

AI-Ready Data 詮釋資料框架指標指引

第1版

數位發展部

中華民國114年10月

目 錄

壹、目的.....	1
貳、應用範圍.....	3
參、作業原則.....	4
肆、參考資料.....	5
伍、定義.....	6
陸、核心架構總覽.....	8
(一)可查找性 (Findability) 構面-資料是否易於搜尋識別.....	8
(二)可近用性 (Accessibility) 構面-資料是否可穩定存取近用.....	10
(三)互通性 (Interoperability) 構面-資料是否利於系統整合.....	11
(四)再利用性 (Reusability) 構面-資料是否具持續應用價值.....	13
(五)可信任性 (Trustworthiness) 構面-資料是否具備信賴基礎.....	16
柒、評估流程與作業方式.....	18
(一)使用時機.....	18
(二)資料收集與前處理要求.....	19
(三)指標執行步驟 (自動/AI 輔助/人工)	19
捌、對應國際標準之建議與實務作法.....	26
玖、結語.....	29
拾、參考文獻.....	30
拾壹、附錄-衡量指標檢查表.....	32

圖目錄

圖 1 AI-Ready Data 詮釋資料框架指標構面示意圖	18
---------------------------------------	----

表目錄

表 1 名詞定義表	6
表 2 可查找性構面表	9
表 3 可近用性構面表	11
表 4 互通性構面表	12
表 5 再利用性構面表	15
表 6 可信任性構面表	17
表 7 指標使用時機對照表	18
表 8 資料品質處理建議表	19
表 9 各構面自動化檢核、AI 輔助及人工檢核對照表	20
表 10 相關國際標準對應表	26

壹、目的

順應近年人工智慧 (Artificial Intelligence, AI) 迅速發展趨勢，AI 技術日漸成熟與廣泛應用，已於公私部門帶來變革性的創新，並正重塑各領域的運作模式與服務型態。面對此波科技革新浪潮，國際間亦普遍認為應就 AI 技術研發、應用倫理、資料品質各面向發展需求挹注資源，以建構良善的 AI 生態系。

訓練資料是 AI 模型發展的重要元素，其品質優劣將直接影響模型的效能與應用可靠性。其品質優劣將直接影響模型效能與應用可靠性。不同類型的 AI 模型對資料需求各有差異，例如：監督式學習需仰賴標註一致且具代表性的資料，非監督式學習則依賴資料間的結構特徵，多模態模型則須確保跨語言、影像、聲音等異質資料間的對應與對齊。因此，若缺少關鍵的詮釋資料 (metadata)，可能導致模型訓練失準、偏差擴大或再利用受限。例如，缺乏時間與地理資訊的資料集，將影響模型對時序變化或區域差異的判斷；缺乏資料來源與標註方式的說明，則難以評估其適用性與可信度。強化措施可能包括：透過再加工補充外部來源資料、進行資料標註一致性檢核，或於應用端建立偏差調整機制。

資料之可用性與再利用價值，除仰賴於高品質、結構化與可機器處理 (machine-actionable) 等特性，完善的詮釋資料 (metadata) 亦不可或缺，不僅有助於資料提供方管理資料資源，更能協助資料使用者正確理解資料內容與使用限制。隨著 AI 應用需求日益提升，國際間也持續展開資料適用性討論，例如歐盟提出 FAIR 原則 (Findable, Accessible, Interoperable, Reusable)，聯合國教科文組織 (UNESCO) 與經濟合作與發展組織 (OECD) 亦強調資料在負責任 AI 發展中的重要角色，倡議資料應具備語意可解釋性與跨域整合能力，以支援 AI 模型訓練與應用。

依據近期國際趨勢見解，世界銀行提出「AI-ready development data」概念，指出資料應能被搜尋、理解、取得與使用，並建立於既有的開放資料及 FAIR 原則之上，同時強調需具備完善的 metadata、治理機制與技術

規格，才能支援 AI 系統與使用者的可信互動 (World Bank, 2023)。此外，美國商務部於「Generative AI and Open Data: Guidelines and Best Practices」文件中，也提出資料發布應兼顧 AI 系統自動擷取與理解的實務需求，以確保政府資料能成為可靠的 AI 訓練基礎 (U. S. Department of Commerce, 2025)。然而，目前國際間對於 AI 訓練資料整備程度尚無具體的衡量指標與技術規範，因此建立明確的詮釋資料框架與評估指標，對於促進 AI 訓練資料使用，具有關鍵意義。

為促進政府資料轉化符合 AI 應用需求，特參考國際間針對資料品質構面、治理趨勢與相關實務經驗，研擬「AI-Ready Data 詮釋資料框架指標指引」（下稱本指引），以評估資料之機器可讀性 (Machine Readability) 與跨資料集整合程度 (Degree of Cross-Dataset Integration)。資料本身具有中立性，其應用風險並非源於資料本體，而是來自特定應用目的下的使用與治理措施，因此，資料使用者應從 AI 應用的生命週期角度審慎評估資料來源、收集目的、結構特性與潛在偏誤，採適當運用。

本指引係為通用性之詮釋資料衡量架構，現階段主要針對結構化資料類型進行設計與描述，期建立標準化的評估準則，促進跨機關、跨領域 AI 訓練資料利用效益，然而各應用領域在資料類型、應用場景等具有相當之差異性，爰各領域主管機關得參考本指引為基礎，進一步發展具領域特性之詮釋資料框架指標，以強化指引落地之可行性。

