

研究資料管理概述

2023-12-21

2023 研究資料管理工作坊

何明諠 | 中央研究院資訊科技創新研究中心 | 專案經理
hmsyuan@citi.sinica.edu.tw



depositar
研究資料寄存所

「研究資料寄存所」是中央研究院資訊所、資創中心的協作專案，部份經費來自國家科學及技術委員會。<https://data.depositar.io/about>



提到研究資料管理，你會想到什麼？



Menti code: 22306763

<https://www.menti.com/>

什麼是「研究資料」的「管理」？

- **研究資料 (Research Data)** 是任何構成研究分析與結果的資料。
- **研究資料管理 (Research Data Management, RDM)** 是一個通用的研究技能，使研究資料能被負責、開放且永續的使用。

General	Social Sciences	Natural/Physical Sciences
<ul style="list-style-type: none">• images• video• mapping/GIS data• numerical measurements	<ul style="list-style-type: none">• survey responses• focus group and individual interview transcripts• economic indicators• demographics• opinion polling	<ul style="list-style-type: none">• measurements generated by sensors/laboratory instruments• computer modeling• simulations• observations and/or field studies• specimen

不做好研究資料管理 (RDM), 是有代價的

研究過程中：

- 資料損毀或遺失
- 延宕研究進度

研究成果發表後：

- 無法再現研究結果
- 他人難以取得資料



nature > news > article

NEWS | 21 June 2022

Many researchers say they'll share data – but don't

Reasons included a lack of informed consent or ethics approval and that others had moved on from the project.

Clare Watson



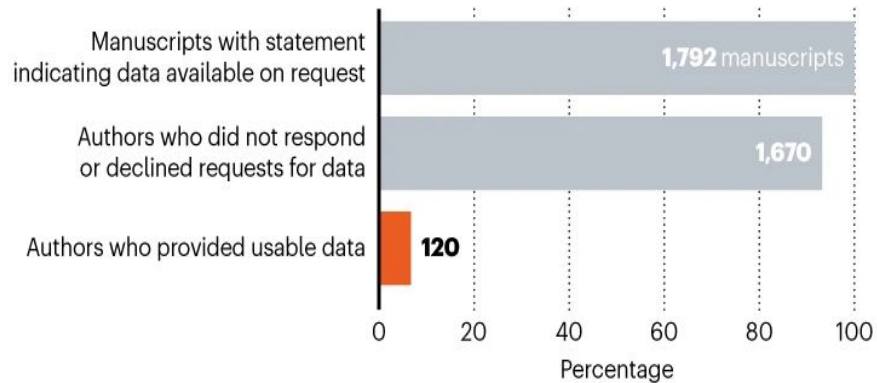
You have full access to this article via Academia Sinica

Download PDF



DATA-SHARING BEHAVIOUR

Of almost 1,800 manuscripts for which the authors stated they were willing to share their data, more than 90% of corresponding authors either declined or did not respond to requests for data. Only about 7% of authors actually handed over data.



不做好 RDM, 是有代價的 (2)

- 2018 年的報告:RDM 的實踐零星且非最佳化。
- 沒有開放/FAIR 資料會產生巨大的成本
- 至少每年額外支出 100 億歐元 (最多可能 270 億歐元)
- 主要耗費在「花費的時間」及「儲存成本」

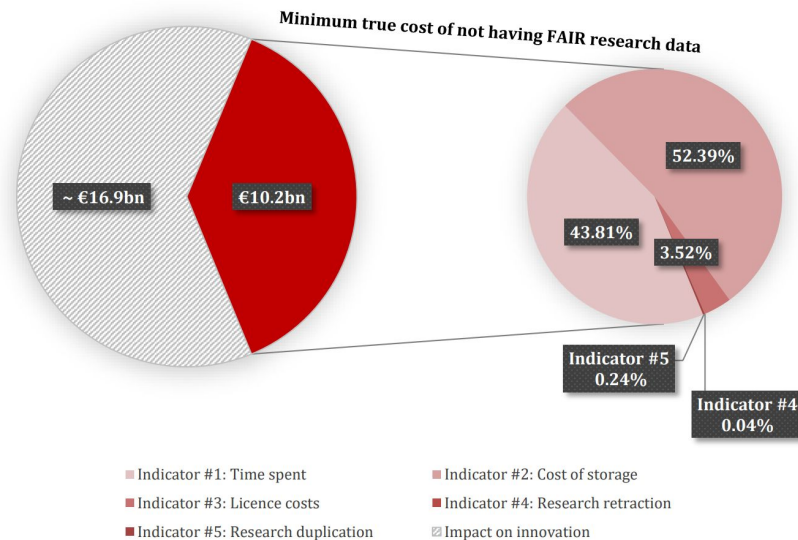
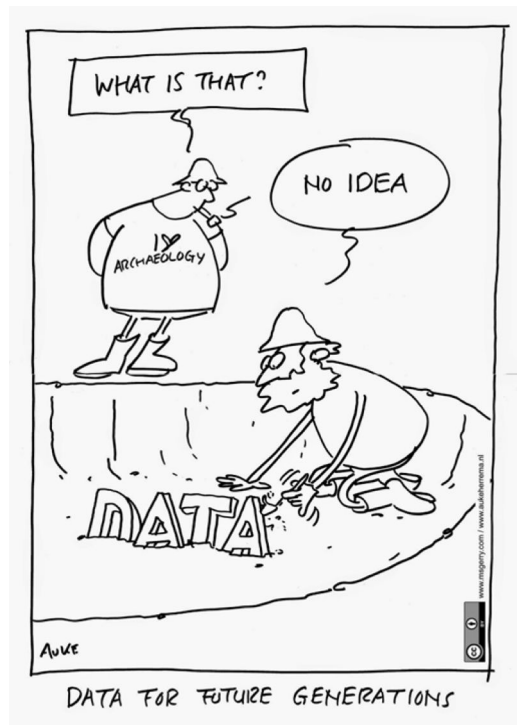


Figure 5: Cost breakdown

為什麼要做 RDM

- 符合科研經費補助單位或期刊的要求
- 讓研究過程更順利, 提升研究效率
- 使研究可再現, 提升研究誠信
- 使資料可被散布, 讓更多人使用
- 加速科學進展
- 帶來更多合作機會
- 讓論文獲得更多引用
 - studies that made data available in a public repository received 9% (95% confidence interval: 5% to 13%) more citations than similar studies for which the data was not made available. ([Piowar HA, Vision TJ. 2013](#))



RDM 的執行：政策面

- (2023/01) National Institute of Health (NIH) - [Policy for Data Management and Sharing](#)
- (2022/10) CERN - [CERN Open Science Policy](#)
- (2022/08) Office of Science and Technology Policy (OSTP) Memo - [Ensuring Free, Immediate, and Equitable Access to Federally Funded Research](#)
- (2021/04) National Science Foundation - [Proposal and Award Policies and Procedures Guide \(PAPPG\)](#)
- (2022/04) Horizon Europe - [Euratom Research and Training Programme General Mode Grant Agreement & EIC Accelerator Contractor](#)
- (2018/11) USGS-CASC - [Data Sharing Policy](#)
- (2015) Belmont Forum - [Open Data Policy and Principles](#)
- (2014) The UK Economic and Social Research Council (ESRC) - [ESRC Research Data Policy](#)

