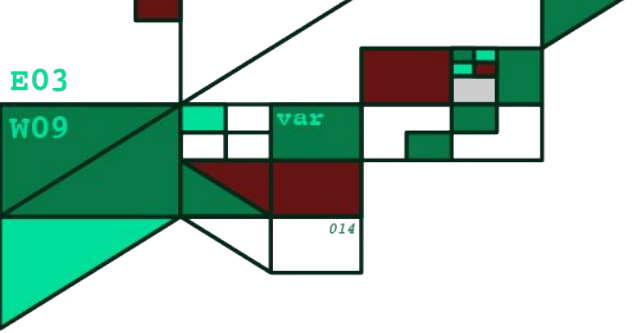


Brief Introduction to Research Data Management

2024-03-29

AMD Computing & User Training Workshop for NSTC Core Computing Service 2024

何明諠 (Ming-Syuan Ho)｜Research Center for Information Technology Innovation｜Project Manager
hmsyuan@citi.sinica.edu.tw

depositar
研究資料寄存所

depositar is a collaborative project supported by the Institute of Information Science and the Research Center for Information Technology Innovation (both at Academia Sinica), and in part by research grants from the National Science and Technology Council of Taiwan.。https://data.depositar.io/about

何明諠

專案經理

目前任職於資訊科技創新研究中心，是協助執行資料管理方案（DMP）的聯絡窗口之一，同時也致力於資料政策分析，並提倡研究資料管理（RDM）。此外，他在中央研究院也參與了另一個關於隱私與資安的計畫。在加入 depositar 團隊前，他在台灣人權促進會致力於推動數位人權議題。

李承鑫

技術經理

現任職於中央研究院資訊科學研究所。目前擔任「研究資料寄存所」(depositar / data.depositar.io) 的技術經理，負責該平台開發工作。近期研究主軸為資料流通標準、資訊系統自動化部署與長期維運，期許透過導入相關技術，以因應研究資料寄存所的持續發展需求。多年 Python 使用者。

王家薰

專案經理

王家薰目前任職於中央研究院資訊科學研究所，曾參與在中央研究院執行的自由軟體鑄造場及台灣創用 CC 計畫。具工程師的背景並熱愛開放自由的風氣，關注數位保存及研究資料管理相關資訊，目前為研究資料管理計畫的專案經理，致力於推廣研究資料管理的概念並向下扎根。

莊庭瑞

計畫主持人

莊庭瑞為中央研究院資訊科學研究所副研究員，同時合聘於人文社會科學研究中心（地理資訊科學研究專題中心）以及資訊科技創新研究中心。

depositar lab:
https://lab.depositar.io/

What is research data?

- Any information that has supported and validated the original research findings.

- Research data is very important, so we try to ensure it can be reused at least by ourselves.

General	Social Sciences	Natural/Physical Sciences
• images • video • mapping/GIS data • numerical measurements	• survey responses • focus group and individual interview transcripts • economic indicators • demographics • opinion polling	• measurements generated by sensors/laboratory instruments • computer modeling • simulations • observations and/or field studies • specimen

Why research data **management** (RDM) matters?

Data can be missed

"Careful examination of the first author's lab notebook then revealed missing contemporaneous entries and raw data for key experiments. The authors are therefore retracting the paper."

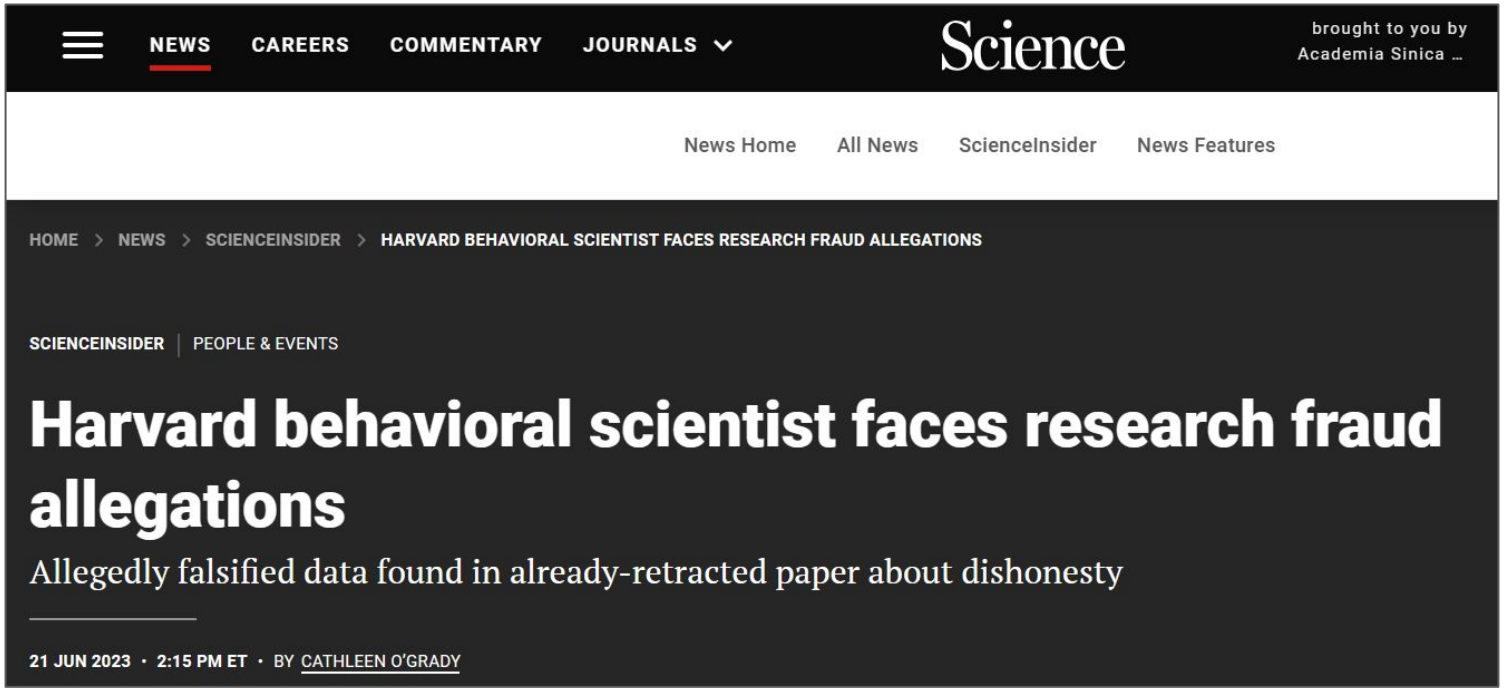
https://www.foxnews.com/science/nobel-prize-winning-scientist-retracts-paper

