

甚麼是研究資料寄存所?

What is the *depositar*?

2023-06-26 @ NCHC

莊庭瑞 Tyng-Ruey Chuang

研究資料寄存所實驗室 the depositar lab
<https://lab.depositar.io>

台灣 中央研究院
Academia Sinica – Taiwan

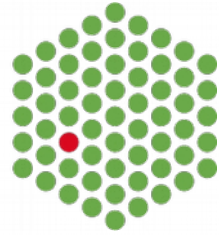


您的資料為何要放在公共的儲存庫？

Why Put Your Data In A Public Repository?

- 資料能找得到、能被引用、並能持續保存、可再使用
- 確認後設資料 (metadata) 與來源資訊 (provenance)
- 因為資料要開放，也同時提昇研究資料管理實務
- 論文出版時的要求；資料（論文）發布的需要
- 研究計畫資助機構的要求
- 資料的協同產製使用；「開放科學」 (“Open Science”)
- 經由開放儲存庫的資料介面，發展新的應用與資訊服務

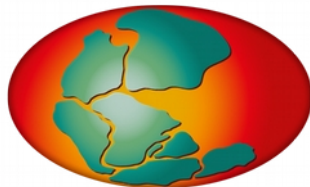
EMBL-EBI



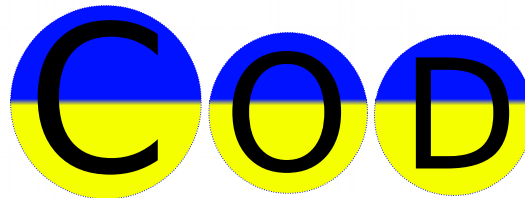
GBIF

Global Biodiversity
Information Facility

HUMANITIES
COMMUNIS



PANGAEA.



GitHub



d. **depositar**
研究資料寄存所

- 
- EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute) @ re3data
 - GBIF (Global Biodiversity Information Facility) @ re3data
 - Humanities Commons (Commons Open Repository Exchange, CORE) @ re3data
 - Pandora @ re3data
 - PANGAEA (Data Publisher for Earth and Environmental Science) @ re3data
 - Crystallography Open Database @ re3data
 - GitHub @ re3data
 - Zenodo @ re3data
 - depositar 研究資料寄存所 @ re3data

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

資料管理方案的核心需求

CORE REQUIREMENTS FOR
DATA MANAGEMENT PLANS

可信賴的儲存庫的選用準則

CRITERIA FOR THE SELECTION OF
TRUSTWORTHY REPOSITORIES

譯註：

資料管理方案 ↔ Data Management Plans (DMPs)
研究資料儲存庫 ↔ Research Data Repositories

提案 ↔ Proposal

專案 ↔ Project ↔ 計畫 (項目)

方案 ↔ Plan



可信賴的儲存庫的選用準則 (1/2)

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

可信賴的儲存庫的選用準則 (2/2)

- 資料取用及使用授權
 - 讓資料取用是在明確的條件下進行
 - 確保資料的真實性及完整性
 - 讓資料可被擷取
 - 提供關於授權及許可的資訊（使用機器可讀格式最為理想）
 - 確保資料當事人及創作者的保密性與權利
- 資料保存
 - 確保後設資料和資料的持續保存
 - 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）

FAIR 資料原則

Fair Data Principles

F

Findable
可被找到

資料以豐富的后設資料描述；
資料在可搜尋的資源中註冊或索引

A

Accessible
可被取用

能使用標準化通訊協定，並以識別碼擷取資料；
該通訊協定是開放的，而且可自由並普遍實施

I

Interoperable
可相互操作

資料係使用正式、可取用、可共享和廣泛接受的
語言以達成知識再現

R

Reusable
可再次使用

以精確且相關的多種屬性，豐富地描述資料；
以清晰、易懂的資料授權條款釋出資料

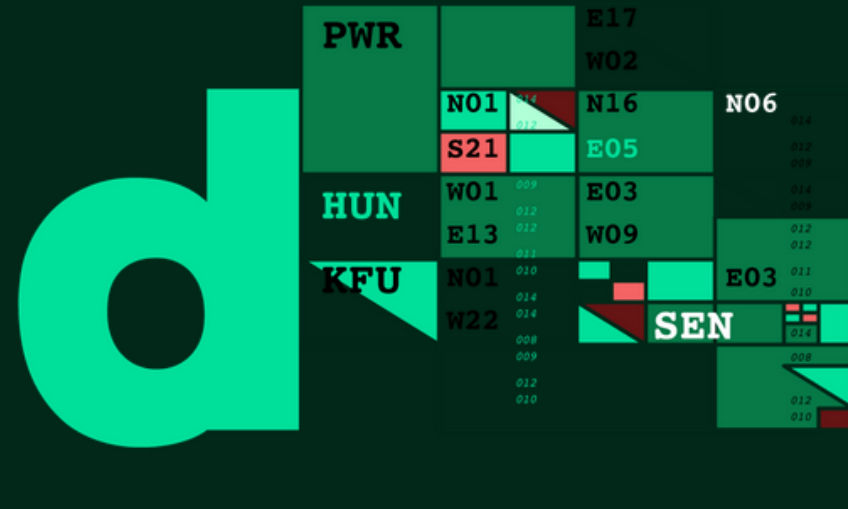
<https://www.force11.org/group/fairgroup/fairprinciples>

儲存 · 尋找 · 分享

資料集列表

上傳資料集

建立新專案



Well managed and preserved research data is the cornerstone of reproducible research.