貳、應用範圍

本指引之應用範圍，為個人、學校、團體、企業或政府機關等資料提供方，可參考本指引內容，完善資料集之詮釋資料，衡量資料作為AI訓練使用之整備情形。

參、作業原則

- 一、AI-Ready Data 詮釋資料框架指標涵蓋資料來源之透明度，良好的資料品質，包含標示資料來源、時間範圍及品質管控方式等要素，有助於提升資料可信度。考量訓練資料對於時效與趨勢分析的依賴，具備時序性、可累積性之資料集常態上更具利用價值，資料提供方透過提高更新頻率、明確標註資料時間或資料版本之作法，將有助於 AI 模型掌握時態變化及進行動態學習。
- 二、對外公開之資料內容原則建議採用 UTF-8 編碼格式，資料如儲存於對外可存取的網路空間，並透過 HTTPS (HyperText Transfer Protocol Secure) 協定提供下載服務，以確保資料傳輸的安全性。同時，宜明確闡述詮釋資料的內容及相關外部取得資訊，方便使用者查詢與連結下載。
- 三、資料提供方於資料產出過程，宜適當考量其資料屬性，優先採用結構化格式（如 CSV、JSON、XML 等），並統一欄位命名規則，同時搭配資料詞彙表 (Data Vocabulary) 與欄位說明文件，使資料易於理解與使用；結構化資料建議採用開放格式，或依數位發展部「共通性應用程式介面指引」提供應用程式介面 (API 或 Web Service) 服務，方便各界查詢與取用，達成機器對機器的識別、讀取及利用。
- 四、AI-Ready Data 詮釋資料框架指標可與結構化資料規則搭配，另行建構制定符合各領域需求、機器可讀的詮釋資料欄位，例如全球資訊網協會 (World Wide Web Consortium，以下簡稱 W3C) 推薦的資料目錄詞彙 (Data Catalog Vocabulary，以下簡稱 DCAT) 檢核規則，以提升資料的完整性與使用效率。
- 五、為權衡兼顧資料利用與隱私保護，AI-Ready Data 詮釋資料框架指標包含可信任指標的評估，以利瞭解資料產出適用範圍，促進適切合規利用。

肆、參考資料

本指引所採用之通用性詮釋資料標準框架，以行政機關電子資料流通實施要點為基礎，並參考國際廣泛使用之都柏林核心詮釋資料項目(DCMI Metadata Terms)中基礎的15個核心欄位之描述詞彙，以及 DCAT 之資料詮釋架構類別。

為確保詮釋資料標準框架具備通用性，亦參考國內 TWSMP3.0 詮釋資料標準(使用 ISO/TC211 標準)、國家圖書館與數位典藏詮釋資料的特性、定位與規範的設計架構。詮釋資料欄位填寫內容方面，則參考行政機關電子資料流通詮釋資料及分類檢索規範之分類詞彙和代碼、W3C XSD 資料型態規範、ESIP Data Readiness Cluster 規範、歐盟 FAIR 原則、及國際標準化組織(ISO)所訂定之語言代碼，確保資料集詮釋資料欄位定義明確。

伍、定義

本節列出本指引使用的名詞及其定義，如下表所示。

表 1 名詞定義表

英文名稱	中文名稱	定義
AI-Ready Data	為人工智慧作好準備的資料	經過優化、具備利於人工智慧分析應用特性的高品質資料
Machine Readability	機器可讀性	資料或詮釋資料採用「具有結構、可由軟體自動辨識並萃取特定資料」的檔案格式，使其內容可被自動化流程理解及處理之特性
Degree of Cross-Dataset Integration	資料的跨資料集整合程度	不同資料集之間能互通整合的程度，代表資料集的結構、識別碼、詮釋資料等採用共同標準，使不同資料集無須大量人工處理即可被系統正確理解及合併
Metadata	詮釋資料	描述資料集的資料
Dataset	資料集	可辨識的資料集合
Web Service	網路服務	透過網路提供功能的服務介面
Application Programming Interface	應用程式介面	經定義允許不同軟體或系統之間交換資料與提供功能的程式介面，簡稱 API
Data vocabulary	資料字典	記錄資料欄位名稱、格式、意義與關聯的文件
Findability	可查找性	指資料可被搜尋與識別的能力

英文名稱	中文名稱	定義
Accessibility	可近用性	指資料能被各種使用者和系統無障礙取得的程度
Interoperability	互通性	指資料利於跨平臺系統交換與使用的程度
Reusability	再利用性	指資料得依不同目的、平臺或情境重複使用的程度
Trustworthiness	可信任性	指資料具備可靠來源，於處理過程中保持透明與可追溯，提供偏差預防資訊，且符合隱私保護與倫理規範的程度

陸、核心架構總覽

AI-Ready Data 詮釋資料框架指標涵蓋5大主軸構面、14個面向指標，旨在強化資料結構的邏輯性與實務應用的可行性。本指引各構面所列衡量指標，依其檢核方式區分為「量化」及「量化/質性」2類：

量化：指標可透過數值化方式直接檢測，例如空值數量、更新頻率、欄位完整度等，適合以程式化工具進行自動化檢核。

量化/質性：指標同時具備可量化與需質性審查的面向，例如可信任性構面下隱私及資料保護、偏差預防、透明度等3項指標，可透過量化檢核相關內容是否存在，但仍需檢視質性內容以確認完整性與可理解性，通常需結合自動化檢核與人工審查並行。

本指引另於各構面中指標分類採資料品質相關性分級概念，其係用以區分衡量指標對於資料品質與 AI 應用之影響程度，分為「低」、「中」、「高」3類：

低：對資料品質之影響相對有限，若未明確標示，對整體資料完整性與可理解性影響不大，主要屬輔助性。

中：對資料品質具有中度影響，若能明確標示，將有助於增進資料的可判讀性與可用性，降低使用成本並改善應用效益。

高：對資料品質具關鍵影響，若能明確標示，對於確保資料的完整性、可信度與再利用價值具關鍵作用，為資料品質判斷之核心依據。

上述分類為通用性概念，後續各領域主管機關或資料提供方可依實際需求，與資料使用方共同協作調整分類或評級階層之數量、定義與細緻度，例如擴充為五級（Level 1 至 Level 5），以呈現更細緻的影響程度區分，俾利實務應用，以作為評估資料集是否符合 AI 應用需求的重要依據，其指引架構詳細說明如下：

（一）可查找性（Findability）構面-資料是否易於搜尋識別

發展人工智慧系統與服務的首要步驟，是能有效查找可用於訓練、驗證與測試的資料集。可查找性係指提升資料被使用者及電腦系統搜尋、識別與定位的機會，可透過建置資料目錄、充實詮釋資料內容等措施逐步落實。