Inha Cho et al., Retraction. Science 367,155-155(2020).DOI:10.1126/science.aba6100

Data can get destroyed

https://securityaffairs.com/115457/breaking-news/ovh-data-centers-fire.html

Data can be manipulated or altered



(Unmanaged) data can bring large costs

- 2018 EU's survey
- at least additional cost € 10 billion for not being FAIR
- at most more than € 27 billion additional costs
- mainly because of "Time spent" and "Cost of storage"

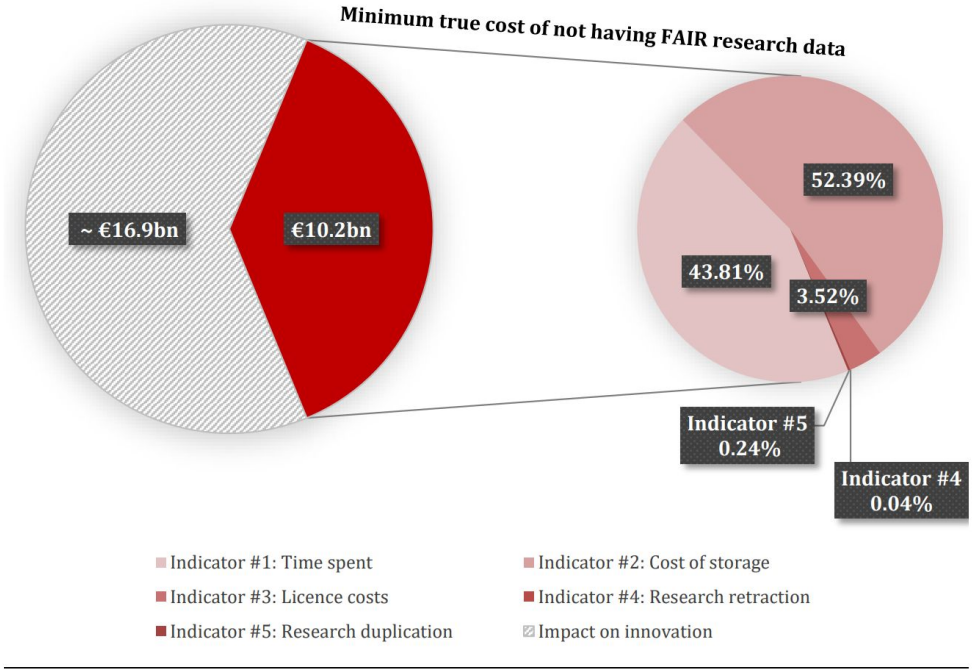


Figure 5: Cost breakdown

Pillars/Principles for good RDM

#BeFAIRandCARE

FAIR Principles

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

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 measureable 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.

FAIR Principles

Findable
可被找到

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

Accessible
可被取用

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

Interoperable
可相互操作

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

Reusable
可再次使用

R1. (meta)data have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

https://force11.org/info/the-fair-data-principles/

CARE Principle for Indigenous Data Governance

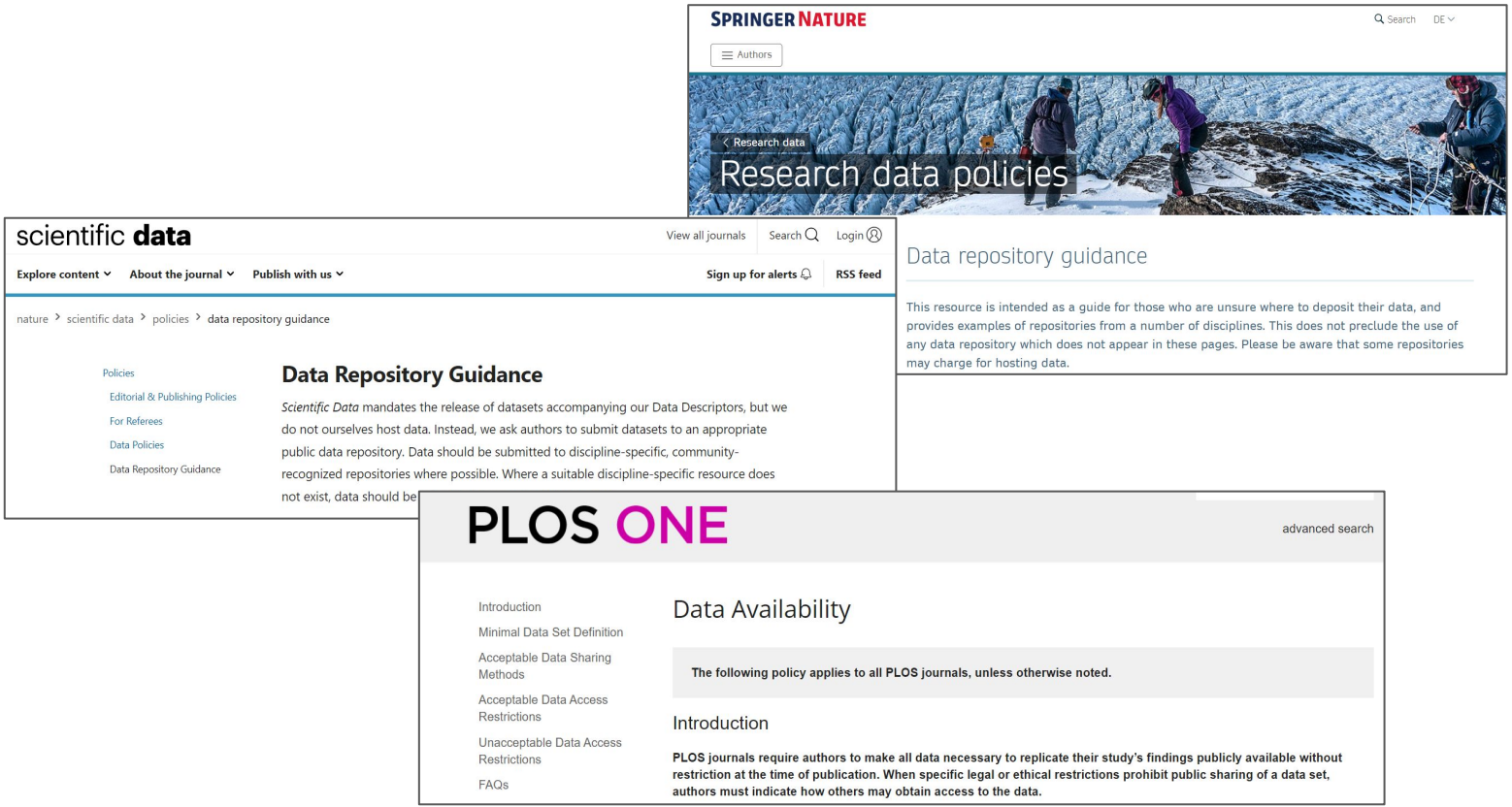
- Initiated by Global Indigenous Data Alliance (GIDA) in 2019

