



資料管理方案介紹

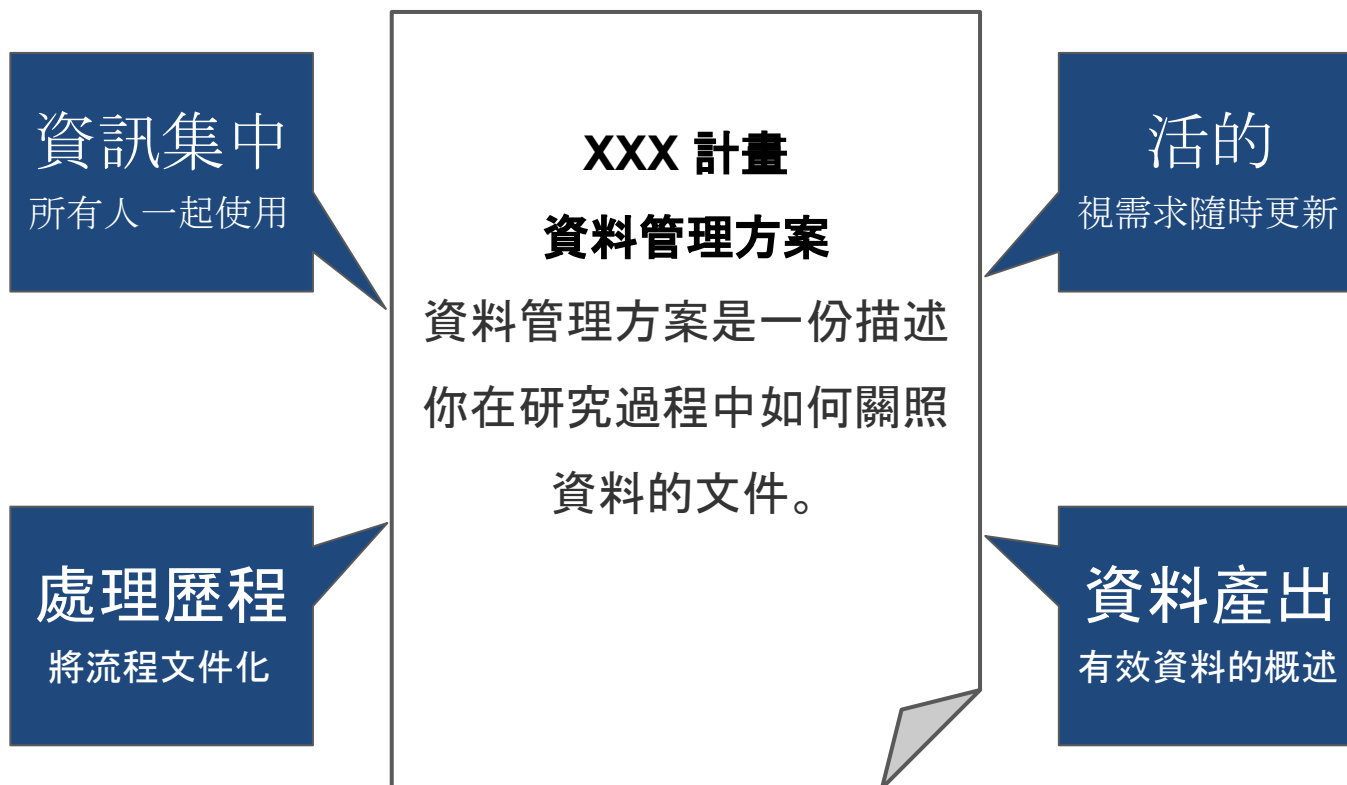
Introduction to Data Management Plan (DMP)

何明誼 Ming-Syuan Ho

hmsyuan@citi.sinica.edu.tw

2023-1-12 @東華大學孫義方老師團隊交流

什麼是「資料管理方案」(DMP)