在此基礎上，資料本體及其詮釋資料若能具備良好的可搜尋性，將更易於使用者與電腦系統發現與存取。其中，具備機器可讀性的詮釋資料在自動化資料探索與整合過程中扮演關鍵角色，能有效降低搜尋成本，提升資料重複利用的可能性，進而擴大資料價值。以下指標可作為評估資料集「可查找性」的重要依據。其構面下包含完整性（Completeness）、可發現性（Discoverability）等2個面向指標，其衡量指標及驗證方式說明如下：

1、完整性（Completeness）

完整性是指資料屬性欄位應盡量避免空值或缺漏，以確保資料在應用過程中具備足夠的資訊基礎。在詮釋資料完整性層面，可能出現該值不存在、確實存在卻未被記錄或不知該值是否存在等空值情形，可透過自動檢測確認資料的完整性，並以總值數及完整值的比例表示資料的完整性。完整性要求充分填寫各項描述欄位，避免留白或不具一致性的資料正規化，確保使用者或系統能準確理解資料的內容、來源與結構，並應視資料內容適當提供資料集大小，例如資料筆數、檔案數量、欄位數量等。高完整性的資料與詮釋資訊，有助於提升資料的可用性、準確性與在 AI 應用中的訓練價值。

2、可發現性（Discoverability）

可發現性是指能被使用者及電腦系統有效發現。其關鍵在於詮釋資料的描述品質，如描述越清晰且具結構性，資料就越容易被定位與識別，特別是運用標準化控制詞彙、關鍵字及一致命名規則，不僅可提升搜尋準確度，也有助於機器自動處理與比對，進而提升資料在跨系統與跨領域環境中的可查找性。

表 2 可查找性構面表

構面	指標	衡量指標 Metrics	質性/ 量化	計算 方式	資料品質 相關性
可查找性 Findability	完整性 Completeness	資料集中資料為空值的數量	量化	百分比	中
		詮釋資料欄位未填的數量	量化	百分比	中

構面	指標	衡量指標 Metrics	質性/ 量化	計算 方式	資料品質 相關性
		在詮釋資料填寫資料集大小	量化/ 質性	二元值	低
		詮釋資料提供資料集名稱/標題	量化/ 質性	二元值	低
	可發現性 Discoverability	詮釋資料填寫關鍵字欄位	量化	二元值	中
		詮釋資料填寫分類欄位	量化	二元值	中
		詮釋資料填寫資料收集時間區間資訊	量化	二元值	中
		詮釋資料填寫空間資訊	量化	二元值	中

(二)可近用性 (Accessibility) 構面-資料是否可穩定存取近用

當使用者成功查找到所需的資料後，接下來的關鍵就是能否順利取得並使用這些資料。可近用性意指降低使用者存取資料之難度，通常可透過採用開放、免費之存取機制與提供明確存取規範等措施加以推動。

在此基礎上，資料若能具備清楚的存取方式與穩定的下載管道，並搭配適當的認證與授權機制，將能有效降低使用者與系統存取資料的難度。可近用性直接影響 AI 系統獲取資料的效率與應用的可行性，是資料應用流程中的核心環節，以下指標可用於評估資料集在可近用性方面的表現，其構面包含可得性 (Availability) 1個面向指標，其衡量指標及驗證方式說明如下：

1、可得性 (Availability)

可得性 (Availability) 指的是使用者或電腦代理程式 (如 API 工具等) 是否能在無障礙、無錯誤或無不當限制的情況下，順利存取資料集內容。高可得性代表資料可穩定且持續地被查詢與下載，並附

有清楚明確的授權條件與存取規範，進而有助於提升資料的整體可用性與應用效率。

表 3 可近用性構面表

構面	指標	衡量指標 (Metrics)	質性/量化	計算方式	資料品質相關性
可近用性 Accessibility	可得性 Availability	提供資料下載網址	量化	二元值	高
		資料下載網址可近用	量化	二元值	高
		無需註冊即可下載資料	量化	二元值	中
		詮釋資料提供使用授權資訊	量化	二元值	高
		詮釋資料以受控制詞彙提供授權資訊	量化	二元值	中
		具備應用程式介面 (API) 或網路服務 (Web Service) 近用資料集	量化	二元值	中
		提供近用資料集之 API 說明文件	量化	二元值	中

(三) 互通性 (Interoperability) 構面-資料是否利於系統整合

在實際應用中，資料往往需與其他資料集進行整合，以順利介接至各類分析、儲存與處理的系統或工作流程。互通性意指便利資料與其他資料、應用或工作流程的整合分析、儲存與處理，通常可透過採用通用格式、標準化結構等措施加以推動。

在此基礎上，以下指標可作為評估資料集互通性的依據，其構面包含合規性 (Conformity/compliance)、機器可讀性/可處理性 (Machine readability/processability)、開放性 (Openness) 等3個面向指標，其衡量指標及驗證方式說明如下：

1、合規性 (Conformity/compliance)

合規性是指資料及其詮釋資料是否依循公認的標準規範，涵蓋資料的獲取、發布、描述及格式定義等層面。合規的資料應符合相關技術與語意標準，例如資料中的日期欄位應使用 ISO 8601格式，詮釋資料則應採用如 DCAT 等通用描述架構，以提升資料的一致性與可交換性，並提供資料字元編碼資訊，可透過自動化檢測機制查核資料是否符合標準，並以符合項目數占檢查項目總數的比例（百分比）檢視其合規程度。資料若具備高合規性，不僅有助於確保跨系統或跨平臺環境中的資料正確解析與整合，進一步提升資料可信度，屬於推動資料標準化與國際接軌的核心基礎。

2、機器可讀性/可處理性 (Machine readability/Processability)

機器可讀性/可處理性是指資料集及其詮釋資料是否具備良好的結構與格式，足以被自動化程序正確解析、理解與處理，具高機器可讀性、具明確欄位定義及一致的資料型態，並搭配結構化的詮釋資料，使資料能無需人工介入即可被各類系統或演算法直接使用。此一特性是實現 AI 模型訓練、資料交換與流程自動化的基礎條件。

3、開放性 (Openness)

開放性指資料宜以非專屬格式（如 CSV、JSON、XML 等開放格式）提供，以利資料使用方適當存取應用，並得視實際應用需求提供壓縮檔案，以提升取用效率。高開放性資料有助於降低使用門檻，促進資料流通與創新應用。