- “CARE” means:
  - Collective Benefit
  - Authority to Control
  - Responsibility
  - Ethics

https://www.gida-global.org/care

Trend for RDM

Policies for sharing research data

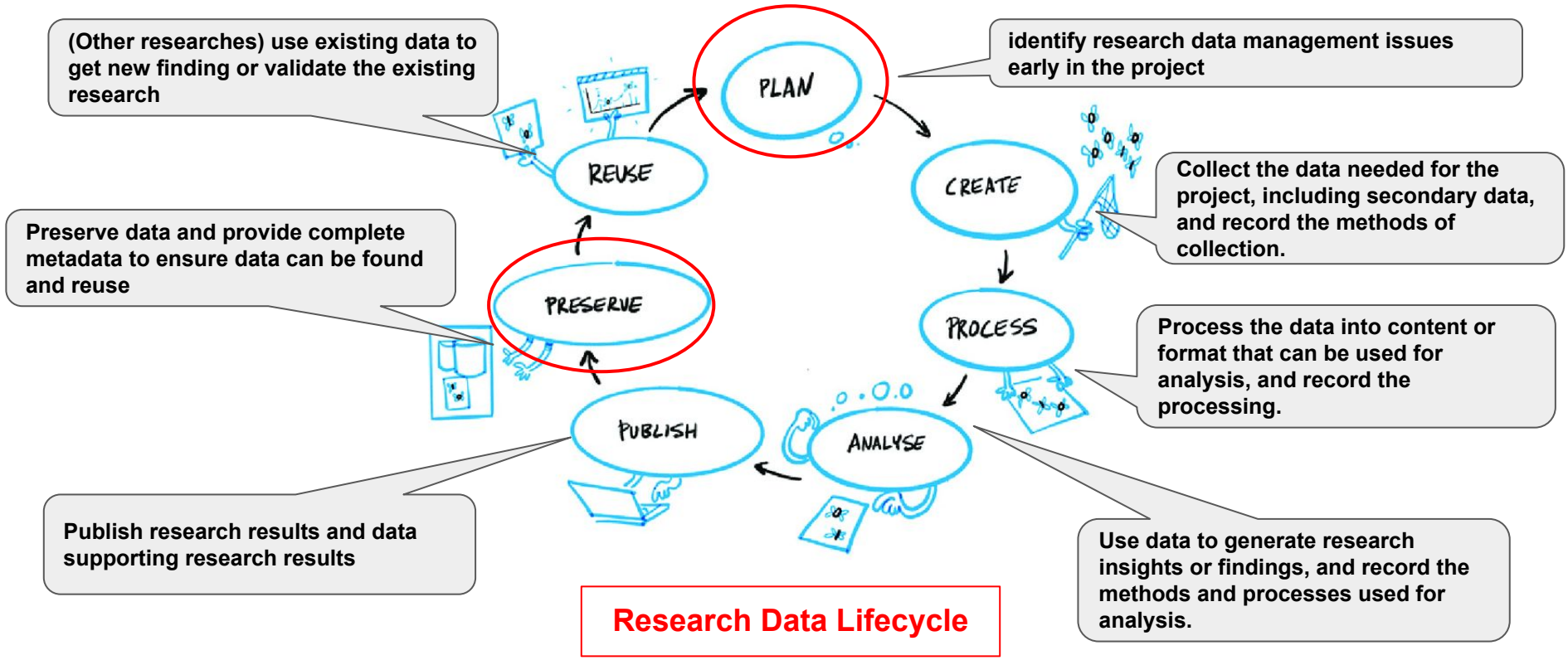


More data sharing policies

- (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

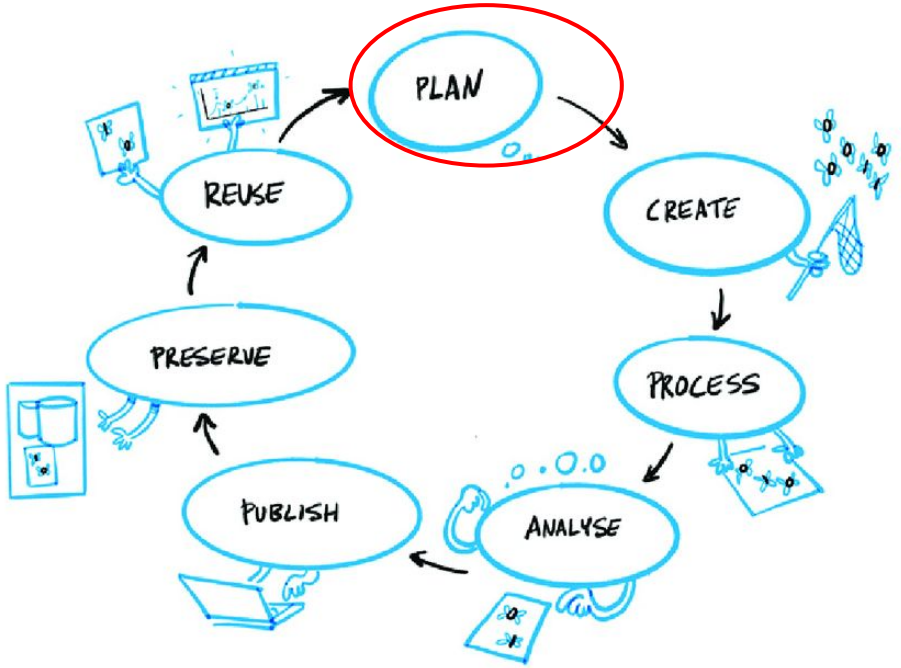
Five tips for RDM

RDM's scope



Engelhardt, Claudia. (2022). How to be FAIR with your data: A teaching and training handbook for higher education institutions. 10.17875/gup2022-1915.

Tip 1. Research funders are looking forward to “data management plan (DMP)”.



DMP is necessary for funding application

Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021

BELMONT FORUM | e-INFRASTR DATA MANA

Data and Digital Outputs Management Plan Templa

Introduction

The Belmont Forum supports multi-national and transdisciplinary collaborative rese together natural sciences, social sciences and the humanities, as well as stakeholde knowledge and solutions for sustainable development. The Belmont Forum Challeng international transdisciplinary research providing knowledge for understanding, mit adaptation to global environmental change.

To meet this challenge, the Belmont Forum emphasizes open sharing of research da 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.

Research data and digital outputs include, but are not limited to:

- Quantitative and qualitative digital information and objects created during research activities such as experiments, analyses, surveys, interviews, meas instrumentation, observations, video, audio, and computer simulations;
- All metadata describing the data and digital outputs, their acquisition (inclu description and related metadata for simulations and workflows), and othe and the reuse of the data;
- Secondary data resulting from data reduction, transformation, analyses, an with the associated code, software, workflows, and provenance informatio
- Stakeholder-oriented digital outputs such as maps (including GIS layers), de tutorials, videos, local language resources, lesson plans, curricula, policy me whitepapers; and