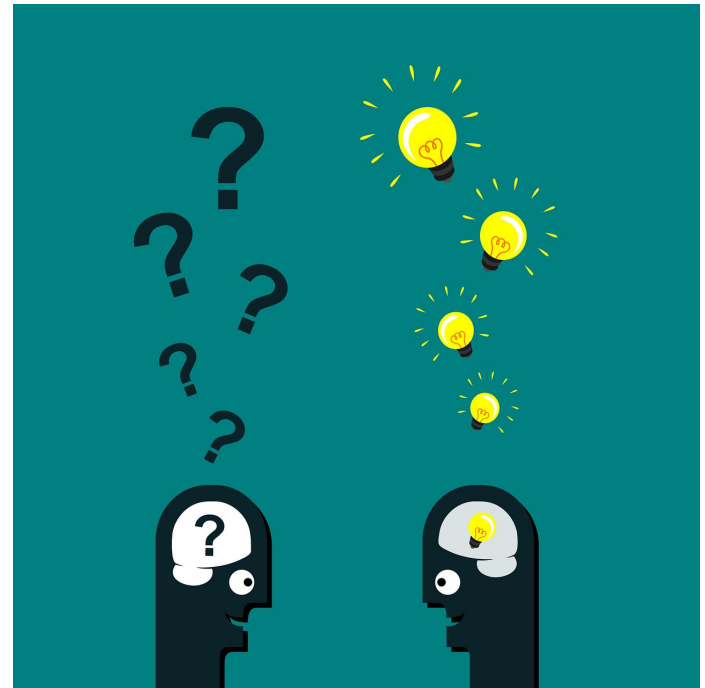


DMP 是分享用的文件

至少你要和未來的自己分享

為什麼寫 DMP？

- 受他方要求 (e.g., 經費資助者 (funder) 的要求)
- 把心目中的資料處理歷程正式明文且記錄下來
- 協助在研究初期辨識資料管理不足
- 節省後續研究過程的時間
- 促進科學研究過程的透明化





DMP & 經費資助者 (Funder) 要求

- INT - Belmont Forum
- EU - Horizon Europe
- US - National Science Foundation (NSF)
- TW - 國科會永續學門
- Others

NSF's DMP Requirement

Dissemination and Sharing of Research Results - NSF Data Management Plan Requirements

NSF DATA SHARING POLICY

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See [Proposal & Award Policies & Procedures Guide \(PAPPG\) Chapter XI.D.4](#).


NSF DATA MANAGEMENT PLAN REQUIREMENTS

Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See [PAPPG Chapter II.C.2.j](#) for full policy implementation.



NSF's DMP Content

- NSF 建議計畫申請者的 DMP 可包含以下內容：
 1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
 2. the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
 3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
 4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
 5. plans for archiving data, samples, and other research products, and for preservation of access to them.



撰寫與執行計畫申請書 DMP 的 Tips

1. **確認要求:** 確認經費資助者是否要求提交 DMP
2. **確認格式:** 確認經費資助者是否有說明 DMP 的文件格式、篇幅及所須包含的內容
3. **確認資料管理的目標:** 確認進行研究資料管理時應符合的資料或研究成果政策
4. **自我評估與撰寫:** (可參考已公開的 DMP) 花數小時至數天撰寫 DMP, 並確認文件所載時程可同時符合資助者和自身研究所需
5. **實踐與修正:** 定期修正 DMP 的內容

DMP 的內容

*"The content of this guide has been developed to support researchers in ensuring that data are **FAIR**, where appropriate."* (p.4)

DMP 的六大核心要求：

1. 資料的描述與蒐集, 或既有資料的再次使用
2. 文件與資料的品質
3. 研究過程中的資料儲存及備份
4. 法律和倫理要求、行為守則
5. 資料共享和長期保存
6. 資料管理的責任與資源



https://data.depositar.io/dataset/se_rdm_guides

1. 資料的描述與蒐集，或既有資料的再次使用

a. 如何蒐集或產生新的資料？如何再次使用既有資料？

- 請說明收集或產生新資料的方法或軟體。
- 陳述既有資料在再次使用上的任何限制或使用條款(如果有的話)。
- 闡述如何記錄資料溯源(data provenance)。
- 若有考慮過重用任何既有資料來源但最終決定不採用，請簡述原因。

b. 要蒐集或產生什麼(種類、格式和數量的)資料？

- 提供資料類型的詳細資訊：數值、文本、圖像、音訊、影片、混合媒體等。
- 提供資料格式的詳細資訊：編碼方式、副檔名等。
- 決定使用某資料格式的理由為何？
- 請優先選用開放、標準的檔案格式。
- 提供資料大小及數量的詳細資訊。