表 4 互通性構面表

構面	指標	衡量指標 (Metrics)	質性/量化	計算方式	資料品質相關性
互通性 Inter-operability	合規性 Conformity/ Compliance	資料／詮釋資料的日期格式符合規範	量化	二元值 /百分比	中
		資料／詮釋資料的EMAIL 格式符合規範	量化	二元值 /百分比	中
		資料字元編碼符合規範	量化	二元值 /百分比	中

構面	指標	衡量指標 (Metrics)	質性/ 量化	計算 方式	資料品質 相關性
	機器可讀性/ 可處理性 Machine readability/ Processability	檔案格式與媒體類 型的可處理性	量化	二元值	高
	開放性 Openness	檔案格式的開放性 與廣泛支援程度	量化	二元值	高

(四)再利用率 (Reusability) 構面-資料是否具持續應用價值

AI-Ready 資料集的核心目標之一，是使資料能在各類人工智慧系統與服務中被重複使用、靈活應用與有效整合。再利用率意指促進資料在不同情境或不同目的下的利用，通常可透過採用明確的授權條款、詳細記錄資料來源與使用限制等措施加以推動。

在此基礎上，若資料能具備清楚、完整且一致的描述，即可於不同平台、工具與應用情境下順利被複製、轉換或組合使用。高度可再使用的資料不僅可支援多元 AI 應用場景，也有助於模型的持續優化、跨任務應用以及知識的累積與延伸。以下指標可用於評估資料集在再利用率方面的表現，其構面包含及時性 (Timeliness)、一致性 (Consistency)、相關性 (Relevance)、可理解性 (Understandability)、可信度 (Credibility) 等5個面向指標，其衡量指標及驗證方式說明如下：

1、及時性 (Timeliness)

及時性指資料集本身及其詮釋資料是否保持更新，並能應對的實務應用需求，避免使用者基於過時資訊進行分析或決策。具備良好及時性的資料，有助於提升 AI 模型的準確性與實用性，特別是在需要反映及時狀態的應用情境中更為關鍵。

2、一致性 (Consistency)

一致性指的是資料與詮釋資料在內容與結構上宜注意保持邏輯一致，避免出現任何相互矛盾的資訊，常見的不一致情況包含資料集內出現重複或衝突的紀錄、詮釋資料中存在多個彼此矛盾的授權聲明，或欄位如「修改日期」早於「建立日期」等邏輯錯誤，資料提供方宜

適度採用自動化資料邏輯檢核機制，分析各欄位內是否僅使用控制詞彙表中包含的允許值、是否有重複資料，並依此計算出含非允許值及重複數值之百分比，並由人工輔以驗證是否具一致性。資料的一致性是其可信度與可用性的基礎，對於 AI 模型訓練與應用的穩定性至關重要。

3、相關性 (Relevance)

相關性係指資料內容與特定應用目的或任務需求之間的實質關聯程度。高相關性的資料能有效支援特定 AI 任務的目標，例如模型訓練、驗證或部署階段所需之資料需求。依據 OECD 所提出之 AI 系統生命週期治理框架，資料使用者應從系統開發與運作的全生命週期角度，審慎評估資料的來源、收集目的、結構特性與潛在偏誤，確保資料內容不僅符合應用語境，亦能避免在特定階段引入不必要的風險或誤導。例如，在模型設計初期應評估資料是否能代表欲解決的問題情境，在訓練與測試階段則須確認資料是否具備足夠的細緻度與標註品質，以支持預期的技術表現與公平性要求。資料筆數與欄位數確實需視應用脈絡而定，單純數量本身不必然代表品質，惟資料應具備足以支撐任務的適量資訊，另非僅以數量作為品質評估，資料之相關性，應涵蓋資料集內容與應用目的之間的合理對應，就應用情境、治理責任與資料本身特性進行綜合判斷。

4、可理解性 (Understandability)

可理解性指資料及其詮釋資料對使用者而言是否清晰、直觀且易於理解。若資料的結構、內容與說明具備良好的表達與邏輯性，使用者將能迅速掌握資料的意涵與用途，進而提升資料的可用性與應用效率。

資料提供方可透過檢視資料集的標題、欄位命名、關鍵字設定及說明文件的完整性與表達品質，強化資料的可理解性。良好的可理解性有助於使用者快速上手，降低誤用風險，促進資料的流通與再利用。

5、可信度 (Credibility)

可信度指的是資料是否提供使用者判斷資料內容是否真實且可靠的程度。當資料來自明確的來源、具備清楚的背景說明與標註，更容易取得使用者的信任。可信的資料不僅有助於提高 AI 模型訓練的品質，也能強化資料驅動決策的信賴基礎。

為提升可信度，資料提供方宜在詮釋資料中明確揭示資料發布單位、聯繫窗口、資料集擁有者及其職責，並標示資料來源與更新頻率，強化使用者對資料品質與責任歸屬的理解。

表 5 再利用性構面表

構面	指標	衡量指標 (Metrics)	質性/量化	計算方式	資料品質相關性
再利用性 Reusability	及時性 Timeliness	詮釋資料提供更新頻率	量化/質性	二元值	中
		詮釋資料提供資料上架日期	量化/質性	二元值	中
		詮釋資料提供資料最近修改日期	量化/質性	二元值	中
	一致性 Consistency	在資料/詮釋資料中非容許值的數量	量化	二元值/百分比	低
		不具重複資料	量化	二元值/百分比	低
		資料採用一致的計量單位、資料類型和參數名稱	量化/質性	二元值	高
	相關性 Relevance	資料筆數	量化/質性	計數/二元值	高
		資料欄位	量化/質性	計數/二元值	高
	可理解性 Under-standability	詮釋資料提供資料集背景目的說明	量化/質性	二元值	低
		詮釋資料提供資料集的資料字典/代碼手冊	量化/質性	二元值	低
	可信度 Credibility	詮釋資料提供聯絡窗口	量化/質性	二元值	低

構面	指標	衡量指標 (Metrics)	質性/ 量化	計算方式	資料品質 相關性
		詮釋資料提供資料集發佈者	量化/ 質性	二元值	低
		詮釋資料提供用戶回饋和建議的機制	量化	二元值	低