Let's practice the FAIR data principles together. May all research data be findable, accessible, interoperable, and reusable!

Open and Free

The data repository is built on top of the open source CKAN package. It has been customized and extended to support research data management. The service is open to all researchers. Registration is free!

More

Flexible

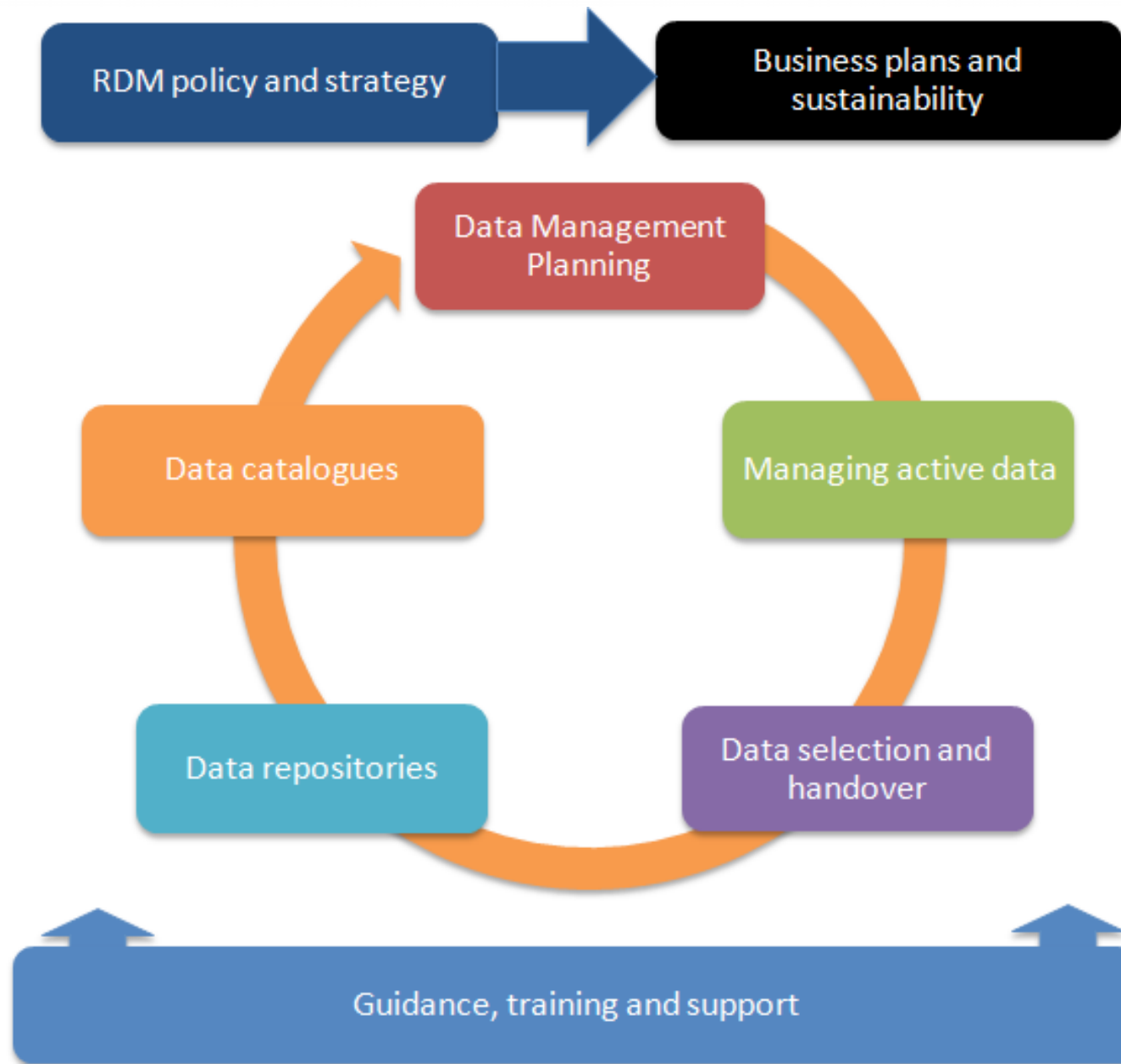
All kinds of data can be deposited. Datasets can be searched by spatiotemporal ranges, data types, keywords, and other conditions. The datasets are indexed by [Google Dataset Search](#).

More

Interoperable

Resource catalogs and data endpoints are supported by Web APIs. APIs to access structured data (e.g. CSV and Excel files) are available too. Programmable data access and analytics can be implemented.

More



US OSTP Memo and NSF Policies

- **"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, ..."
 - OSTP "Nelson Memo", The White House (2022-08-25)
- **"[DMPs] are currently a required component of an NSF proposal.** To ensure that funded researchers have considered data deposition and appropriate repositories, NSF plans to more explicitly and systematically require researchers to include, as part of these DMPs, **details about how the data produced will be shared publicly and with other researchers and anticipated repositories for the data, ..."**
 - NSF "Public Access Plan 2.0" (2023-02-21)

@_depositar



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

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、何明諠、李承鑫、王家薰。