Solution 格式資料類型與偏好格式

Data Type	Recommended formats	Acceptable formats
-----------	---------------------	--------------------

資料類型*§:

-  壓縮檔資料
-  原始碼
-  組態資料
-  資料庫
-  辦公軟體文件
-  結構化圖形
-  影像
-  影音資料
-  網路通訊資料
-  原始輸出資料
-  科學與統計資料
-  應用程式
-  結構化文字資料
-  純文字資料
-  其他

資料集管理頁面

	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • CSV (.csv) 	<ul style="list-style-type: none"> • Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDFS (.hdfs, .h5)
---	--	--

本平台預設會依據資源後設資料的「格式」項目設定資料預覽（參閱 [新增資料集](#) 步驟 5 內容），下列為本平台支援之預覽格式：

- 文字類：txt, html, xml, json, geojson。
- 影像類：png, jpg, jpeg, gif。
- 影片類：MP4, WebM, Ogg。
- 音訊類：MP3, WAV, Ogg。
- 表格類：csv, xls(x)。
- 空間資料：WMTS, WMS, Shapefile（Shapefile 請標記為「shp」，否則將無法預覽）。
- 其他：PDF, 一般網頁連結。

 Video	<ul style="list-style-type: none"> • Matroska (.mkv) 	<ul style="list-style-type: none"> • MPEG-2 (.mpg, .mpeg, .m2v, mpg2) • AVI (.avi) • QuickTime (.mov, .qt)
--	---	---

2. 文件與資料的品質

a. 有哪些後設資料和文件 (如資料蒐集方法和組織資料方式的描述) 會伴同資料？

- 將提供哪些後設資料來協助他人識別和尋得資料。
- 將使用哪些後設資料標準 (例如 DDI, TEI)
- 說明如何組織資料以及相關之專業標準、版本控制、資料夾結構。
- 需附哪些其他文件記錄才能使資料可被再用？
- 如何擷取以上提到的資訊以及將其記錄在何處？

b. 將採取哪些措施以維持資料品質？

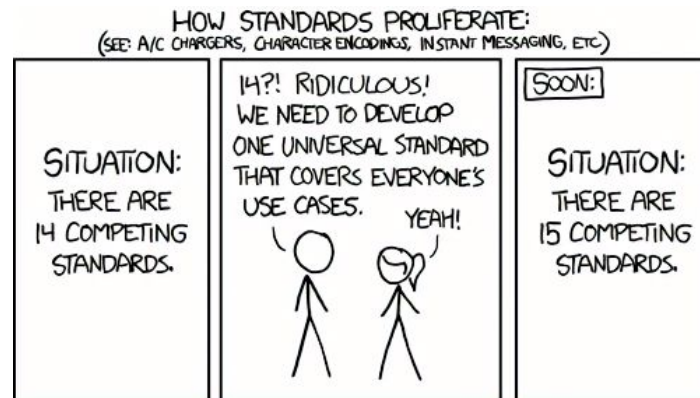
- 闡述收集回來的資料是一致、優良，及如何記錄這一切。這過程可能包括：校準，重複收集樣本或測量，標準化的資料擷取，資料驗證與勘誤，資料的同儕評審，使用的控制詞彙表述等。

如何找到適合你的 metadata 標準？

- 各學科領域所慣用的標準不同，亦有很多學科目前並無相關標準

- 慣用標準可查詢：

- [RDA Metadata Directory](#)
- [DCC discipline Metadata list](#)
- [Bath Metadata Standard Catalogue](#)



- 確認機構或偏好的資料儲存庫 (data repository) 在 data 或 dataset level 上的慣用標準

- e.g.1, GBIF: Darwin Core @data level | EML @dataset level
- e.g.2, Zenodo - DataCite @dataset level
- e.g.3, CKAN/depositar - DCAT @dataset level