(五)可信性 (Trustworthiness)構面-資料是否具備信賴基礎

人工智慧雖為多個領域帶來顯著效益，其快速發展亦伴隨若干社會關切，包括隱私侵犯與監控風險、行為操控、不透明性以及資料與演算法偏見等問題 (Müller, 2023; United Nations System, 2022)，根據 Gartner 研究與多項國際文獻的建議，在資料應用及 AI 模型開發過程中，資料的來源資訊、描述方式與結構透明度均可能影響結果的可解釋性與信任程度。

可信性意指資料能在來源可靠、處理過程透明可追溯、並兼顧隱私保護與偏差預防等情形下得被使用者信賴，通常可透過揭示來源資訊、建立偏差揭露機制、強化資料透明度與責任歸屬等措施加以推動。在此基礎上，提升可信性可協助資料使用者理解資料特性，評估其代表性與適用性，並依應用脈絡進行必要的判斷與調整。高度可信的資料不僅能降低模型偏差與不當使用風險，也有助於提升 AI 系統的可解釋性與問責性，進而促進資料的負責任應用。

本指引在建構相關詮釋資料框架指標時，除聚焦技術與品質面向外，亦納入資料來源揭露與結構描述的要素，以協助資料使用者理解資料特性，審慎評估其代表性與適用性，並依應用脈絡進行必要判斷與調整，以強化資料應用的適切性與負責任的使用行為。本詮釋資料框架指標建議 AI-Ready 資料集宜涵蓋1份可信性內容說明文件，說明資料在隱私、偏差及透明性等面向的處理與揭露情形，包含以下3個面向：

1、隱私及資料保護 (Privacy and Data Protection)

說明是否涉及個人資料，是否依照個人資料保護法（以下簡稱個資法）及其子法或國際間通用資料保護規定（例如歐盟通用資料保護規則，General Data Protection Regulation, GDPR），落實相關保護機制與法規遵循措施。

2、偏差預防 (Bias Prevention)

明確揭示資料集中涉及人口變項（如年齡、性別、地區、族群等）之資訊，協助使用者理解資料的代表性與潛在偏誤來源。此舉有助於在資料早期階段即揭露可能影響 AI 模型公平性與判斷結果的結構性問題。同時，建議資料使用者在發展 AI 系統或服務時，應審慎評估資料特性可能帶來的限制與風險，並積極進行偏差識別與緩解措施，降低資料偏差對模型預測結果的影響。

3、透明度 (Transparency)

透明度係指資料集的收集、處理、發布過程是公開且可驗證的，高度透明的資料集能增加使用者的信任，並有助於 AI 系統的可解釋性和可問責性。資料提供方宜清楚說明資料的來源、收集方法、處理流程、更新頻率等資訊，同時建立資料使用追蹤機制，以確保資料的使用符合原始目的倫理與法規規範。

表 6 可信任性構面表

構面	指標	衡量指標 (Metrics)	質性/量化	說明文件
可信任性 Trustworthiness	隱私及資料保護 Privacy and Data Protection	資料經過隱私保護程序之說明	量化/ 質性	因應人工智慧及資料分析技術的快速發展，建議資料提供方 提供1份說明文件 針對資料集於隱私保護、偏差預防及透明度3個構面，提出相應的指標說明，以確保資料的合規使用並降低潛在風險。
	偏差預防 Bias Prevention	偏差預防說明(提供正確描述之資料集內容，包含時間範圍、區域、類型等，減少資料集可能存在的偏差)		
	透明度 Transparency	資料集透明度說明(提供資料的來源、收集方法、處理流程、更新頻率等資訊)		

柒、評估流程與作業方式

為協助資料提供及使用雙方在資料釋出與利用過程中，初步瞭解其資料品質與結構特性是否具備支援 AI 應用之潛力，本節說明 AI-Ready Data 詮釋資料框架指標的評估流程與作業建議。內容涵蓋評估的啟用時機、資料預備要件、執行方法到成果彙整方式，提供一套可資參考之標準作業流程，供機關依實際需求採用與調整，作為資料品質管理與應用準備之輔助工具，進一步推動制度化作業與實務應用之銜接。資料提供方得依領域專業，使用衡量指標檢查表（如附錄）進行各構面下指標逐項評估，俾利落實追蹤資料優化補強歷程，AI-Ready Data 詮釋資料框架指標構面示意圖如下圖1。

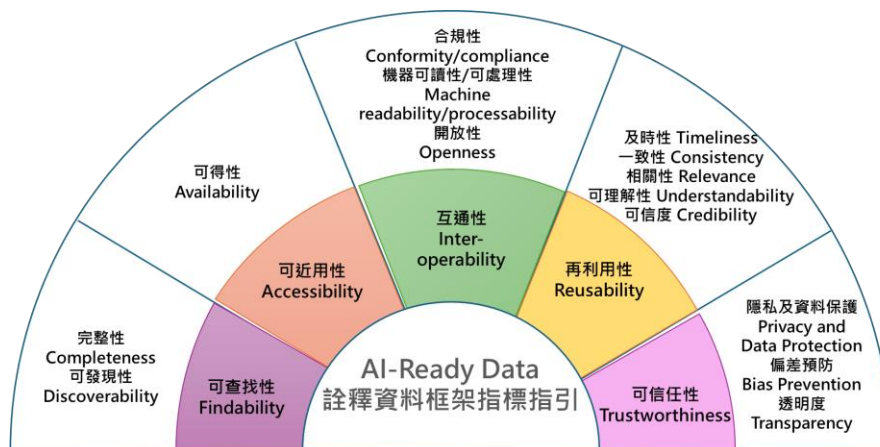


圖 1 AI-Ready Data 詮釋資料框架指標構面示意圖

(一)使用時機

推動 AI-Ready Data 詮釋資料框架指標進行資料評估實務中，清楚界定「何時應進行評估」可導入作為內部流程的一環，避免資料釋出後再補救的效率損失與風險。建議針對以下2情境執行 AI-Ready Data 評估：

