求及風險調適，一年至少進行上千次AI機器學習，且各建案建立獨立的「線上視覺化互動圖表」，精進相關資料集已超過數千萬個標籤（labels）。

關於AI軟體證明其可信任的管理作法，可得出要點在於管理數位資料歷程，有助AI軟體公司留存完整、真實、透明的AI資料證據鏈，彰顯公司對於可信任AI的研發、選用態度。

結語

AI系統面臨難以證明資料可信度、可用性的風險，建議業界亦可參考我國EDGS的「資料管理機制」發展「負責任的AI」，留意AI的風險評估、風險處理，以及數位資料的生成、保護以及維護作法，透過EDGS管理AI生命週期所涉及的數位資料，強化AI的資料可溯性、降低企業可能面臨的AI風險、增進跨域合作機會。●