RDM 的目標：讓資料符合「FAIR 原則」

Findable
可被找到

人或機器能輕鬆找到資料或後設資料。

Accessible
可被取用

人或機器可存取或下載資料或後設資料。

Interoperable
可相互操作

不同來源的資料可被交換或結合。

Reusable
可再次使用


資料能容易被未來的研究複製或結合。

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, ... Barend Mons  + Show authors

Scientific Data **3**, Article number: 160018 (2016) | [Cite this article](#)

463k Accesses | 4223 Citations | 1991 Altmetric | [Metrics](#)

 An [Addendum](#) to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

完整的 FAIR 原則

Findable
可被找到

人或機器能輕鬆找到資料或後設資料。

- F1. 配發持續識別碼 (PIDs)
- F2. 使用豐富的后設資料 (metadata)
- F3. 資料在可搜尋的資源中註冊或索引
- F4. 後設資料指明資料的識別碼

Accessible
可被取用

人或機器可存取或下載資料或後設資料。

- A1. 資料的識別碼使用標準化的通訊協定
- A2. 即使資料不可取得，後設資料仍可得

Interoperable
可相互操作

不同來源的資料可被交換或結合。

- I1. 使用正規、共享及廣泛使用的語言以達成知識再現 (e.g. RDF, JSON-LD)
- I2. 使用符合 FAIR 原則的語彙 (vocabulary)
- I3. 資料包括對其它資料的合適參照

Reusable
可再次使用

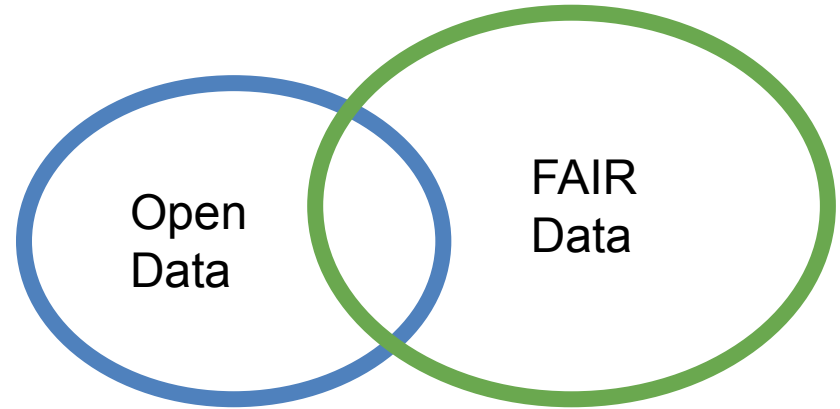
資料能容易被未來的研究複製或結合。

- R1. 以精確且相關的多種屬性，豐富地描述(後設)資料
 - R1.1 以清晰、可取用的資料授權條款釋出(後設)資料
 - R1.2 (後設)資料連結到詳細的溯源資訊
 - R1.3. (後設)資料符合該領域相關的社群標準