表 7 指標使用時機對照表

使用情境	目的	建議頻率	實例情境
資料上架預檢	確保符合 AI 應用基本門檻與品質規範	每次資料上架或更新	資料首次產出、重大內容調整後、例行資料統計

使用情境	目的	建議頻率	實例情境
定期品質盤點	作為資料年度品質管理之依據	每年或每半年一次	年度資料盤點

(二)資料收集與前處理要求

為提升資料的自動檢核效率，需在評估前完成基礎的資料標準化處理，以確保評估結果客觀且具代表性為重要前提。以下為各面向之處理建議：

表 8 資料品質處理建議表

資料面向	資料品質面向	處理建議說明
資料結構與格式	資料欄位	建立清楚命名原則、資料型別、單位一致、欄位定義文件（資料字典）
	編碼規範	使用國際通用之標準編碼（例如 UTF-8 編碼格式、日期使用 ISO 8601 格式等）
資料內容與語意	缺漏值	建立標示缺值（如 NA、null）之標準化處理規則
	控制詞彙	採用控制詞彙（如 Chinese National Standards, CNS）、都柏林核心詮釋資料（Dublin Core Metadata Initiative, DCMI）或自建詞彙表控制描述語意
資料機敏風險	隱私及資料保護	涉及個資資料應提供適當隱私保護措施與保護說明

(三)指標執行步驟（自動/AI 輔助/人工）

本詮釋資料框架指標評估建議兼採自動化檢核、導入 AI 輔助與人工審查等 3 種模式，以提高效率並兼顧評估品質。首先，可搭配自動化檢核工具，進行結構化規則比對、欄位格式檢查、自動欄位對應分析等，以提升可程式化檢核項目之評估效率；其次，針對需進一步優化或補齊之詮釋資料內容，可導入 AI 輔助機制，由 AI 工具先依據預設指令（prompt）產出檢測提示（附帶評估係數或符合指標程度說明），若 AI 輔助評估結果高於使用情境之預設門檻，則完成評估或另採抽樣確認，若 AI 輔助評估結果低

於使用情境之預設門檻，則逐筆續交由人工進行覆核，並衡量建立通案性評估基準；針對語意一致性、描述完整性、可信任性及涉及資料敏感相關內容，建議仍應由專責人員進行審查，以確保資料之正確性與適用性。

表 9 各構面自動化檢核、AI 輔助及人工檢核對照表

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
可查找性 Findability	完整性 Completeness	1. 資料集中資料為空值的數量	可透過 Python 搭配 JSON Schema Validator 或其他套件工具組合，檢核欄位完整程度	導入 AI 輔助輸出檢測提示，指出欄位缺漏或完整性狀況	確認 AI 提示是否符合實際資料狀態
		2. 詮釋資料欄位未填的數量			
		3. 在詮釋資料填寫資料集大小			
	可發現性 Discoverability	1. 詮釋資料提供資料集名稱/標題		導入 AI 輔助輸出檢測提示，標示標題與關鍵字的可辨識度或缺漏	檢視是否符合專業語境與使用需求
		2. 詮釋資料填寫關鍵字欄位			
		3. 詮釋資料填寫分類欄位			

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
		4. 詮釋資料 填寫資料 收集時間 區間資訊 5. 詮釋資料 填寫空間 資訊			
可近用性 Accessibility	可得性 Availability	1. 提供資料 下載網址 2. 資料下載 網址可近 用 3. 無需註冊 即可下載 資料 4. 詮釋資料 提供使用 授權資訊 5. 詮釋資料 以受控制 詞彙提供 授權資訊 (泛指詮 釋資料所 使用詞彙 是透過經 整理之受 控制的權 威詞彙) 6. 具備應用 程式介面	可透過網 址驗證工 具(例如 curl、 wget)檢 核資料 下載是 否有效	AI 產出檢 測提示， 指出 API 或授權資 訊是否完 整	人工確認 授權正確 性與近用 合理性

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
		(API) 或網路服務 (Web Service) 來近用資料集			
		7. 提供近用資料集之 API 的說明文件			
互通性 Interoperability	合規性 Conformity / Compliance	1. 資料／詮釋資料的日期格式符合規範	可透過自訂正規式表達檢查工具檢查資料描述合理性	AI 產出檢測提示，標示格式異常或不一致之處	人工覆核特殊格式或領域慣例
		2. 資料／詮釋資料的 EMAIL 格式符合規範			
		3. 資料字元編碼符合規範			

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
	機器可讀性/ 可處理性 Machine Readability / Processability	1. 檔案格式 與媒體類 型的可處 理性	可透過資 料格式驗 證工具檢 查 (如 CSV JSON XML)	AI 產出檢 測提示， 指出是否 為常見可 處理類 型，以及 應用領 域是否 廣泛支 援	人工確認 格式開 放性與 實務適 用性
	開放性 Openness	1. 檔案格式 的開放性 與廣泛支 援程度			
再利用性 Reusabil- ity	及時性 Timeliness	1. 詮釋資料 提供更新 頻率	可利用時 間設定函 式自動比 對時間戳 記	AI 產出檢 測提示， 提示更新 週期是否 與預期一 致	人工判斷 檢查資料 提供歷程 合理性 (例如統 計月報應 以每月為 更新週 期、資料 上架日期 與最近 修改日期 不應為未 來時間等)
		2. 詮釋資料 提供資料 上架日期			
		3. 詮釋資料 提供資料 最近修改 日期			

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
	一致性 Consistency	1. 在資料／詮釋資料中非容許值的數量	可透過資料值檢查工具(例如 Open-Refine、Pandas)，檢核資料空值、重複及一致程度	AI 產出檢測提示，標示數值或單位不一致的情況	人工個案研判內容是否異常
		2. 不具重複資料			
		3. 資料採用一致的計量單位、資料類型和參數名稱			
	相關性 Relevance	1. 資料筆數	可利用自件對資料筆數與欄位數	AI 產出檢測提示，指出資料筆數或欄位與應用需求的匹配度	人工判斷是否符合實際使用場景
2. 資料欄位					
可理解性 Understandability	1. 詮釋資料提供資料集背景目的說明	可透過布林值 (Boolean) 條件自動檢核是否有填寫	AI 產出檢測提示，指出資料筆數或欄位與應用需求的匹配度	人工判斷是否符合實際使用場景	
	2. 詮釋資料提供資料集的資料字典/代碼手冊				
可信度 Credibility	1. 詮釋資料提供聯絡窗口		AI 產出檢測提示，標示聯絡與回饋	人工確認資訊正確性與可追溯性	

