

政府資料品質提升機制運作指引

數位發展部

中華民國 112 年 1 月

壹、前言

一、目的

為提升政府資料開放內容之正確性、易用性、即時性及採適當格式開放，使政府資料開放邁向結構化，爰訂定本運作指引，以期協助各機關建立資料品質概念、逐步提升政府資料之可用性。

二、適用對象

行政院及所屬各級機關（以下簡稱各機關）適用之；行政院以外之其他政府機關、公營事業機構、公立學校及行政法人，得準用；地方政府得參照。

三、適用範圍

本運作指引以符合國際「開放定義」（Open Definition）之政府資料，範圍含括各機關於職權範圍內取得或做成，且依法得公開之各類電子資料，包含文字、數據、圖片、影像、聲音、詮釋資料（metadata）等。

貳、資料品質提升流程介紹

一、資料品質相關標準

以資料使用流程為出發點，分為資料可直接取得、資料易於處理、資料易於理解三個面向：

（一）資料可直接取得

包含2項指標：

1. 資料資源連結有效性：資料資源連結可回傳連結成功狀態（HTTP Status Code）。
2. 資料資源可直接下載：使用者能透過資料資源連結直接獲取資料，無需透過登入或任何額外的操作形式。

（二）資料易於處理

就資料描述方式可區分為結構化與非結構化2種形式。

1. 結構化資料：

指經分析後可分解成多個互相關聯的組成部分，各組成部分間有明確的層次結構：

（1）固定欄位結構化資料：

單一行 (Row) 標題的表格式資料，每筆資料的欄位數均相同，且無合併儲存格、無公式、無空行、無小計等。

(2)非固定欄位結構化資料：

每筆資料欄位非固定之結構化資料，符合 W3C 之 XML、JSON、JSON-LD 標準等結構化資料。

2. 非結構化資料：

排除結構化資料以外均為非結構化資料。例如一般文字文件、圖檔、影音等。

為利資料易於處理，如資料型態允許，應優先提供固定欄位結構化資料。

(三)資料易於理解

針對結構化之資料資源，於詮釋資料須依「資料集詮釋資料標準規範」提供資料資源之編碼及其主要欄位之描述。

二、資料品質檢測方式

7 項檢測指標列示如下：

檢測構面	檢測指標	計算基礎	判斷方式	判斷時點	輸出結果
資料可直接取得	連結有效性	全部資料資源	機器測試	定期	True/False
	資料資源能否直接下載	全部資料資源	機器測試	資料資源異動時	檔案載點(含 API 回傳資料)/ 網頁連結
資料易於被處理	是否屬結構化檔案類型	全部資料資源	機器測試 (非僅檢測副檔名)	資料資源異動時	結構化檔案 / 非結構化檔案
資料易於理解	詮釋資料編碼描述與資料相符	結構化資料資源	機器測試	資料資源異動時	True/False
	詮釋資料欄位描述與資料相符	結構化資料資源	機器測試	資料資源異動時	True/False
	資料更新時效性	結構化資料資源	機器測試	定期	無逾期/有逾期
民眾意見回饋	民間回饋意見之回復效率	全部資料集	人工檢核 (是否於 14 個工作天內進)	定期	

檢測構面	檢測指標	計算基礎	判斷方式	判斷時點	輸出結果
			行回復) / 民間評分低 於中間值者		

三、資料品質管理重點

- (一)各部會管理者參考結果報告，協助資料集提供者進行改善；倘資料產製來源為系統者，可配合於維運週期提供符合結構化資料定義之資料資源。
- (二)非結構化但可轉為結構化之資料資源（如：具有合併儲存格之 CSV、具報表原始資料然僅以 PDF 檔提供者），應以資料使用者的角度思考其提供方式是否妥適並安排期程改善。
- (三)資料內容為非結構化格式者，應避免進行無意義的檔案格式轉換（如：以 Word 另存為 XML 格式）。
- (四)政府資料開放平臺之資料集均應適用政府資料開放授權條款，各項資料之開放前均應確認其權利完整性（如：著作權等）。
- (五)民眾回饋意見，機關應於 14 個工作天內回復；如屬資料有誤，或重大事件相關之資料資源未能符合民間需求品質，除依上開期限回復外，數位發展部得協調資料提供機關改善。

參、資料品質提升作法說明

一、建議資料提供方式

(一)通則

1. 資料集詮釋資料注意事項

- (1)「資料集名稱」是資料資源集合的正式名稱，請以可一目瞭然之名稱命名之，必要時可加上機關全銜。
- (2)「資料集描述」為資料集簡明陳述，應避免與資料集名稱相同，讓使用者可進一步瞭解資料集內容。
- (3)如有說明文件（readme.*、schema.*）、說明網頁連結，應放置於「資料集相關網址(landingPage)」欄位，或與資料資源一併置於壓縮檔內。