Solution: 後設資料標準(資料集層級)

資料集層級

欄位	對應語彙	範圍	規範
標題	dct:title	rdfs:Literal	DCAT 2
網址	dct:identifier	rdfs:Literal	DCAT 2
摘要	dct:description	rdfs:Literal	DCAT 2
資料類型	dct:type	rdfs:Class	DCAT 2
Wikidata 關鍵字	dcat:theme	skos:Concept	DCAT 2
標籤	dcat:keyword	rdfs:Literal	DCAT 2
語言	dct:language	dct:LinguisticSystem	DCAT 2
備註	(無對應)	(不適用)	(不適用)
時間解析度	dcat:temporalResolution	xsd:duration	DCAT 2
(時間資訊類型)	dct:temporal	dct:PeriodOfTime	DCAT 2

3. 研究過程中的資料儲存及備份

a. 研究過程中，如何儲存資料和後設資料，並建立備份？

- 哪裡儲存和備份資料，以及每隔多久做備份。
- 優先選用具自動備份功能、運作穩固而且有管理人員掌管的託管存儲空間，例如由研究人員原屬機構/單位的資訊支援服務提供的備份空間。

b. 研究執行期間，如何確保資料安全並保護敏感資料？

- 說明一旦發生意外時該如何復原資料。
- 如何管理使用者角色權限設置。
- 保護資料免受安全威脅。評估及描述主要風險以及如何管控。
- 說明機構內有哪些現行的資料保護政策。

Solution: 版本備份與比較



資料集



主題



動態牆



展示案例



管理



  何明誼 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
2 週前 | [查看此版本](#) | [變更](#) **點選「查看此版本」，可查看資料集過去的版本**

  Cheng-Jen Lee 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
3 週前 | [查看此版本](#) | [變更](#)

  Cheng-Jen Lee 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
3 週前 | [查看此版本](#) | [變更](#)

  Tyng-Ruey Chuang 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
3 週前 | [查看此版本](#) | [變更](#)

  Tyng-Ruey Chuang 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
3 週前 | [查看此版本](#) | [變更](#)

  Tyng-Ruey Chuang 更新了「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集
3 週前 | [查看此版本](#) | [變更](#)

You're currently viewing an old version of this dataset. To see the current version, click [here](#).

「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集

什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？

課程描述：

提出研究計畫時，被要求同時提交「資料管理方案」(Data Management Plan)，詳細說明在計畫執行期間，研究資料在蒐集、紀錄、管理、保存的實施方法，並需規劃研究資料的釋出使用，這已是研究團隊經常面臨的議題（如國科會永續學門目前已有要求）。提出論文時，亦需說明研究資料可被取用的情形，常見的方式是將資料寄存於公眾可取用的「研究資料儲存庫」(Research Data Repository)。而如何選用合適的研究資料儲存庫？他們又應該具備哪些理想特點呢？本短期課程將簡要介紹研究資料寄存以及資料管理方案，以實務角度分享我們的經驗。課程中將以中央研究院資訊所和資創中心所支持發展的「研究資料寄存所」(<https://data.depositar.io>) 為例，解說研究資料保存與釋出的作法，以及其與研究資料管理的關聯。相關文件與討論請參見「研究資料管理入門」(<https://data.depositar.io/organization/rdm-basics>) 專案資料集以及「研究資料管理推進室」

編輯

資料集

成員

瀏覽

新增資料集

10 datasets

搜尋資料集



找到 10 個資料集

排序依照:

根據名稱遞增排序

<input type="checkbox"/>	<input type="checkbox"/> 設為公開	<input type="checkbox"/> 設為非公開	<input type="checkbox"/> 刪除
<input type="checkbox"/>	「什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？」課程資料集		編輯
	什麼是研究資料儲存庫？什麼是資料管理方案？為何您需要關心？課程描述：提出研究計畫時，被要求同時提交「資料管理方案」(Data Management...		
<input type="checkbox"/>	COSCUP 2022 — 資料長期保存的好所在：用 CKAN 打造的「研究資料寄存所」(depositor)		