- Descriptions of, and metadata relating to, physical samples connected with the actual physical samples.

Each project awarded through a Collaborative Research Action is required to develo Data and Digital Outputs Management Plan to ensure ethical approaches and comp Belmont Forum Open Data Policy and Principles, as well as the FAIR Data Principles Accessible, Interoperable, and Reusable).

The Belmont Forum is in the process of gradually integrating its Open Data Policy an CRA process. This evolving process includes both the acculturation of researchers an agencies to open data practices, and the increasing movement toward transdisciplin change research. For example, the Belmont Forum recognizes that some funding ag their own data and digital output management requirements and that further speci guidance may be needed for both proposers and funding agencies to address poten However, the Belmont Forum expects that proposers will make every effort to thor

National Institute of Standards and Technology (nist.gov): NIST: National Institute of Standards and Technology Data Management Plan

Summary of activities for data generation

summary of activities that generate data

Guidance:

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.

Consider these questions:

- How will you capture or create the data?
- Are you pulling from existing data sources?

Data types and classification

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.

Guidance:

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.

Consider these questions:

- 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?

Preservation

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,

NSTC Department of Natural Science and Sustainable

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研擬計畫內容，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)」。

https://www.nstc.gov.tw/nat/ch/detail/ba70ec0e-0dfa-4a6b-8ac8-efab4bd5f520

(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：

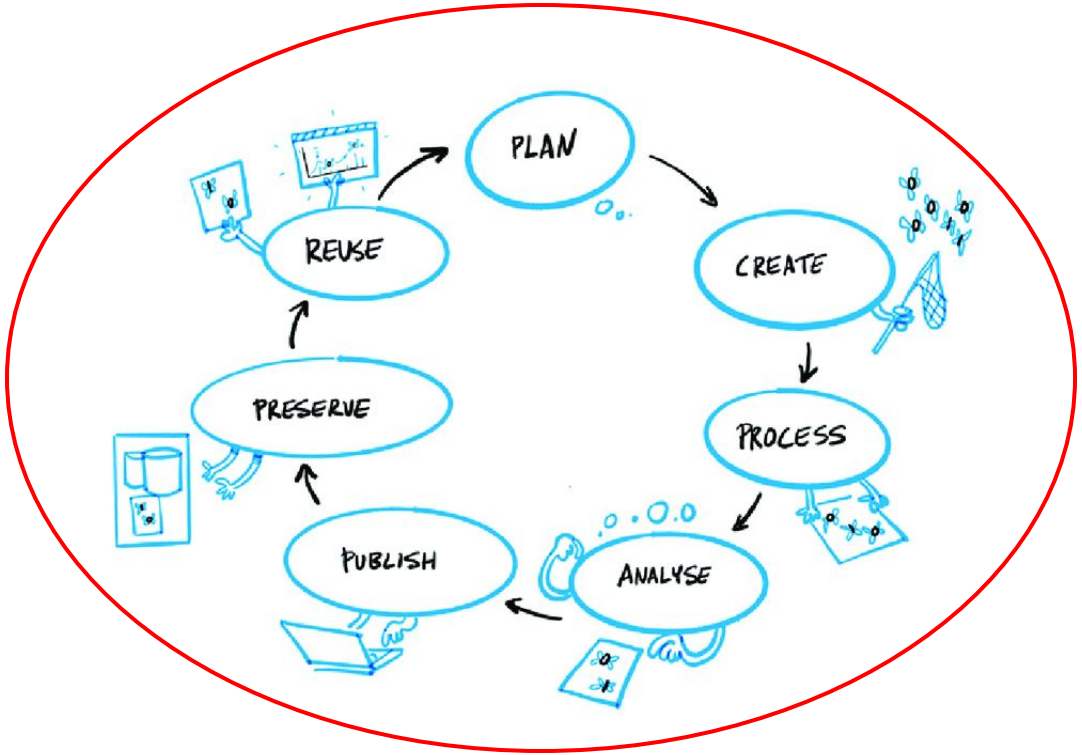
1.DMP 之規劃請以專章列於計畫書中；整合型計畫請列入「四、整合型研究計畫項目及重點說明(CM04)」、個別型計畫則列入「研究計畫內容(CM03)」。