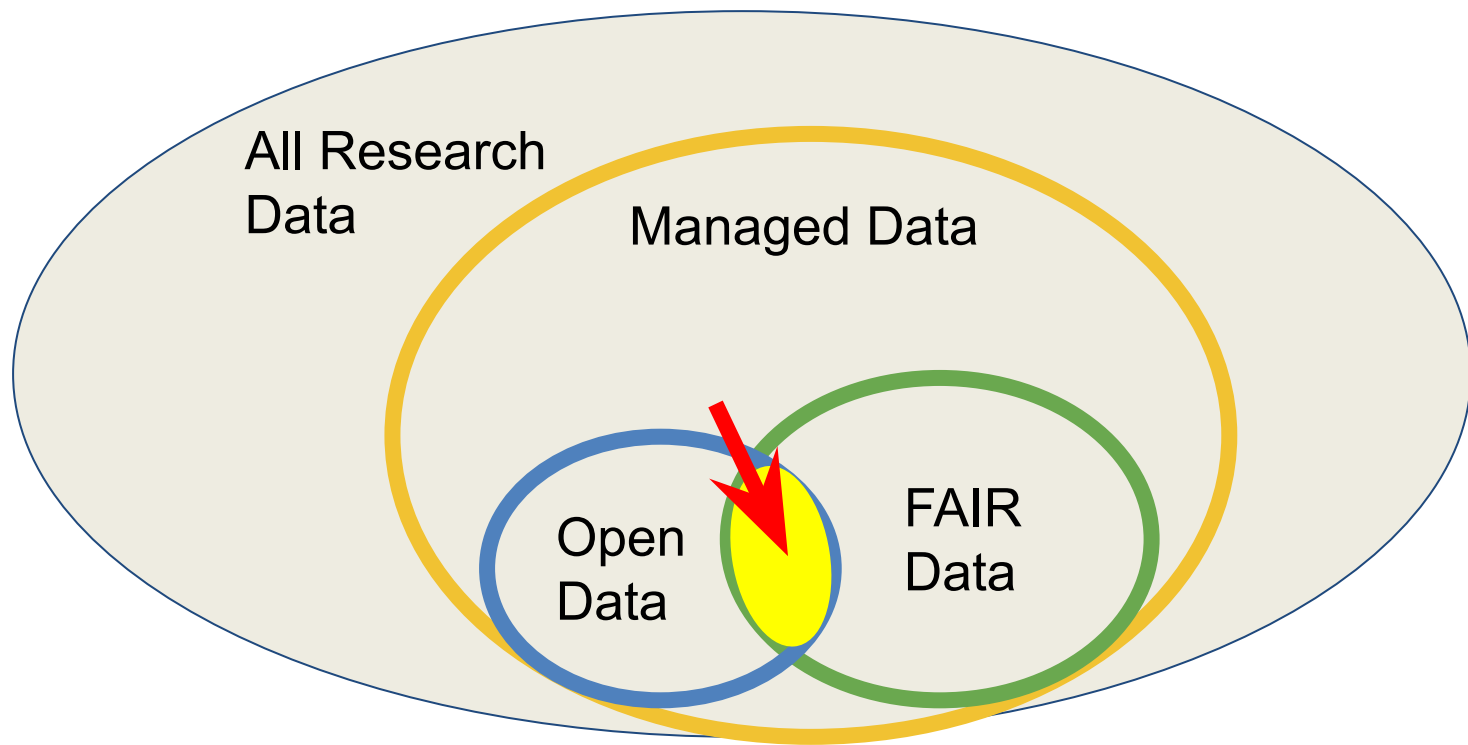
FAIR Data

≠

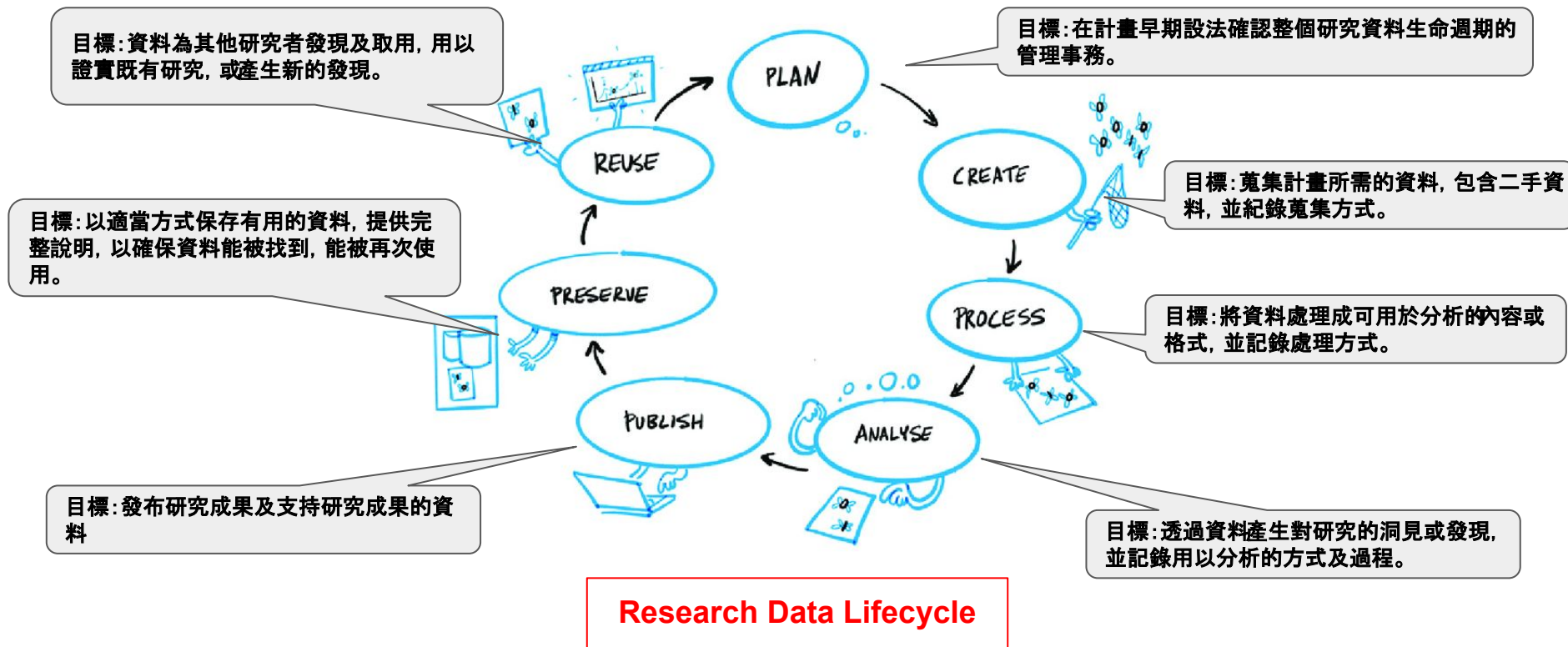
Open Data



RDM 最好的狀況：增加既 Open 又 FAIR 的研究資料



具體來說, RDM 要做什麼



Q: 在 Research Data Lifecycle 中,
哪些環節對你來說比較重要?

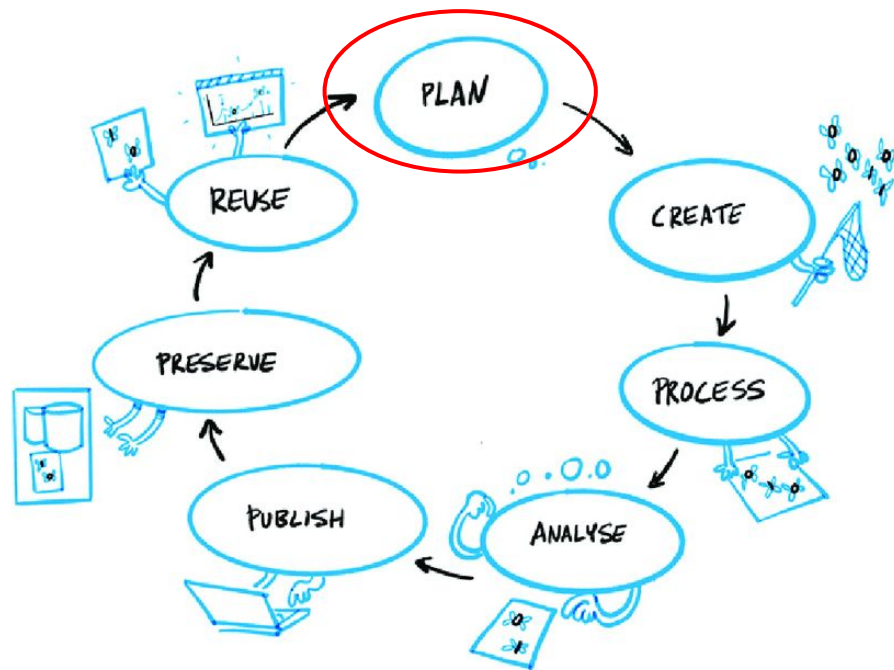


Menti code: 22306763

<https://www.menti.com/>




為了讓資料 FAIR, 可記住以下五個要點

1. 讓 DMP 成為對自己有用的工具



Tool for Plan Phase -

資料管理方案 (Data Management Plan, DMP)

  <p>Horizon Europe</p> <p>Data Management Plan Template</p> <p>Version 1.0 05 May 2021</p>	 <p>Data and Digital Outputs Management Plan Template</p> <p><u>Introduction</u></p> <p>The Belmont Forum supports multi-national and transdisciplinary collaborative research together natural sciences, social sciences and the humanities, as well as stakeholder knowledge and solutions for sustainable development. The Belmont Forum Challenge international transdisciplinary research providing knowledge for understanding, mitigation and adaptation to global environmental change.</p> <p>To meet this challenge, the Belmont Forum emphasizes open sharing of research data outputs to stimulate new approaches to the collection, reuse, analysis, validation, a data and information, thus increasing the transparency of the research process and results.</p> <p>Research data and digital outputs include, but are not limited to:</p> <ul style="list-style-type: none">• Quantitative and qualitative digital information and objects created during research activities such as experiments, analyses, surveys, interviews, measurement instrumentation, observations, video, audio, and computer simulations;• All metadata describing the data and digital outputs, their acquisition (including description and related metadata for simulations and workflows), and other and the reuse of the data;• Secondary data resulting from data reduction, transformation, analyses, and with the associated code, software, workflows, and provenance information;• Stakeholder-oriented digital outputs such as maps (including GIS layers), datasets, tutorials, videos, local language resources, lesson plans, curricula, policy memos, whitepapers; and• Descriptions of, and metadata relating to, physical samples connected with the actual physical samples. <p>Each project awarded through a Collaborative Research Action is required to develop a Data and Digital Outputs Management Plan to ensure ethical approaches and comply with the Belmont Forum Open Data Policy and Principles, as well as the FAIR Data Principles Accessible, Interoperable, and Reusable).</p> <p>The Belmont Forum is in the process of gradually integrating its Open Data Policy into its CRA process. This evolving process includes both the acculturation of researchers and agencies to open data practices, and the increasing momentum toward transdisciplinary change research. For example, the Belmont Forum recognizes that some funding agencies have their own data and digital output management requirements and that further specific guidance may be needed for both proposers and funding agencies to address potential issues. However, the Belmont Forum expects that proposers will make every effort to</p>	<p>National Institute of Standards and Technology (nist.gov): NIST: National Institute of Standards and Technology Data Management Plan</p> <p>Summary of activities for data generation</p> <p>summary of activities that generate data</p> <p><i>Guidance:</i></p> <p>Data sharing in this policy refers to final research data. These data are the recorded factual materials commonly accepted in the scientific community as necessary to document and support research findings. This policy applies to new data collection as well as to data obtained through transforming or linking existing datasets. For most studies, an electronic file will constitute the final research data. This dataset will include both raw data and derived variables, which will be fully described in accompanying documentation section.</p> <p>Consider these questions:</p> <ul style="list-style-type: none">• How will you capture or create the data?• Are you pulling from existing data sources? <p>Data types and classification</p> <p>a summary of the data types generated by the identified activities. Data should be categorized, at a minimum, according to the data categories presented in the NIST Data Taxonomy and Actions/Consequences for Data Categories, provided in Appendix A of this Order, as applicable.</p> <p><i>Guidance:</i></p> <p>Describe how your data will be "documented." Think about what information is needed for the data to be read and interpreted in the future. What would someone else need to be able to use these files? The documentation should include a summary of the purpose of the data collection, methodology and procedures used to collect the data, timing of the data collection, as well as details of the data codes, definition of variables, variable field locations, and frequencies. The data documentation should be a comprehensive and stand alone document that includes all the information necessary to replicate the analysis performed by the original research team.</p> <p>Consider these questions:</p> <ul style="list-style-type: none">• What data will you collect or create in the research?• What data types will you be creating or capturing and what data will be shared?• What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?• What contextual details (metadata) are needed to make the data you capture or collect meaningful?• How will you create or capture these details? <p>Preservation</p> <p>a plan for storage and maintenance of the data generated by the identified activities, in both the short-term and long-term (if relevant). Data should be preserved, at a minimum, according to the preservation consequence levels defined in the NIST Data Taxonomy and Actions/Consequences for Data Categories, provided in Appendix A of this Order, as applicable.</p>
--	--	--