構面	指標	衡量指標	自動化檢核建議	AI 輔助建議	人工檢查建議
		2. 詮釋資料 提供資料 集發佈者		資訊是否完備	
		3. 詮釋資料 提供用戶 回饋和建 議的機制			
可信任性 Trustworthiness	隱私及資料保護 Privacy and Data Protection	1份質性文件 針對資料集 於隱私保 護、偏差預 防及透明度 提出說明， 以利資料合 規使用並降 低潛在應用 風險	可透過布林 值 (Boolean) 條件自動檢 核是否有填 寫	AI 產出檢測 提示，指出 隱私保護文 件是否說明 隱私保護做 法、標示資 料分布或樣 本是否可能 偏差、透明 度資訊是否 完整 (來源 、限制、 處理方法)	建議人工檢 視質性文 件，並依據 資料應用目 的，確認資 料內容是否 足具代表性 且偏誤風險 可控
	偏差預防 Bias Prevention				
	透明度 Transparency				

捌、對應國際標準之建議與實務作法

為強化資料品質一致性與跨域整合能力，AI-Ready Data 詮釋資料框架指標參考相關國際資料治理與品質標準，有助於跨平臺、產業間的資料整合與合作，本節整理國際常見資料標準與本指標間的對應關係，供作制度設計與內控規範參考。

表 10 相關國際標準對應表

國際標準	核心內容／規範重點	對應 AI-Ready data 詮釋資料 框架指標構面	對應建議與實務作法
FAIR 原則	Findable, Accessible, Interoperable, Reusable 四大原則為資料可被 AI 與人理解與利用之基礎	可查找性、可近用性、互通性、再利用性	導入資料標準欄位（如標題、分類、授權）、標示 API、時間與空間欄位
W3C DCAT (v2)	國際資料集詮釋資料架構標準，支援 RDF/JSON-LD 語意結構	可查找性、可近用性、互通性	使用 dcat:distribution, dct:license, dct:temporal 等標準欄位結構
DCMI Core	Dublin Core Metadata Initiative，定義通用的15個資料描述欄位	再利用性（可理解性、可信度）	使用 title, creator, subject, date, identifier 等欄位描述
ISO 8000	資料品質標準，涵蓋準確性、完整性、一致性、唯一性、語意清晰等八大維度	再利用性（及時性、一致性）、互通性（合規性）	檢查欄位格式、空值、重複紀錄、欄位說明，建立資料一致性與準確性驗證程序

國際標準	核心內容／規範重點	對應 AI-Ready data 詮釋資料 框架指標構面	對應建議與實務作法
ISO/IEC 11179	資料元素管理與語意控制標準，用於定義欄位名稱、單位、定義、值域等	再利用性（可理解性、一致性、可信度）	建立機關層級資料欄位資料字典（Data Element Dictionary），提供欄位定義與屬性說明
GS1 標準	商品與物流資料交換國際標準，如 GTIN、GLN、GPC 等分類與資料格式	互通性（開放性）、再利用性（一致性）	導入 GS1 分類碼、標準欄位命名（如商品條碼）、資料格式（如 EPCIS XML）
ISO 8601	國際日期時間格式標準，例如：YYYY-MM-DD	互通性（合規性）、再利用性（及時性）	日期欄位統一使用 YYYY-MM-DD 或 RFC 3339 格式（如 2024-06-01T12:00:00Z）
ISO 639-1 / 639-3	語言標準代碼，支援多語系資料描述	可查找性、互通性	zh, en, ja 等標準代碼，建議於「資料語言」欄位中統一使用
GDPR / 個資法	歐盟通用資料保護規則（GDPR）及我國個資法，規範個人資料的合法使用與匿名化處理方式	可信任性（隱私保護、偏差揭露、透明度）	資料若涉及個資，需進行隱私強化處理並提供處理說明
ISO/IEC 27001	資訊安全管理標準，強調資料保護、權限控管、存取紀錄	可近用性、可信任性（透明度）	對於 API 存取設立身分驗證、紀錄存取行為，標示授權條款、敏感資料等級
Schema.org / OpenAPI	資料描述與 API 說明標準，支援機器可讀性與自動化發現資料	互通性（機器可讀性）、可近用性、再利用性	提供 OpenAPI 格式說明文件、JSON Schema 定義欄位型態與限制條件

國際標準	核心內容／規範重點	對應 AI-Ready data 詮釋資料 框架指標構面	對應建議與實務作法
ESIP Data Readiness Checklist	評估資料集 AI 可用程度的檢核清單（由美國地球科學資訊合作組織提出）	可查找性、互通性、再利用性（可理解性）、可信任性	檢查是否標示來源、分類、欄位說明、資料更新頻率、偏差與使用限制說明

為落實上開標準與建議於實務操作，建議資料集在描述層面採用 DCAT 與 DCMI 的結構化格式，強化詮釋資料的一致性與語意互通性；在欄位屬性管理方面，可對照 ISO/IEC 11179 與 ISO 8000 建立資料字典，明確定義欄位名稱、資料型別、允許值與單位，提升資料一致性與準確性；資料品質評估則可參照 ISO 8000 的八大品質維度與 ESIP Data Readiness Checklist，進行結構完整性與使用適用性的檢核；在格式與語意標準方面，建議全面導入 GS1 商品分類、ISO 8601 日期格式與 ISO 639-1 語言代碼，確保資料格式統一並具備跨域可交換性；最後，在授權與隱私管理上，應審慎對應 GDPR、我國個資法及 ISO 27001 等標準，於資料集中清楚標註授權條款、機敏處理與存取紀錄，以強化資料合規性與使用信任度。

玖、結語

AI-Ready Data 詮釋資料框架指標指引旨在協助政府機關以系統性檢視結構化資料在人工智慧應用中的整備程度，涵蓋資料可查找性、可近用性、互通性、再利用性與可信任性等關鍵面向，提供具操作性的衡量準則評估框架，適用範圍不限於開放資料，亦可擴及各類供 AI 訓練應用之結構化資料，協助資料提供方與資料使用方評估及提升資料作為 AI 應用之整備情形。

