

可信賴的資料儲存庫 的選用指引

下方表格提供了一份指引，用以列出挑選可信賴的資料儲存庫的準則，這些準則是由四個主要議題構成：

1 提供持續及唯一的識別碼 (Persistent and Unique Identifiers, PIDs)

可信賴的資料儲存庫應該：

1a 能進行資料探索和識別

- 確保 PIDs 包含在對應的後設資料裡。

1b 能搜尋、引用、和擷取資料

- 一致、有系統地編配 PIDs（如 DOI,⁵ URN,⁶ ARK⁷）予持有的資料，使得即使資料的存放位置有變，其對應的資料和後設資料仍可被找到、參照、及擷取。

1c 可支援多個資料版本

- 確保儲存庫中儲存的資料版本，都有被明確標示並以永久的稽核軌跡 (Audit Trail) 進行紀錄，以追蹤其來源出處。

注意：不是所有的儲存庫都使用上述被認可及普遍的 PID 系統。相反地，他們會使用儲存庫自己維護的編碼系統或管理編號。在這情況下，一旦資料存放的位置改變，或儲存庫不復存在、重組、或變更治理方式，資料無法尋獲的風險將會增加。

2 後設資料

應使用豐富的後設資料來描述資料。後設資料應記錄資料的產生方式、所用的授權條款、以及可再次利用的方式，後設資料並應提供脈絡，讓其他研究者能適當地解釋資料。

一個可信賴的資料儲存庫應該：

2a 讓資料可供查找

- 以語法結構完善及容易理解的語言，提供資料及後設資料，以確保資料可相互操作及可再次使用。資料和後設資料應使用標準的詞彙和格式加以描述，以使電腦系統可以搜尋、可以自動合併、以及可自研究資料檔案區分出後設資料。

2b 可以參照至相關聯的資訊

- 確保後設資料的資訊，可連結至其他相關資訊，這些連結方式包含提供PID，以及描述相關資訊和資料的科學關聯。其中一類的資訊，即是相關研究人員的細部資訊，研究人員的永久識別碼即是為此而生（例如：ORCID⁸、ISNI⁹、或DAI¹⁰）。

2c 應為資料提供公開有效、可維護的資訊，即使該資料尚未出版、受保護、被撤回、或已刪除

- 即使研究資料內容（因隱私限制、法律義務、或其他保護措施等原因）暫時或已不能取用，儲存庫亦要確保相關的後設資料有被長期典藏，並始終保持可擷取的狀態。
- 因不良的研究成果、不當的研究行為等緣由而撤回的資料，應確保該資料仍能透過後設資料被找到，且仍能被保存，以便研究紀錄的查考。

2d 使用被（科學社群）廣泛接受的後設資料格式

- 應確保儲存庫維護的後設資料被（科學社群）廣泛認可，並可用機器擷取。

- 確保資料處理方式有遵循社群標準或最佳實務（若有的話）。專門為特定研究領域服務的儲存庫在上傳資料或後設資料時，請留意其或許設有社群標準。

2e 確保後設資料能夠以機器擷取

鼓勵後設資料所含的資訊是依機器可擷取的方式進行結構編排，例如，儲存庫提供固定欄位的表單以供使用者填寫。

3 資料取用及使用授權

一個可信賴的資料儲存庫應該：

3a 資料的取用是在明確的條件下進行

- 清楚闡明資料再次使用的條款。此類（授權）資訊通常包含在後設資料中。

3b 確保資料的真實性及完整性

- 確保後設資料包含關於資料來源出處的詳細資訊，包括資料的產生及處理的方式，資料在何種情況下可被再次使用、以及資料的可靠程度。

3c 讓資料可被擷取

- 允許使用開放、標準化協定（而非私有的通訊協定）以擷取資料，或最少允許擷取後設資料。

3d 提供關於授權及許可的資訊（使用機器可讀格式最為理想）

- 授權資訊應以結構化的方式被參照，最好的狀況是對人及機器而言，使用條件都很清楚。盡量採用可透過 URL（網址）參照至授權內文的通用或廣泛接受之授權系統（例如創用 CC 授權）。

3e 確保資料當事人及創作者的保密性與權利

- 提供對人類及機器使用者驗證身份和授權的方法，並允許針對保密事項或限制事項，設置使用者（或群組）特定的取用權限。

4 資料保存

一個可信賴的資料儲存庫應該：

4a 確保後設資料及資料的持續保存

- 確保使用者所委託的資料和後設資料，持續有效並可被取用。

4b 資訊透明：列明儲存庫的任務、範圍、資料保存政策和方案（包括治理、財務永續性、資料保存期限、以及營運維持方案）

- 以書面形式紀錄資料和後設資料的管理與保存，尤其應該有保存策略，詳細說明資料儲存庫的任務和範圍、治理面向、財務永續性、外包合作夥伴和保存期間（保存的時間範圍）。
- 制定可公開取得的應變計劃，以確保資料儲存庫的生命週期完結時，其所保有的資料和後設資料也能獲得保存（例如，通過便利的方式提取資料和後設資料，並將其轉移到別的儲存庫）。