4. 法律和倫理要求、行為守則

a. 如何確保涉及個人資料的處理過程能符合相關法規並保障資料安全？

- 確保在處理個人資料時遵守資料保護法規
 - 獲得個人資料的知情同意。
 - 建議個人資料匿名化 (anonymisation) 以進行保存和共享。
 - 考慮將個人資料擬匿名化 (pseudonymisation)。
 - 考慮將個人資料加密。
 - 說明是否有為具存取個人資料權限的使用者訂立管理資料取用的程序。

b. 如何處理智慧財產權與所有權等各類可能衍生的法律問題？有哪些適用的法規？

- 說明誰是資料的擁有者。
- 是否會影響到智慧財產權。若有，如何處理？
- 闡明對第三方資料的再使用是否有任何限制。

c. 如何考慮可能的倫理問題，並遵守行為準則？

- 闡明倫理問題是否影響到資料的儲存、傳輸、探索、使用及保存期限，以及相對應的方案。
- 遵守國內和國際對資料處理及保護的行為守則和機構倫理準則，並檢查蒐集研究資料的過程是否需要執行倫理審查。

Solution: 個資與授權

depositar 使用條款 §2:

您所上傳的資料(無論設定為公開與否), 您已確認有權利上傳並儲存該資料在本服務。您所上傳的資料均適合廣為散布, 並合乎個人資料保護、資通安全、及智慧財產權等法規要求。**您對所上傳的資料負完全責任。**

depositar 使用條款 §5:

本服務得就下列情形為逕行處理, 且不對該處理之結果負責:

- a. 刪除或隱藏不適當或未受充分保護之資料;
- b. 撤銷或限制妨礙本服務運作、違反本使用條款、或違反法律之使用者帳號

5. 資料共享和長期保存

a. 資料該在何時、以何種方式共享？在資料共享上是否有特定限制？若延後開放的期限，其理由為何？

- 說明如何讓資料更容易被尋獲及以何種方式共享。
- 規劃長期保存資料的方案。
- 說明資料可供開放使用的預期發佈時間。是否限定某些資料只供某使用者取用，如果需要，請闡述原因以及時限。說明是否需要延遲發佈或限制資料共享，以保護智慧財產權或申請專利。
- 指明誰有使用資料的權限。若有必要訂立資料共享協議，請說明方式和原因。說明將採取什麼措施來克服限制或將限制減至最小。

b. 如何甄選需要長期保存的資料？在何處（如資料儲存庫或典藏庫）長期存放？

- 列明出於契約、法律或法規目的而必須保留或銷毀哪些資料。
- 列明如何決定要保留哪些資料及需要長期保存的資料。
- 評估資料的潛在研究用途和使用者。
- 說明資料存放的位置。若不打算使用已建立的資料儲存庫，則請於DMP中闡明，在計畫經費到期後，資料仍能被有效地策展。建議需闡明資料儲存庫的政策和程序（包括任何後設資料標準和涉及的成本）已受檢核。

5. 資料共享和長期保存_2

c. 需要透過哪種方法、軟體或工具才能取用資料嗎？

- 說明潛在使用者是否需要特定工具來取用、使用或再次使用資料？請考量用以取用資料的軟體之永續性。
- 說明資料是否能透過資料儲存庫、直接遞交請求、或使用其他機制等方式達成共享？

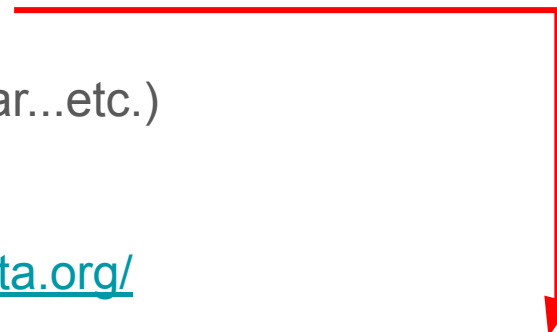
d. 如何確保編配具唯一性及永久性的識別碼（如數位物件識別碼 DOI）給每一個資料集？