隨著 AI 技術快速演進，資料型態日益多元，未來的 AI 應用將不僅侷限於結構化資料，更廣泛依賴文字、影像、音訊、影片等多模態資料，以及報告文件、圖像檔案等非結構化內容。本指引為通用性衡量架構，後續可由各領域主管機關參考本指引為基礎，進一步發展具領域特性之詮釋資料框架指標，以強化指引落地之可行性，並可因應 AI 的發展，持續延伸與擴充框架指標，以及鼓勵各機關納入相關資料管理政策，促進跨域資料流通與再利用。

拾、參考文獻

- ESIP Data Readiness Cluster (2022). Checklist to examine AI-readiness for open environmental datasets. Retrieved from <https://doi.org/10.6084/m9.figshare.19983722.v1>
- Fraunhofer FOKUS (2020). Definition and description of data quality indicators and metrics. Data Quality Guidelines. Retrieved from <https://joinup.ec.europa.eu/sites/default/files/solution/documentation/2020-08/PRJ004%20-%20ISA%20Fraunhofer%20Fokus%20-%20Data%20Quality%20Guidelines%20-%20Deliverable%20B1%20-%20v1.2.pdf>.
- Fraunhofer FOKUS (2020). Recommendations for providing high quality data. Data Quality Guidelines. Retrieved from <https://joinup.ec.europa.eu/sites/default/files/solution/documentation/2020-08/PRJ004%20-%20ISA%20Fraunhofer%20Fokus%20-%20Data%20Quality%20Guidelines%20-%20Deliverable%20C1%20-%20v1.3.pdf>.
- GO FAIR (2022). FAIR Principles. Retrieved from <https://www.go-fair.org/fair-principles/>
- Greenberg, J. (2024, February 23). AI-Ready Data: Navigating the Dynamic Frontier of Metadata and Ontologies. Metadata Research Center. <https://mrc.cci.drexel.edu/2024/02/23/ai-ready-data-navigating-the-dynamic-frontier-of-metadata-and-ontologies/>.
- Hiniduma, K., Byna, S., & Bez, J. L. (2024). Data Readiness for AI: A 360-Degree Survey (arXiv:2404.05779). arXiv. Retrieved from <http://arxiv.org/abs/2404.05779>.
- Madiaga, T. A. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. Think Tank, European Parliament.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf).

Müller, V. C. (2023). Ethics of artificial intelligence and robotics. The Stanford Encyclopedia of Philosophy (Fall 2023 Edition). In Edward N. Zalta & Uri Nodelman (eds.). <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=ethics-ai>.

Publications Office of the European Union. (2021). Data.europa.eu data quality guidelines. Publications Office of the European Union. Retrieved from <https://data.europa.eu/doi/10.2830/333095>.

Rotaru, T. Ş., & Amariei, C. (2023). Ethical Issues in Research with Artificial Intelligence Systems. In M. Radenkovic (Ed.), Ethics - Scientific Research, Ethical Issues, Artificial Intelligence and Education. <https://www.intechopen.com/chapters/1127065>.

United Nations System (2022). Principles for the ethical use of artificial intelligence in the United Nations System. https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.

U. S. Department of Commerce. (2025). Generative AI and Open Data: Guidelines and Best Practices. Retrieved from <https://www.commerce.gov/sites/default/files/2025-01/GenerativeAI-Open-Data.pdf>.

World Bank. (2023, June 26). From open data to AI-ready data: Building the foundations for responsible artificial intelligence. World Bank Blogs. Retrieved from <https://blogs.worldbank.org/en/opendata/from-open-data-to-ai-ready-data--building-the-foundations-for-re>.

拾壹、附錄-衡量指標檢查表

構面	指標	衡量指標	評估結果	
可查找性 (Findability)	完整性 (Completeness)	1. 資料集中資料為空值的數量		
		2. 詮釋資料欄位未填的數量		
		3. 在詮釋資料填寫資料集大小		
	可發現性 (Discoverability)	1. 詮釋資料提供資料集名稱/標題		
		2. 詮釋資料填寫關鍵字欄位		
可近用性 (Accessibility)	可得性 (Availability)	3. 詮釋資料填寫分類欄位		
		4. 詮釋資料填寫資料收集時間區間資訊		
		5. 詮釋資料填寫空間資訊		
		1. 提供資料下載網址		
		2. 資料下載網址可近用		
		3. 無需註冊即可下載資料		
		4. 詮釋資料提供使用授權資訊		
互通性 (Inter-operability)	合規性 (Conformity / Compliance)	5. 詮釋資料以受控制詞彙提供授權資訊		
		6. 具備應用程式介面(API) 或網路服務(Web Service) 來近用資料集		
	7. 提供近用資料集之API的說明文件			
機器可讀性/可處理性 (Machine readability / Processability)	開放性 (Openness)	1. 資料/詮釋資料的日期格式符合規範		
		2. 資料/詮釋資料的EMAIL格式符合規範		
再利用性 (Reusability)	及時性 (Timeliness)	3. 資料字元編碼符合規範		
		1. 檔案格式與媒體類型的可處理性		
		1. 檔案格式的開放性與廣泛支援程度		
	一致性 (Consistency)	相關性 (Relevance)	1. 詮釋資料提供更新頻率	
			2. 詮釋資料提供資料上架日期	
			3. 詮釋資料提供資料最近修改日期	
	可理解性 (Understandability)	可信度 (Credibility)	1. 在資料/詮釋資料中非容許值的數量	
			2. 不具重覆資料	
	可理解性 (Understandability)	可信度 (Credibility)	3. 資料採用一致的計量單位、資料類型和參數名稱	
			1. 資料筆數	
2. 資料欄位				
可理解性 (Understandability)	可信度 (Credibility)	1. 詮釋資料提供資料集背景目的說明		
		2. 詮釋資料提供資料集的資料字典/代碼手冊		
		1. 詮釋資料提供聯絡窗口		
可信任性 (Trustworthiness)	透明度 (Transparency)	2. 詮釋資料提供資料集發佈者		
		3. 詮釋資料提供用戶回饋和建議的機制		
		1份質性文件針對資料集於隱私保護、偏差預防及透明度提出說明，以利資料合規使用並降低潛在應用風險。		