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）。

https://www.nstc.gov.tw/nstc/attachments/38e60072-0b8b-4b8d-b08c-b184479d0e54

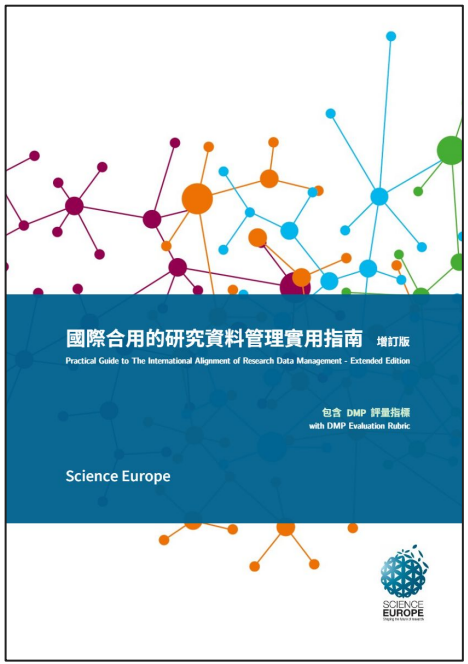
a DMP will cover…



The entire research data lifecycle

Engelhardt, Claudia. (2022). How to be FAIR with your data: A teaching and training handbook for higher education institutions. 10.17875/gup2022-1915.

The DMP template we recommended



Practical Guide to the International Alignment of Research Data Management – extended edition

國際合用的研究資料管理實用指南— 增訂版

2021-01：published by Science Europe

2021-07：translated to Mandarin by depositar team

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

Public DMPs (not in Chinese)

- **DMP Online**
  - UK, maintained by Digital Curation Center
  - https://dmponline.dcc.ac.uk/
- **DMPTool**
  - US, maintained by multiple archiving institutions
  - https://dmptool.org/
- **DMP Assistant**
  - Canada, maintained by Portage Network
  - https://assistant.portagenetwork.ca/

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

Public Chinese DMPs

2022

2023

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
指南——增訂版》(Practical Guide to The International Alignment of Research Data Management - E
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
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
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
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、
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，...

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。

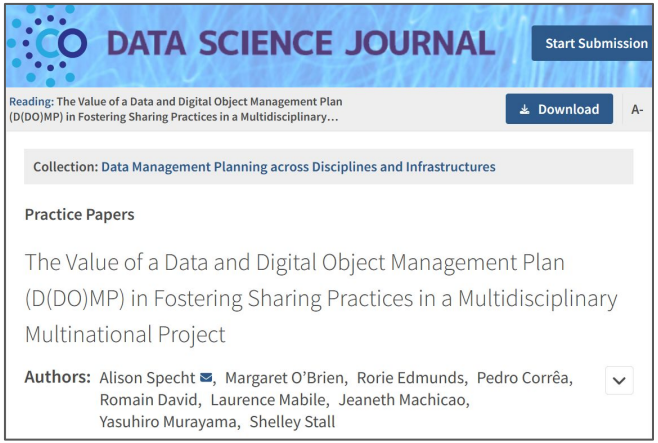
資料管理方案共計 23 份，課程講師依最後的評比、學員意願及授權，篩選出 5 份資料管理方案釋出於此資料集。期盼透過開放及分享，能啟發有研究資料管理需求的研究者有所參考依據，追求研究資料發展、分享與維護最佳實踐。

https://data.depositar.io/dataset/2022-rdm-course-dmps

What makes a good DMP?

- keeping FAIR (& CARE) in the mind
  - FAIR is a process. There are many stages between FAIR and NOT FAIR
- keeping “openness/sharing” in the mind
  - as open as possible, as closed as necessary
- “feasible” is better than “theoretically good”
  - do not promise something you can’t keep
- “concrete steps” is more important than “abstract visions/ideas”
  - make sure you know how to implement what you mentioned
- always considering multi-stakeholders in your RDM process
  - e.g., what a library or librarians can help you?
- regularly updating your DMP
  - DMP is a tool to check and improve your workflow!

Making DMP useful



https://datascience.codata.org/articles/10.5334/dsj-2023-038

(When evaluation)

" For example, it quickly became clear that the D(DO)MP is substantial (~20 pages) and unwieldy for regular use by busy researchers, so a reference summary of resources was added to the 2021 version (Stall et al. 2021b) and modified later (e.g., Stall et al. 2023b)."