2. 資料資源注意事項

- (1)如有檔案壓縮之需求，應於壓縮檔內提供 manifest 表單，以利使用者理解壓縮檔內各檔案之關聯性。

(2)檔案名稱建議以英數、連字號、底線、括號為限，避免使用中文檔名以免轉換時產生亂碼。

(3)若資料資源內容有所更新，或者有多個來源相同、欄位之資料資源，建議上架於同一資料集，不新建新資料集存放。

(二)API

1. 適用資料類型：高更新頻率資料或已有系統可即時產製資料者。

2. 建議資料格式：JSON、XML 格式。

3. 資料集詮釋資料注意事項：

建議優先提供符合 OpenAPI 3.0 含以上版本之說明文件。

(三)固定欄位結構化資料

1. 適用資料類型：定期產製且未經統計彙整之原始資料 (raw data)、統計資料、地理圖資、已有系統可產製資料者。

2. 建議資料格式：

CSV、JSON、XML、GeoJSON、KML、KMZ、SHP 等格式。

3. 資料內容注意事項：

(1)CSV 內容請以半形逗號","作為資料區隔，欄位標題以一列為原則 (可加註另一列為英文)。

(2)結構化資料內容無多表合併、無合併儲存格、無空行、無小計。

4. 資料集詮釋資料注意事項：

主要欄位說明請依「資料集詮釋資料標準規範」填列，欄位區隔請一致採用全形頓號"、"。

(四)非固定欄位結構化資料

1. 適用資料類型：會議紀錄、法規函釋等。

2. 建議資料格式：建議以 JSON 格式優先。

(五)非結構化資料

1. 適用資料類型：不限制。
2. 建議提供資料型態：不限制。

二、檢測流程概要

- (一)資料品質檢核功能於政府資料開放平臺後臺提供，後臺上架資料集（手動/批次）需於資料上架前進行資料資源檢測流程；透過介接方式上架之資料集，需於上架後於後臺或用品質檢測API執行檢測。單筆檢測結果可於介面瀏覽，批次檢測結果則以電子郵件通知測試結果，便利資料提供者據以提升資料品質。
- (二)當資料資源有所異動時，皆須重新檢測資料集，以確保更新資料之品質。
- (三)連結有效性測試以每週至少一次的頻率進行測試，放置資料資源之檔案伺服器應同時支援IPv4與IPv6。
- (四)資料資源是否屬結構化之判斷流程：

[start]承辦人增修資料資源

[A]判斷資料資源檔案格式

- 如果是 壓縮檔，到 [B]
- 如果是 JSON, XML, GeoJSON, KML, KMZ, SHP 檔，到 [C]
- 如果是 CSV 檔，到 [D]
- 如果是 XLS, XLSX, ODS 檔，到 [E]
- 如果是 API, Webservice 到[F]
- 如果是 其他檔，直接到 [Z]

[B] 壓縮檔 (zip, rar, 7z...) file

- 解壓縮後，先排除部分忽略清單(例如 readme.* schema.*等)
 - 如果只剩一個檔案，回到 [A]
 - 如果有多個檔案則到 [U]

[C] JSON, XML, GeoJSON, KML, KMZ, SHP file

→ 進行 syntax check

→ 失敗到 [Z]

→ 通過到 [T]

[D] CSV file

→ 進行 UTF-8 check

→ 屬 UTF-8，到 [T]

→ 非屬 UTF-8，轉存為 UTF-8 後到 [T]

[E] XLS, XLSX, ODS file

→ 有合併儲存格、繪製框線、儲存格背景色或使用公式等，到 [Z]

→ 有使用多分頁，到 [Z]

→ 無以上情形者，到 [T]

[F] API / Webservice

→ 若 API / Webservice 提供 html 介面，則到 [X]

→ 若直接提供 json 輸出結果，則回到 [C]

[T] 進行是否為 table 格式判定

→ 嘗試轉換為 CSV 格式

→ 屬 table 格式，取出 CSV 欄位值，填至資料集詮釋資料之 fieldDescription 欄位，於政府資料開放平臺保留轉換後之 CSV 檔，到 [Y]

→ 非屬 table 格式，到 [X]

[U] 判斷壓縮檔是否包含 manifest 描述檔

→ 無 manifest 描述檔，到 [Z]

→ 有 manifest 描述檔，採人工判斷 [end]

[X] 將這個檔案上架，標記為結構化資料 [end]。

[Y] 將原始檔案上架，標記為結構化資料，一併將轉存之 CSV 檔案上架，並註記轉換日期時間[end]。

[Z] 將檔案上架，標記為非結構化資料，並錄是從哪個流程進到這步，供部會管理者參考[end]。

肆、附錄

一、API 參考範例：

- (1) [交通部運輸資料流通服務平臺](#)
- (2) [政府資料開放跨平臺介接規範](#)

二、固定欄位結構化資料：

- (1) [消費力統計-各分位載具消費張數金額-資料集](#)
- (2) [社會經濟資料服務平臺](#)

三、非固定欄位結構化資料

- (1) [數位發展部資料開放諮詢小組會議紀錄](#)
- (2) [法務部全國法規資料庫](#)