- 解釋可再次使用資料的其他情況。資料應該編配持續的識別碼 (PIDs)，這可讓查找和引用資料更可靠及有效率。PIDs 還有助於追蹤資料被引用和再次使用的記錄。
- 闡明是否會替資料編配 PIDs。通常信賴度高、可長期提供服務的資料儲存庫，會提供持續的識別碼。

你需要什麼資料儲存庫？

- 選用原則: Science Europe〈可信賴的資料儲存庫的選用準則〉
 - 領域的習慣
 - 期刊、機構、補助機構推薦或指定
 - 通用型 (Zenodo, 4TU, DANS, depositar...etc.)
- 名錄:
 - re3data (Datacite) : <https://www.re3data.org/>
 - FAIRsharing: <https://fairsharing.org/>

以 Nature 為例

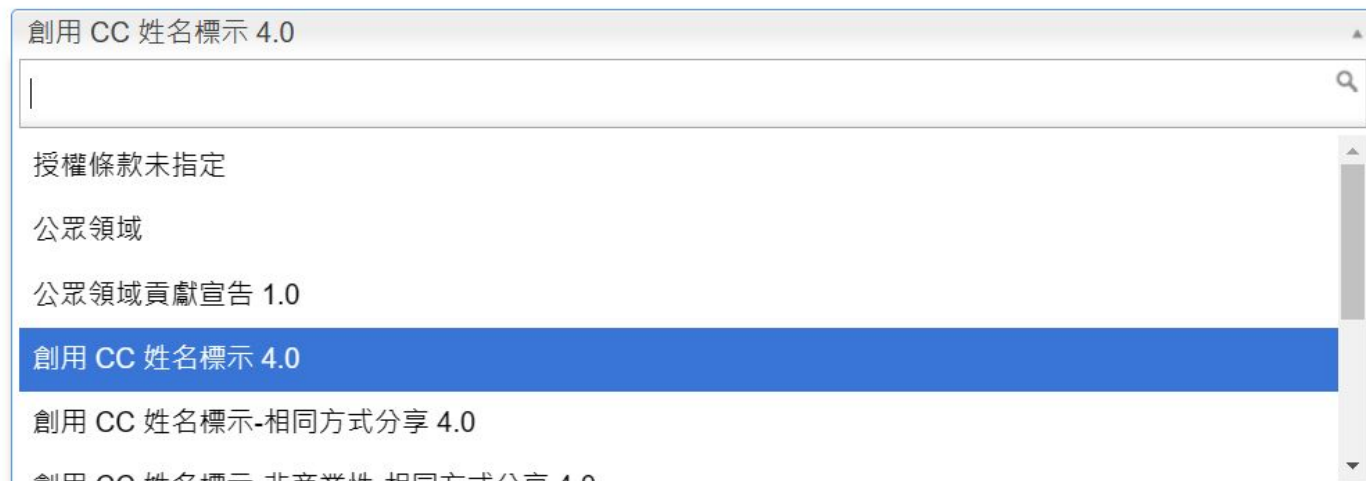


View data repositories

- **Biological sciences:** Nucleic acid sequence; Protein sequence; Molecular & supramolecular structure; Neuroscience; Omics; Taxonomy & species diversity; Mathematical & modelling resources; Cytometry and Immunology; Imaging; Organism-focused resources
- **Health sciences**
- **Chemistry and Chemical biology**
- **Earth, Environmental and Space sciences:** Broad scope Earth & environmental sciences; Astronomy & planetary sciences; Biogeochemistry and Geochemistry; Climate sciences; Ecology; Geomagnetism & Palaeomagnetism; Ocean sciences; Solid Earth sciences
- **Physics**
- **Materials science**
- **Social sciences**
- **Generalist repositories**

Solution: 使用資料的權限

● 授權*:



資料集的管理頁面



資料集的瀏覽頁面

Solution: 如何讓資料更容易被尋獲及共享？

開放且自由使用

以開放原始碼軟體 CKAN 為基底建構的資料儲存服務，已客製化並擴增功能，並支援研究資料管理。所有研究者都可註冊使用，用來管理和開放研究資料集。

[更多資訊](#)