國科會永續學門 - 鼓勵納入 DMP

五、重要注意事項：(詳如附件1中說明)

(一)整合型計畫：整合團隊必須有三位以上之總/子計畫主持人組成，並由各主持人服務單位送出計畫書申請。計畫書審查後，每一整合型團隊必須有三位以上(含總/子計畫主持人)通過，總主持人計畫通過為必要條件。並將「整合型研究計畫自我檢核表」納在CM04中。

(二)計畫書請依附件所列議題與研究主題擬計畫內容，並於計畫書中敘明所提研究內容所依據之研究議題、主題與研究方向。整合型計畫須於「整合型研究計畫項目及重點說明(表CM04)」及「研究計畫中文摘要(表CM02)」中列明所依據之議題與研究主題；個別型計畫則於「研究計畫中文摘要(表CM02)」中列明。

(三)為落實跨領域研究(TDR)精神，並強化原住民族相關研究合宜性。若研究內容涉及「原住民或原住民族土地或部落及其周邊一定範圍內土地」之計畫，請依原住民族基本法第21條第1項中所述原則，於計畫書內規劃相關事宜。

(四)為促進研究計畫之科學資料產製品管、資料保存及研究成果資料分享。鼓勵於研究計畫納入「研究資料管理方案(Data Management Plan, DMP)」。

(四)為促進研究計畫之科學資料產製品管、資料保存及研究成果資料分享。鼓勵於研究計畫納入「研究資料管理方案(Data Management Plan, DMP)」，說明如下：

1.DMP 之規劃請以專章列於計畫書中；整合型計畫請列入「四、整合型研究計畫項目及重點說明(CM04)」、個別型計畫則列入「研究計畫內容(CM03)」。

2.DMP 並非資料庫建置，撰擬原則詳參中研院研究資料管理推進室 (Research Data Management Hub; <https://rdm.depositar.io>) 或洽詢學門承辦人。

3.DMP 所需經費編列於計畫經費中並於說明註明，學門將就獲多年期核定之計畫且通過 DMP 審核之計畫提供經費支持（個別型計畫係自行編列，整合型計畫則由總計畫統籌編列，額度以該計畫/整合型團隊核定額度總數十分之一為限）。

資料管理方案 (data management plan, DMP)

資訊集中

所有人一起使用

XXX 計畫

資料管理方案

資料管理方案是一份描述
你在整個研究生命週期的
過程中，如何關照資料的
文件

活的

視需求隨時更新

處理歷程

將流程文件化

資料產出

有效資料的概述

DMP 是分享用的文件

至少你要和未來的自己分享

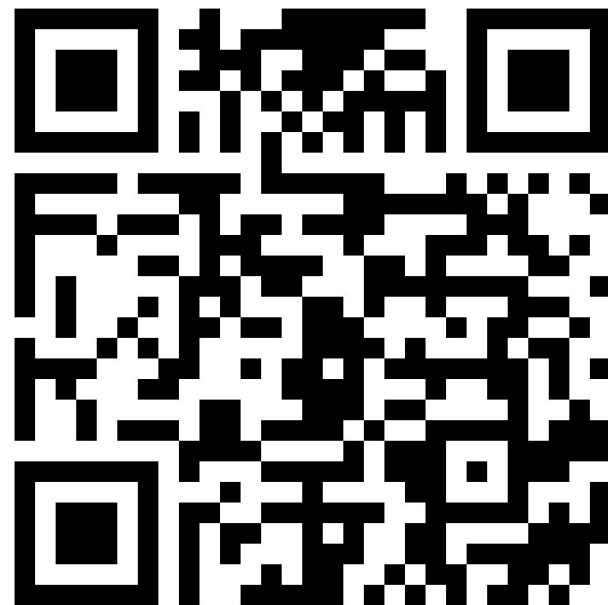


國際合用的研究資料管理實用指南 一增訂版

2021/01 : Science Europe 出版

2021/07 : 研究資料寄存所 (depositor) 團隊中文化完成

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



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

DMP 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



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。

若想參考公開的 DMP...

- **DMP Online**

- 英國, 數位策展中心維護
- <https://dmponline.dcc.ac.uk/>

- **DMPTool**

- 美國, 多所典藏機構共同維護
- <https://dmptool.org/>

- **DMP Assistant**

- 加拿大, 由 Portage Network 維護
- <https://assistant.portagenetwork.ca/>

Establishing a collection of 841 publicly available Horizon 2020 Data Management Plans

scholarly communication | Gerda McNeill | 22 Jan 2021 | 6954 Hits | 0 Comments



<https://www.openaire.eu/blogs/establishing-a-collection-of-841-horizon-2020-data-management-plans>

公開的中文 DMP

資料管理方案 (2022 TaiBIF 生物多樣性資料發布與應用工作坊) | Data Management Plan (2022 TaiBIF Biodiversity Data Release and Application Workshop)

本資料集為 2022/09/28 TaiBIF 生物多樣性資料發布與應用工作坊第三天資料管理課程，學員實作資料管理方案 (DMP) 所產出的資料集。

資料管理方案 (DMP) 表單 (Template) 內容取自 2021 年 Science Europe 出版的《國際合用的研究資料管理指南——增訂版》(Practical Guide to The International Alignment of Research Data Management - Extended Edition) 的資料管理方案「核心要求」與「根據核心要求而設計的資料管理方案範本」，其包含基本六大要求共 15 小題。

此資料集所釋出的資料管理方案，皆由學員依自身研究資料管理上實作經驗分享，並同意以 CC 授權釋出供大眾參考再次利用。

資料管理方案共計 25 份，課程講師依最後的評比、學員意願及授權，篩選出 2 份資料管理方案釋出於此資料集。期盼透過開放及分享，能啟發有研究資料管理需求的研究者有所參考依據，追求研究資料發展、分享與維護最佳實踐。

資料與資源



DMP_郭景嘉_v2.0

此為資料管理方案為東方毛腳燕 (Delichon dasypus) 的繁殖生態學調查，...



DMP_郭景嘉_v2.0

此為資料管理方案為東方毛腳燕 (Delichon dasypus) 的繁殖生態學調查，...