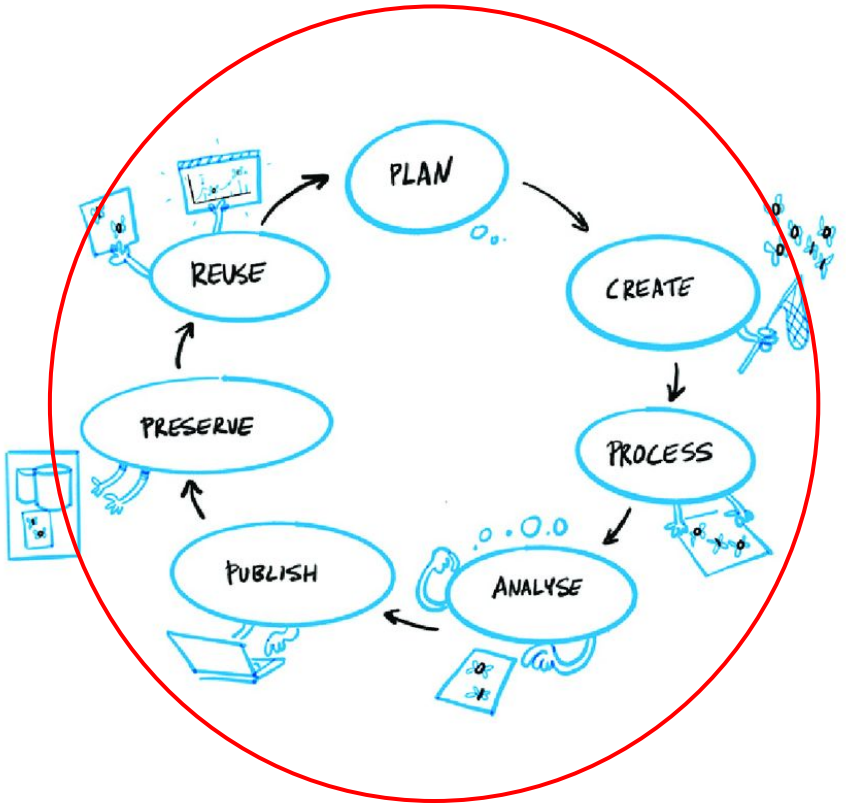
(internal impact of D(DO)MP)

"Did the existence of the D(DO)MP make a difference to the members of the PARSEC project? Establishing a code of conduct, expectations for and standards of citation early in the project, has benefitted and protected all members throughout. Having a clear and central location for communication and temporary storage has facilitated transparency, with all members of the team utilising the PARSEC Google Drive. Establishing the PARSEC Zenodo Community, the Zotero Group Library, and GitHub organisation for the project has enabled effective document and code sharing across the multinational team, and reduced duplication of effort. [...] Without the D(DO)MP to stimulate us to make these decisions and continually evaluate their utility, it is doubtful that the project would have been as organised and productive. "

(external impact)

The versions of the D(DO)MP have been well-viewed, with 510 unique downloads from our Zenodo community as of 12 August 2023 (Stall et al. 2020; 2021a; 2023a).

Tip 2. Providing as much **metadata** as possible



What is metadata?

- Data of data (*Understanding Metadata*, National Information Standards Organization(NISO), 2017)
- Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. (NISO, 2004)
    - “Structured information” means the information is completely composed of “key-value” items, such as: title, author, abstract, keywords, format, provenance, etc,.

https://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf

Why metadata matters?

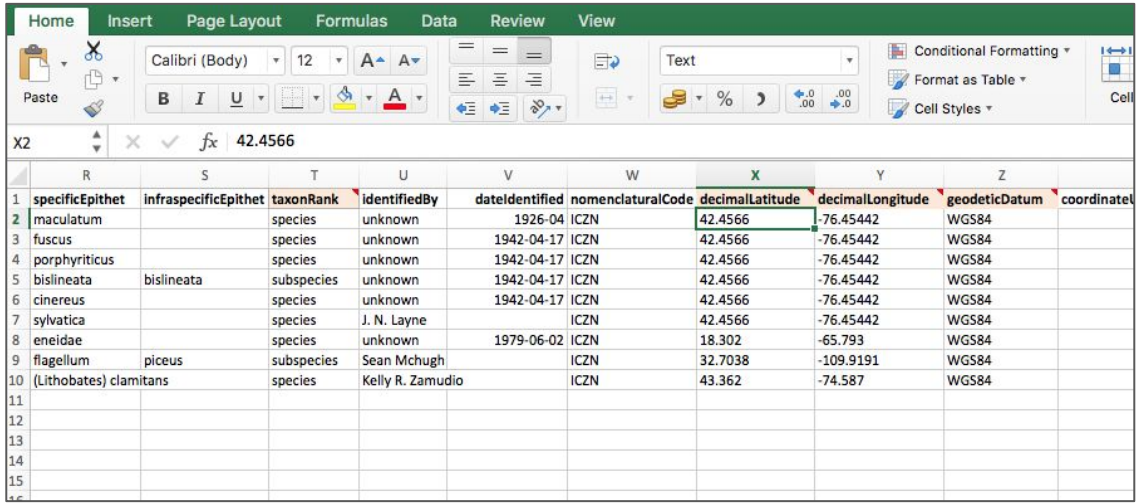
- making data easier to share, reuse, preserve
- making data easier to understand
- preventing others misusing your data
- making data more discoverable

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]

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)
- Adopting metadata standard can improve machine readability, and therefore improve data interoperability, findability.



	R	S	T	U	V	W	X	Y	Z
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	bislineata	bislineata	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									

https://www.gbif.org/news/82852/new-darwin-core-spreadsheet-templates-simplify-data-preparation-and-publishing

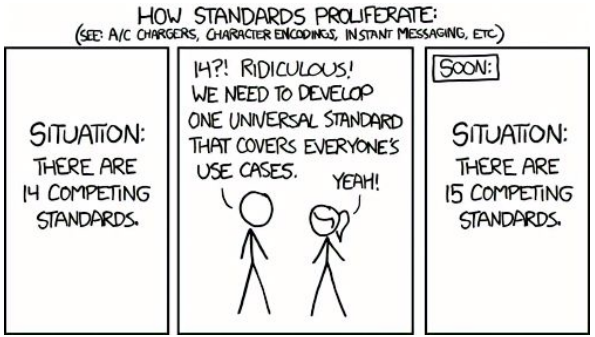
How to find appropriate metadata standards?