資料集易取易得

註冊沒有資格限制、寄存不限資料類型，使用不須安裝軟體。透過時空範圍、資料類型、關鍵字等條件，可快速找到資料集。資料集並由 [Google 資料集搜尋](#) 所收錄索引。

[更多資訊](#)

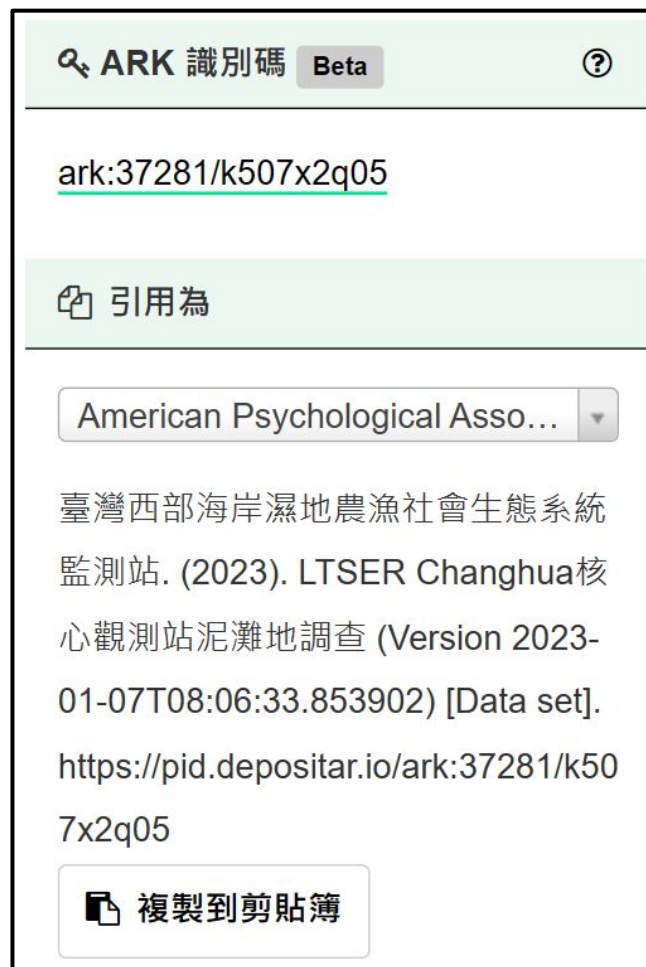
具程式存取介面

可透過程式存取介面 (API)，自動取用並匯入資料集。API 亦可用於存取結構化的資源內容，如 CSV 與 Excel 表格，方便開發者進行資料取用與分析。

[更多資訊](#)

Solution: 持續識別碼 (Persistent Identifier, PID)

- 在數位場域, PID 通常伴隨特定連結, 用以長期指向某個**公開**資源。即使該資源消失, 連結內容也會保留基本的後設資料。(就像書的ISBN)
- 常見的PID: DOI、ARK、Handel等
 - e.g. doi:10.1000/182
 - e.g. ark:70795/k507x2q05
- 李承鑫 (2022)。認識網路資料的持續識別碼。RDM Hub。



The screenshot shows the ARK Beta interface. At the top, there is a search bar with the text "ARK 識別碼" and a "Beta" label. Below the search bar, the identifier "ark:37281/k507x2q05" is displayed. Underneath, there is a section titled "引用為" (Cited as) with a dropdown menu showing "American Psychological Asso...". Below the dropdown, the citation text is displayed: "臺灣西部海岸濕地農漁社會生態系統監測站. (2023). LTSER Changhua核心觀測站泥灘地調查 (Version 2023-01-07T08:06:33.853902) [Data set]. https://pid.depositar.io/ark:37281/k507x2q05". At the bottom, there is a button labeled "複製到剪貼簿" (Copy to clipboard).