2022



2023

資料管理方案 (2023/05/13 TaiBIF 生物多樣性資料發布與應用 工作坊) | Data Management Plan (2023/05/13 TaiBIF Biodiversity Data Release and Application Workshop)

本資料集為 2023/05/13 TaiBIF 生物多樣性資料發布與應用工作坊第三天資料管理課程，學員實作資料管理方案 (DMP) 所產出的資料集。資料管理方案 (DMP) 表單 (Template) 內容取自 2021 年 Science Europe 出版的《國際合用的研究資料管理實用指南——增訂版》(Practical Guide to The International Alignment of Research Data Management - Extended Edition) 的資料管理方案「核心要求」與「根據核心要求而設計的資料管理方案範本」，其包含基本六大核心要求共 15 小題。

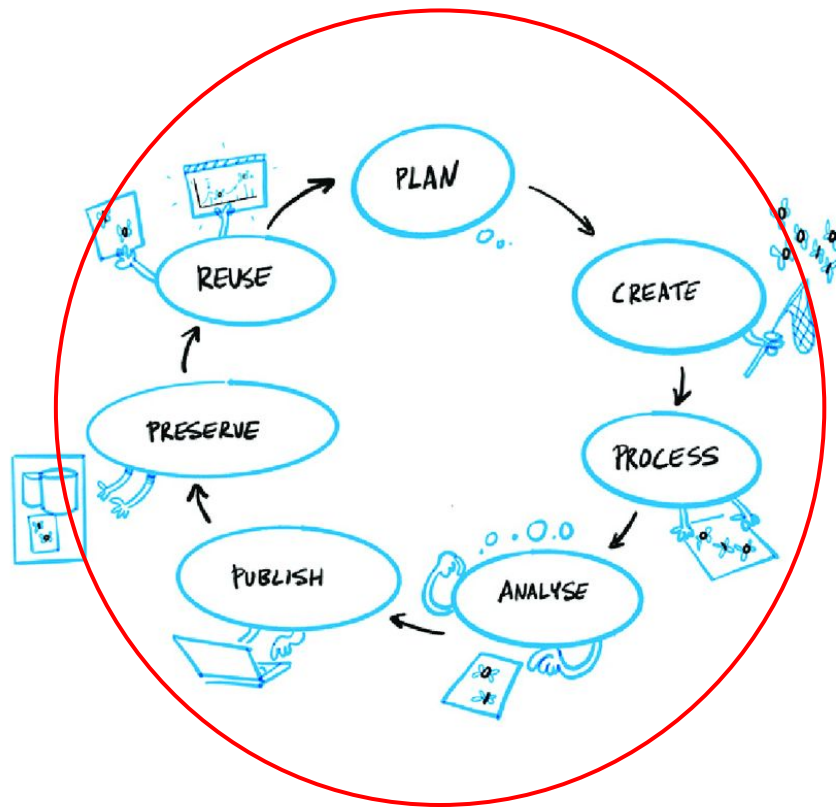
此資料集所釋出的資料管理方案，皆由學員依自身研究資料管理上實作經驗分享，並同意以 CC 授權釋出供大眾參考再次利用。

資料管理方案共計 23 份，課程講師依最後的評比、學員意願及授權，篩選出 5 份資料管理方案釋出於此資料集。期盼透過開放及分享，能啟發有研究資料管理需求的研究者有所參考依據，追求研究資料發展、分享與維護最佳實踐。

什麼是一份好的 DMP?

- 將 FAIR 放在心裡
 - FAIR is not just the goal, it's also the process. There are many stages between FAIR and NOT FAIR
- 將 Openness 放在心裡
 - as open as possible, as closed as necessary
- 「可行」比「理論上很好」重要
 - do not promise something you can't keep
- 「具體」比「我想要...」重要
 - make sure you know how to implement what you mentioned
- 考量利害關係人的角色
- 定時更新
 - DMP is a tool to check and improve your workflow!

2. 儘可能為你的資料提供充分的後設資料



什麼是「後設資料」(metadata)

- 後設資料是用以描述、解釋、查找或使一份資訊更容易被擷取、使用、管理的**結構化**資訊。後設資料時常被稱為**有關資料的資料**。

(*Understanding Metadata*, National Information Standards Organization, 2004)

- 「結構化」意指使用「屬性名-屬性值」(key-value)的方式來描述該資料:如標題、作者、摘要、關鍵字、格式、主題、PID、版權、溯源等。

The image shows a screenshot of a metadata form interface. It contains several sections:

- 標題:** A text input field with the placeholder text "例如：一個描述性的標題" and a URL "網址: data.depositar.io/dataset/<dataset>" with a "編輯" button.
- 摘要:** A larger text area with the placeholder text "關於資料集簡單扼要的說明" and a note "您可以在這裡使用 Markdown 格式".
- 資料類型:** A list of checkboxes for various data types: 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, and 其他.
- Wikidata 關鍵字:** A text input field with the placeholder text "加入維基資料項目以描述資料集..." and a note "搜尋 Wikidata (維基資料) 項目並取關鍵字以描述資料集。只用於您專業或資料集的標記 (如：計畫編號)，請填寫於「標識、欄位」。
- 標識:** A text input field with the placeholder text "例如：第一期計畫、與 Joe 的訪談" and a note "只用於您專業或資料集的標記。".
- 語言:** A text input field with the placeholder text "搜尋 ISO 639-3 語言編碼..." and a note "資料集內容所使用之語言。ISO 639-3 語言編碼請參見列表。".

為什麼後設資料很重要

- 資料更容易被分享、再利用、保存
- 資料更容易被理解
- 別人不會誤會你的資料或誤用你的資料
- 別人可以藉著這些後設資料，更容易找到你的資料



<https://images.app.goo.gl/pWfmkzE6E2iAVpDx9>

"data without the contextual information needed to interpret it (and ultimately reproduce the results) is useless" - B. Marshall [Metadata for Research Data](#)

台中市南屯區鎮平溪 - 劉厝溪航攝影像

追蹤者

0

+ 跟隨

📖 專案



Asian Ecological Observation Network

此專案沒有任何說明

🔗 社交

🐦 Twitter

🗂 資料集

👥 主題

🕒 動態牆

📄 展示案例

🔧 管理

資料與資源



航拍規劃中心線

↪ 探索 ▾



2021-04-13 台中市南屯區鎮平溪 - 劉厝溪航攝影像 🔥

↪ 探索 ▾



OAM正射影像連結 (Link to OpenAerialMap) 🔥

↪ 探索 ▾



2021-04-13 正射影像Google圖磚 (Google Earth tiles) 🔥

↪ 探索 ▾

標籤

南屯區

台中市

地景變遷

筏子溪

都市重劃

Wikidata 關鍵字

正射影像

河景

筏子溪

光球

南屯區

無人航空載具

基本資訊

資料類型

- 壓縮檔資料
- 結構化圖形
- 影像
- 網路通訊資料
- 科學與統計資料

語言

中文 (zho)

產製者

王豫煌

資料產製時間

2021-04-18

資料處理歷程

航攝製圖影像拍攝

採用DJI Phantom 4 RTK無人載具及D-RTK 2 Mobile Station連接國土測繪中心DGPS定位服務進行影像拍攝；航線規劃採用Linear飛航模式，沿中心線KML檔向左右兩側80公尺設定為航拍範圍；鏡頭俯角90度，航高保持離起降點地面60公尺，影像重疊率前後80%、左右70%。

影像處理

採用中央研究院網格計算中心WebODM 測試平台，設定High Resolution計算模式，輸出正射影像、DSM、3D點雲和模型，再經由中央研究院網格計算中心WebODM平台發布計算成果網頁連結。