Different disciplines have different standards, and many disciplines currently do not have relevant standards. You can check standards catalog below: :

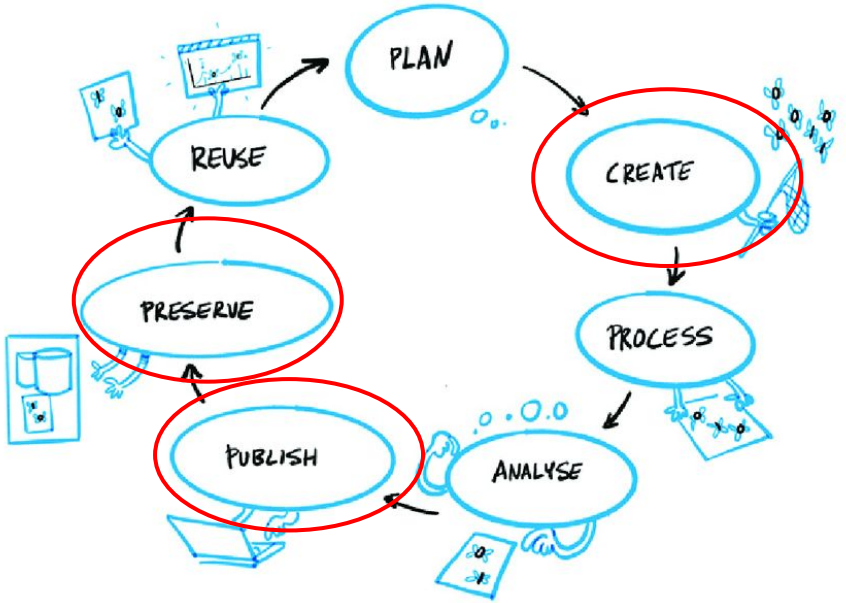
- RDA Metadata Directory : https://rd-alliance.github.io/metadata-directory/standards/
- Bath Metadata Standard Catalogue: https://rdamsc.bath.ac.uk/

Tip: Check the data or dataset standards that used by institutional or preferred data repository:

- 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



Tip 3. Ensure the data retrieved or released does not have license, legal, ethical problems.



Different types of sensitive data

- data concerning human participants: this kind of data is often regarded as “personal data”.
- data relating to species of plants or animals: include information on rare or endangered species, or other conservation activities
- commercial sensitive data: disclose this kind of data will cause economic harm or infringe others interests. E.g., trade secrets, negotiations, commercial agreement.
- data that poses a threat to others: data would pose a threat to national security or would have a negative public impact

Basic principles for handling personal data

- Purpose Specification
  （目的特定）
- Minimization（個資最小）
- Informed and Consent
  （知情同意）
- Appropriate Safeguard（適當的安全措施）

License

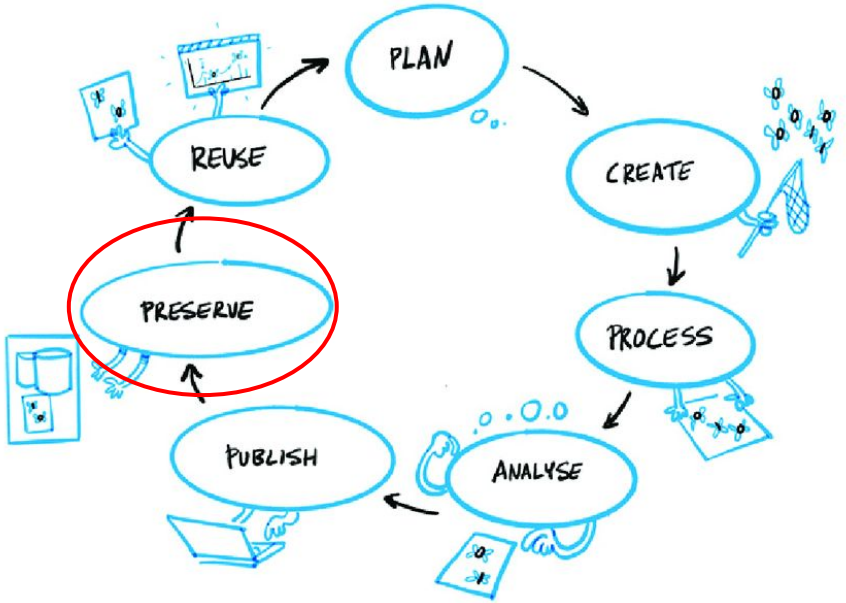
Warning: Infringement of intellectual property rights carries criminal liability in Taiwan. Ensure your data usage is legal.

Common data licnese：Creative Commons

CC0 Public Domain Dedication and Certification
(公眾領域貢獻宣告)

- Tools for selecting open license:
  - https://creativecommons.org/choose/
  - https://ufal.github.io/public-license-selector/
  - https://choosealicense.com/

Tip 4. Planning **long-term data preservation** **with an appropriate data repository** can save lots of efforts



What data should be preserved?

- Is the data related to published research or report?
- Is the data easily reproducible? Or will it cost a lot to reproduce?

- Data specified by regulations or laws.
- Unique or difficult-to-reproduce data(raw data, data analysis processes, etc.)
- Data required for research reproduction (organized data, instrument configuration files, program code, etc.)
- data with potential future reuse
- data that is beneficial to society

https://researchdata.emory.edu/preserve/preservation.html

Why use data repositories?