6. 資料管理的責任和資源

a. 由何人(如角色、職位及機關)負責資料管理(如資料託管員)？

- 概述資料管理 (management) 和託管 (stewardship) 的角色和職責, 例如: 何方負責資料擷取、產出後設資料、把關資料品質、維護儲存空間和備份、處理資料歸檔以及資料共享。盡量列明負責人的名稱。
- 至於多方協作的計畫, 請說明合作夥伴之間如何分配及協調資料管理的職責。
- 列明誰負責實施 DMP、確保方案有被評核並在必要時做修訂。
- 考慮定期更新 DMP。

b. 哪些資源(如經費及時間)將專用於管理資料及確保資料的 FAIR 特質(即可被找到、可被取用、可相互操作、可再次使用)？

- 估算將資料準備為可共享、保存(即資料策展)的必要資源(如時間), 並解釋其估算方式。仔細考量並說明交付資料所需的資源及其理由。這包括儲存成本、硬體設備、人力時間、準備資料寄存的成本、與資料儲存庫的費用。
- 闡明為了準備資料寄存、或為了支應資料儲存庫要求的費用時, 是否需要額外的資源? 若需要, 請說明所需資源的多寡, 以及如何滿足這些開支。

Solution: 哪些資源將被用於確保資料的FAIR特質

開放且自由使用

以開放原始碼軟體 CKAN 為基底建構的資料儲存服務，已客製化並擴增功能，並支援研究資料管理。所有研究者都可註冊使用，用來管理和開放研究資料集。

更多資訊

資料集易取易得

註冊沒有資格限制、寄存不限資料類型，使用不須安裝軟體。透過時空範圍、資料類型、關鍵字等條件，可快速找到資料集。資料集並由 [Google 資料集搜尋](#) 所收錄索引。

更多資訊

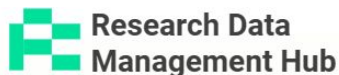
具程式存取介面

可透過程式存取介面 (API)，自動取用並匯入資料集。API 亦可用於存取結構化的資源內容，如 CSV 與 Excel 表格，方便開發者進行資料取用與分析。

更多資訊

討論 DMP (或RDM) 的疑難雜症: RDM Hub Mailing List

<https://rdm.depositar.io>



資源 實踐 消息 關於

經驗分享



當資料管理作為 NGO 提倡企業綠色轉型的推手——專訪綠色公民行動聯盟

本文介紹綠盟研究團隊如何善用政府開放資料，建置「透明足跡」環境資料庫，並在透明足跡資料庫的基礎上，推出「掃了再買」行動應用程式、「ESG 檢測儀」企業環境永續資料平台，透過向大眾揭露企業環境表現的方式，...

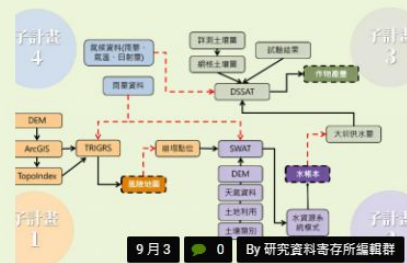
[READ MORE >](#)



研究資料管理經驗分享：個人保養產品之抗紫外線成份環境荷爾蒙暴露、健康風險與管理策略研究

作為首次執行 DMP 計畫的團隊，為有效建立資料蒐集、使用、管理、保存、分享的 SOP，並在研究過程中可隨時修正 DMP，橫向整合與溝通協調亦花費不少額外時間。此前期建置的磨合期長，且增加研究者的文書行政工作...

[READ MORE >](#)



資料管理計畫與科學研究傳承——臺灣氣候智慧調適與跨領域氣候風險評估之研究

由於資料產製過程步驟繁瑣且耗時，加上如雨型設計及不同重現期下延時 24 小時的設計時雨量等資料，屬分析常用的基本氣象資料，未來研究極有可能會使用到相同的資料，且未來若有計畫針對相關主題進行更深入的研究，...

[READ MORE >](#)

[VIEW ALL](#)

訂閱線上論壇

您可在本論壇提出任何應用情境進行討論與知識分享。(更多資訊)

訂閱

Thank You!

<https://data.depositar.io>

<https://rdm.depositar.io>

data.contact@depositar.io

rdm.contact@depositar.io