360影像拍攝與發佈

使用DJI Mavic 2 Pro無人載具和DJI GO4 App拍攝空中360全景影像，並將影像發佈至Google Street View。

聯絡人

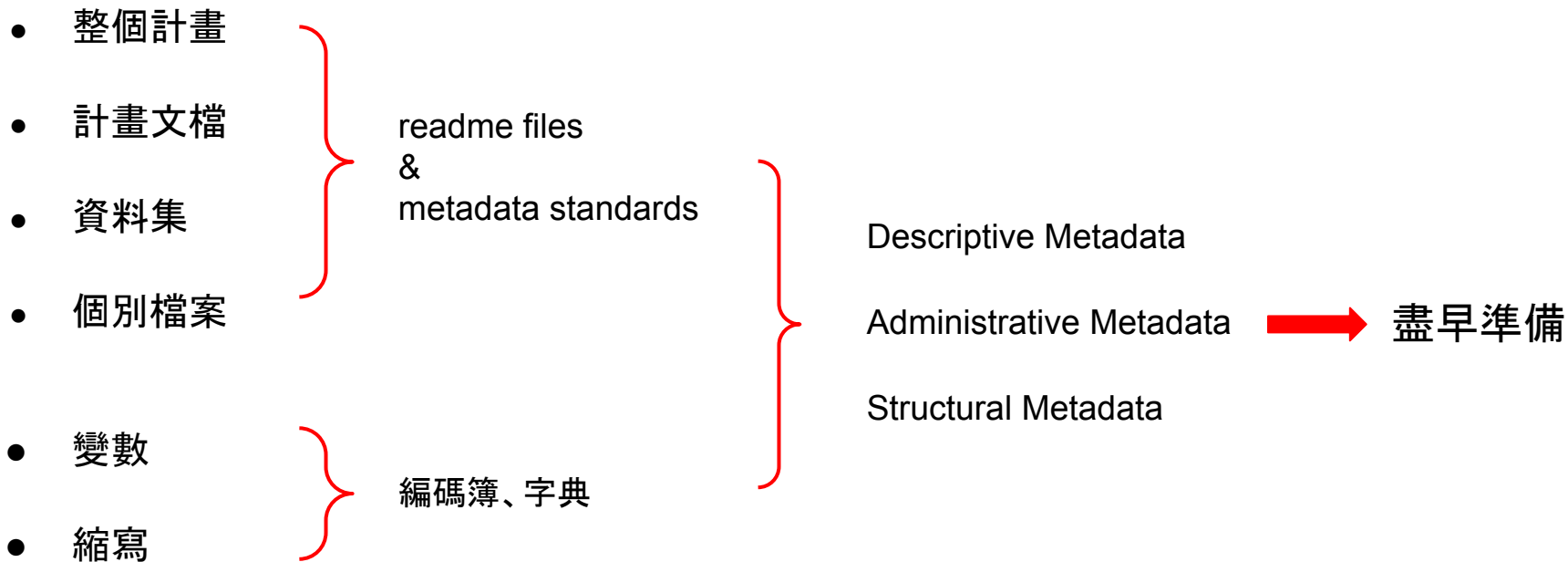
Yu-Huang Wang

聯絡人的電子郵件

yuhuangwang@gmail.com

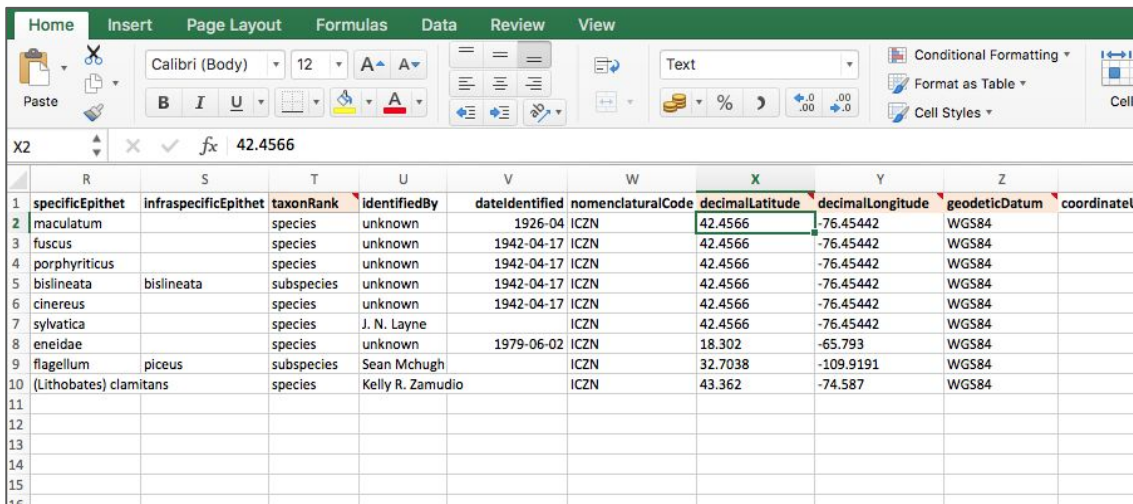
不同層次的 metadata

你想在哪個層次上描述你的資料？如何描述？



Metadata Standards

- Metadata standards often start as schemas developed by a particular user community to enable the best possible description of a resource type for their needs. (DCC)
- 採納後設資料標準，可提升 machine readability，從而改善 interoperability, findability



The screenshot shows an Excel spreadsheet with a metadata table. The table has columns for specific epithet, infraspecific epithet, taxon rank, identified by, date identified, nomenclatural code, decimal latitude, decimal longitude, geodetic datum, and coordinate system. The data rows include species like *Imaculatum*, *fuscus*, *porphyriticus*, *bisilineata*, *cinereus*, *sylvatica*, *eneidae*, and *flagellum piceus*.

	R	S	T	U	V	W	X	Y	Z	coordinate
1	specificEpithet	infraspecificEpithet	taxonRank	identifiedBy	dateIdentified	nomenclaturalCode	decimalLatitude	decimalLongitude	geodeticDatum	
2	maculatum		species	unknown	1926-04	ICZN	42.4566	-76.45442	WGS84	
3	fuscus		species	unknown	1942-04-17	ICZN	42.4566	-76.45442	WGS84	
4	porphyriticus		species	unknown	1942-04-17	ICZN	42.4566	-76.45442	WGS84	
5	bisilineata	bisilineata	subspecies	unknown	1942-04-17	ICZN	42.4566	-76.45442	WGS84	
6	cinereus		species	unknown	1942-04-17	ICZN	42.4566	-76.45442	WGS84	
7	sylvatica		species	J. N. Layne		ICZN	42.4566	-76.45442	WGS84	
8	eneidae		species	unknown	1979-06-02	ICZN	18.302	-65.793	WGS84	
9	flagellum	piceus	subspecies	Sean Mchugh		ICZN	32.7038	-109.9191	WGS84	
10	(Lithobates) clamitans		species	Kelly R. Zamudio		ICZN	43.362	-74.587	WGS84	
11										
12										
13										
14										
15										