- Reduce the burden on researchers to respond to data reuse requests and make data accessible
- Provide assurance of data access and secure storage
- Make your data findable (e.g., make your data found in Google Dataset Search)
- Data repository can continuously make your data FAIR (Findable, Accessible, Interoperable and Reusable)
- Data repositories often care more about the sustainability of data than individual researchers do

https://www.ugent.be/en/research/datamanagement/after-research/data-repositories.htm

How to choose a data repository? (1)

- recommended or designated by journal, institution, funding agency
- disciplines preference
- general purpose data repository (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**

depositar
研究資料寄存所

GBIF
Global Biodiversity Information Facility

How to choose a data repository? (2)

data repository catalog:

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

TRUST Principles for digital repositories

Principle	Guidance for Repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

scientific **data**

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › comment › article

Comment | Open access | Published: 14 May 2020

The TRUST Principles for digital repositories

Dawei Lin ✉, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, Hervé L'Hours, Wim Hugo, Reyna Jenkyns, Varsha Khodiyar, Maryann E. Martone, Mustapha Mokrane, Vivek Navale, Jonathan Petters, Barbara Sierman, Dina V. Sokolova, Martina Stockhause & John Westbrook

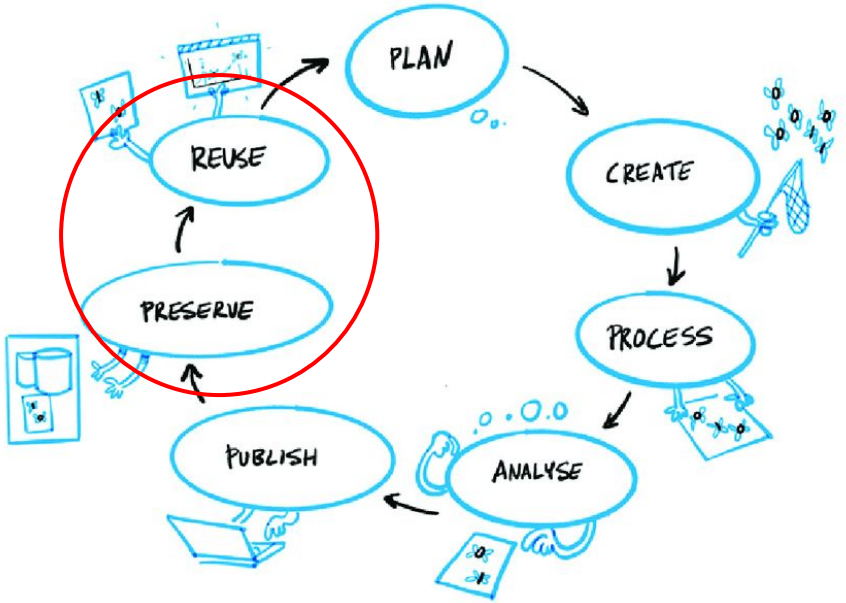
Scientific Data **7**, Article number: 144 (2020) | Cite this article

39k Accesses | **140** Citations | **153** Altmetric | Metrics

As information and communication technology has become pervasive in our society, we are increasingly dependent on both digital data and repositories that provide access to and enable the use of such resources. Repositories must earn the trust of the communities they intend to serve and demonstrate that they are reliable and capable of appropriately managing the data they hold.

https://www.nature.com/articles/s41597-020-0486-7

Tip 5. Setting an appropriate timeframe for data sharing



When to Share

Considering to share data when:

- you have data
- publishing research results
- the end of the project
- the time specified by the funder
- embargo
    - currently usually no more than 1 year, and moving towards immediate open

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。

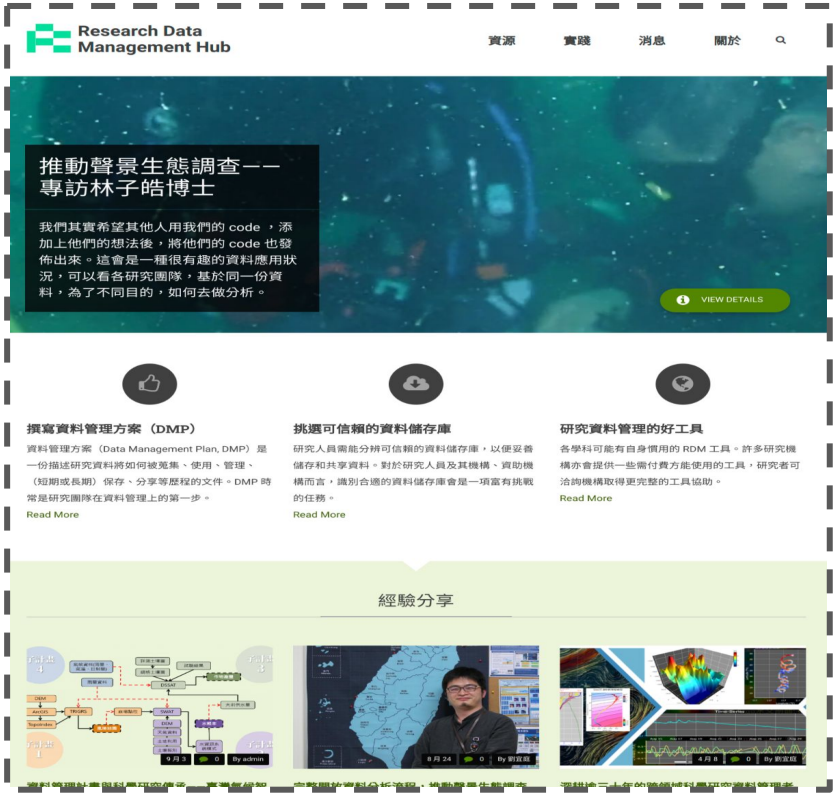
(4) 資料使用者應依學術倫理給予資料提供者適當的尊重（如列為共同作者、致謝等）。

https://www.odb.ntu.edu.tw/regulation/

Research Data Management Hub (RDM Hub)

https://rdm.depositar.io

- A space for sharing and promoting RDM practices in Taiwan
  - Practices sharing
  - Resources collection
  - Training



Contact us

for

depositar or RDM tutorial

@_depositar

謝謝！　Thank You!

depositar
研究資料寄存所

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、何明諠、李孟馨、王家薰。