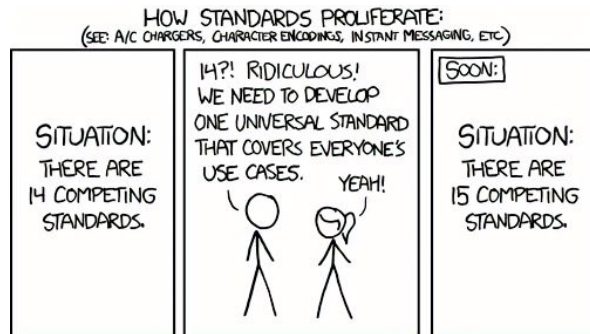
如何找適合的 metadata 標準？

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：

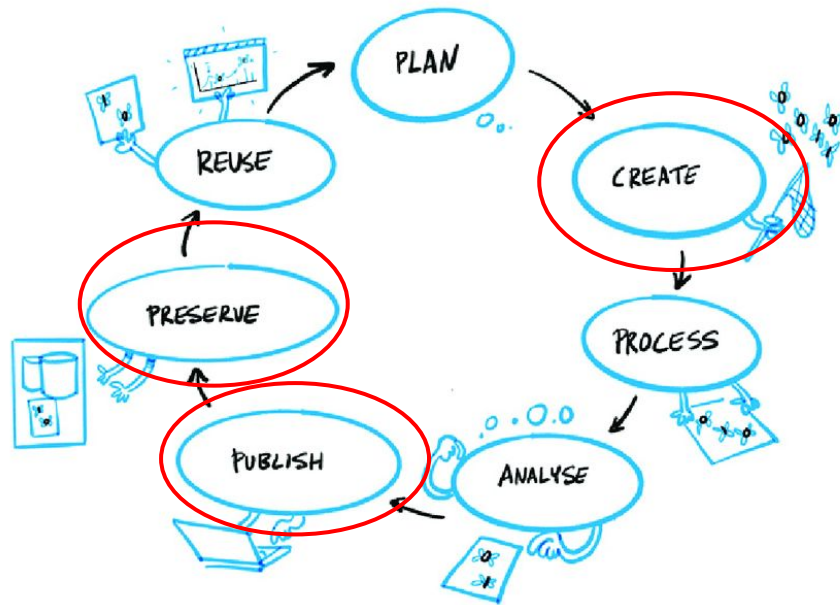
- RDA Metadata Directory: <https://rd-alliance.github.io/metadata-directory/standards/>
- Bath Metadata Standard Catalogue: <https://rdamsc.bath.ac.uk/>

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 - DCAT @dataset level



3. 確保你已取得或將釋出的資料在法律、授權、倫理上是沒問題的





[認識行政院 +](#)

[新聞與公告 +](#)

[政策與計畫 +](#)

[資訊與服務 +](#)

[便民服務 +](#)

[任務編組](#)



[首頁](#) > [政府網站資料開放宣告](#)

政府網站資料開放宣告

政府網站資料開放宣告



一、授權方式及範圍：

為利各界廣為利用網站資料，行政院網站上刊載之所有資料與素材，其得受著作權保護之範圍，採政府資料開放授權條款-第1版發布，以無償、非專屬、得由使用者再授權之方式提供公眾使用，使用者得不限時間及地域，重製、改作、編輯、公開傳輸或為其他方式之利用，開發各種產品或服務（簡稱加值衍生物），此一授權行為不會嗣後撤回，使用者亦無須取得本機關之書面或其他方式授權；然使用時應註明出處。

二、相關事項說明：

- (一) 本宣告範圍僅及於著作權保護之範圍，不及於其他智慧財產權利（包括但不限於專利、商標、及機關標誌之提供）。
- (二) 當事人自行公開或依法令公開之個人資料是否得被蒐集、處理，及利用，使用者須自行依照個人資料保護法之相關規定，規劃並執行法律要求之相應措施。
- (三) 部分的影音、圖像、樂譜、專人專案撰文或其他著作，經機關特別聲明須經同意方可使用者，不在本宣告所及範圍，其後續使用應另經機關同意。

三、應注意尊重第三人之著作人格權（包括姓名表示權及禁止不當變更權）。

四、使用本宣告提供之資料與素材，不得惡意變更其相關資訊，若利用後所展示之資訊與原資料與素材不符，且得被依法歸責，使用者須自負民事、刑事上之法律責任。

五、本網站之宣告，並不授與使用者代表本機關建議、認可或贊同其加值衍生物之地位。

個人資料處理的基本原則

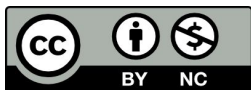
- 目的特定原則
- 個資最小原則
- 知情同意
- 適當的安全措施



FIPs: 基本的資訊處理原則

如何授權你的資料？

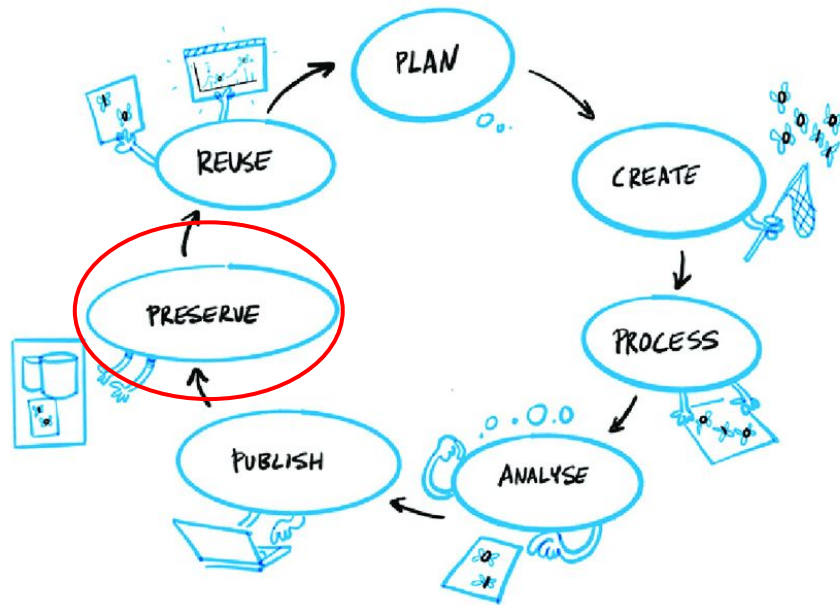
建議使用開放授權，如 Creative Commons



CC0 (公眾領域貢獻宣告)

- 協助選擇開放授權的工具：
 - <https://creativecommons.org/choose/>
 - <https://ufal.github.io/public-license-selector/>
 - <https://choosealicense.com/>

4. 提早設想適合長期保存資料
的資料儲存庫，可節省許多心力



提早設想，因為還需考量： 要保存什麼？

- 資料是否和已發表的研究或報告有關？
- 資料是否容易再次生產？或者需要耗費大量成本？



- 規定應保留的資料
- 獨特或難以再次生產的資料(原始資料、資料分析的流程等)
- 研究再現所需的資料(整理後的資料、儀器設定檔、程式碼等)
- 未來有再次使用潛力的資料
- 對社會整體有益處的資料

資料要保存在哪？



自己的電腦、網站、雲端硬碟



資料儲存庫

表五 研究資料儲存位置分析

研究資料儲存位置（複選題）	儲存位置		供長期保存與 取用儲存位置	
	樣本數	觀察值百分比	樣本數	觀察值百分比
自己的儲存裝置	1,030	96.3	841	78.7
雲端硬碟	600	56.1	670	62.7
服務機構所建置的研究資料儲存庫	219	20.5	497	46.5
學科領域常用的研究資料儲存庫	153	14.3	372	34.8
研究專案之主要資助機構所建置或支持的 研究資料儲存庫	-	-	362	33.9

為何要使用 data repositories

什麼是資料儲存庫？

資料儲存庫(至少)是可搜尋的資料集目錄。研究者可以在開放的資料儲存庫中描述其資料，並使其資料有效。

- 減輕研究者應付資料再利用請求的負擔，並讓資料可被取用
- 提供資料取用和安全儲存的保證
- 讓你的資料可以被找到 (e.g., 讓資料在 Google Dataset Search 被找到)
- 可以持續讓你的資料 [FAIR](#) (Findable, Accessible, Interoperable and Reusable)
- 通常會比個別研究者更在意資料可否被持續保存下去

你需要什麼資料儲存庫(data repository)？

- 期刊、機構、補助機構推薦或指定
- 領域的習慣
- 通用型 (Zenodo, 4TU, DANS, depositar...etc.)

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**



GBIF

Global Biodiversity
Information Facility

怎麼找資料儲存庫？

名錄：

- re3data (Datacite) :
<https://www.re3data.org/>
- FAIRsharing: <https://fairsharing.org/>
- DataCite Commons:
<https://commons.datacite.org/>



挑選合適的資料儲存庫 - Science Europe 選用準則

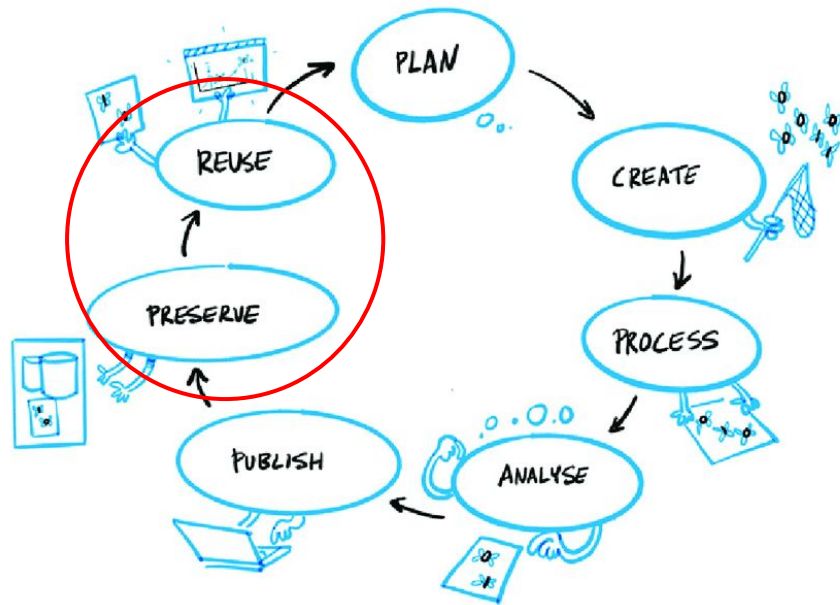
- 持續識別碼 PID
(persistent identifier, e.g.,
DOI, ARK)
- 後設資料
- 資料取用及授權
- 資料保存

可信賴的資料儲存庫 的選用準則

可信賴的儲存庫應滿足以下最低準則：

- 1. 提供持續及唯一的識別碼 (Persistent and Unique Identifiers, PIDs)
 - a. 能進行資料探索 (discovery) 和識別
 - b. 能搜尋、引用、和擷取資料
 - c. 提供資料版本控制的支援
- 2. 後設資料
 - a. 讓資料可供查找
 - b. 可以參照至相關聯的資訊，如其它資料、出版物等
 - c. 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
 - d. 使用被（科學社群）廣泛接受的後設資料格式
 - e. 確保後設資料能夠以機器擷取

5. 為資料訂定一個合適的 釋出時間



When to share

- 有資料就釋出
- 著作出版時
- 計畫結束時
- 資助者規定的時間
- 設定閉鎖期 (embargo)
 - 目前通常不超過一年，並朝立即開放的趨勢前進

OSTP 2022 Memo - Scientific data underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication, unless subject to limitations as described in Section 3(c)(i) and should be subject to federal agency guidelines for researcher responsibilities regarding data management and sharing plans, consistent with Section 3(c) of this memorandum.

NIH - Shared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of performance period, whichever comes first.

Horizon Europe - as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle ‘as open as possible as closed as necessary’, unless providing open access would in particular: [下略]

ESRC - formally deposit all data created or repurposed during the lifetime of the grant with a responsible data repository within three months of the end of the grant

9307 自然處海洋學門製

1030304 修

1030430 修

1110322 修

1110811 修

國家科學及技術委員會海洋學門海洋量測資料繳交與釋出規定

1. 原則

- (1) 本規定所指之資料為接受國家科學及技術委員會(以下簡稱本會)補助之研究計畫所蒐集的海洋量測資料，並僅適用於本規定施行後所核定的研究計畫。
- (2) 資料之繳交為計畫執行的一部份，資料繳交的履行為未來研究計畫核定的參考依據。
- (3) 研究計畫主持人應於規定之繳交期限內提供完整之量測報告與數據至本會指定的海洋資料庫(以下簡稱資料庫)。報告中請詳述資料應有的基本訊息，如資料型式、量測時間、處理過程等。資料庫將以統一格式處理繳交之資料，並適時公佈之。
- (4) 資料使用者應依學術倫理給予資料提供者適當的尊重(如列為共同作者、致謝等)。

研究資料管理推進室 (RDM Hub)

<https://rdm.depositar.io>

- 分享與探討各領域所遭遇的資料議題與困難，以促進研究資料管理在台灣的發展。
 - 實踐分享
 - 資源蒐集
 - 教育訓練

Research Data Management Hub

資源 實踐 消息 關於

推動聲景生態調查——專訪林子皓博士

我們其實希望其他人用我們的 code，搭配上他們的想法後，將他們的 code 也發佈出來。這會是一種很有趣的資料應用狀況，可以看各研究團隊，基於同一份資料，為了不同目的，如何去做分析。

VIEW DETAILS

撰寫資料管理方案 (DMP)

資料管理方案 (Data Management Plan, DMP) 是一份描述研究資料將如何被蒐集、使用、管理、(短期或長期) 保存、分享等歷程的文件。DMP 時常是研究團隊在資料管理上的第一步。

Read More

挑選可信賴的資料儲存庫

研究人員需能分辨可信賴的資料儲存庫，以便妥善儲存和共享資料。對於研究人員及其機構、資助機構而言，識別合適的資料儲存庫會是一項富有挑戰的任務。

Read More

研究資料管理的好工具

各學科可能有自身慣用的 RDM 工具。許多研究機構亦會提供一些需付費方能使用的工具，研究者可洽詢機構取得更完整的工具協助。

Read More

經驗分享

Grid of video thumbnails showing various research data management topics and speakers.

謝謝！ Thank You!

<https://data.depositar.io/> 研究資料寄存所
<https://rdm.depositar.io/> 研究資料管理推進室

data.contact@depositar.io
<https://lab.depositar.io/>

The depositar is a collaboration at the Institute of Information Science, the Research Center for Information Technology Innovation, and the Research Center for Humanities and Social Sciences (GIS Center) in Academia Sinica, Taiwan. The project has been supported, in part, by grants from Taiwan's National Science and Technology Council.

The *depositar* project team: T-R Chuang, M-S Ho, C-J Lee & C-H Ally Wang.

「研究資料寄存所」是中央研究院資訊科學研究所、資訊科技創新研究中心、人文社會科學研究中心(地理資訊科學研究專題中心)的協作專案, 部份經費來自台灣國科會的專題研究計畫。

研究資料寄存所計畫成員: 莊庭瑞、何明諳、李承鑫、王家薰。

