

# Introduction and resources on data management

*Becky Lee*  
*HKUST Library*

# Funder Requirements

Many funding agencies have guidelines and policies for research data management **on proposal submission**.



The screenshot shows the NSF Directorate for Engineering (ENG) website. The top navigation bar includes links for HOME, FUNDING, AWARDS, DISCOVERIES, NEWS, PUBLICATIONS, STATISTICS, ABOUT NSF, and FASTLANE. The main header features the NSF logo and the text 'National Science Foundation Directorate for Engineering (ENG)'. A search bar and a 'QUICK LINKS' dropdown menu are also visible. Below the header, a secondary navigation bar lists 'ENG HOME', 'ENG FUNDING', 'ENG AWARDS', 'ENG DISCOVERIES', 'ENG NEWS', and 'ABOUT ENG'. The main content area displays a section titled 'Engineering' with an image of a circuit board, and a prominent heading for 'NSF ENG Data Management Plan Requirements'.

Proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted.



# Data Management Plan (DMP)

- ▶ DMP is a document that **describes what data will be collected or generated** and **how they will be handled** during or after the project.
- ▶ Some **common components** in a DMP include:
  - ▶ Types of data produced or collected
  - ▶ Access, sharing and re-use policies
  - ▶ Confidentiality and intellectual property
- ▶ Some software tools can help on compiling DMP
  - ▶ **DMPTool** (by University of California)
  - ▶ **DMPonline** (by UK Digital Curation Centre)



# Data Publishing

- ▶ Many **journal publishers** now require article authors to provide **well-documented data** used in the research and make these data **easily available** to other researchers.
- ▶ Examples of major publishers and their policies:
  - ▶ Elsevier - [Research Data](#)
    - a platform to support researchers to store, share, discover and use data
  - ▶ Nature - [Availability of data and material and methods](#)
    - ... authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications
  - ▶ Taylor & Francis - [Enhancing your article with supplemental material](#)
    - a number of initiatives to ensure supplemental material is effectively included, or linked to, within the article abstract.



# Data Repositories

## Social Sciences

- [Harvard Dataverse](#) - Opened to all scientific data from all disciplines worldwide. It includes the world's largest collection of social science research data.
- [ICPSR \(Inter-university Consortium for Political and Social Research\)](#) - An international consortium of more than 700 academic institutions and research organizations, and archives more than 500,000 files of research in the social sciences.

## Life Sciences

- [Genbank](#) - The NIH genetic sequence database, an annotated collection of all publicly available DNA sequences.
- [Protein Databank](#) - A worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids.

## Earth Sciences

- [DataOne](#) - Comprises a distributed network of data centers, science networks or organizations, with a focus on Earth observational data.

# Data Citation

- ▶ Citing data brings the **same benefits** as citing articles and books because it
  - ▶ helps data creators and users **find, access, verify and reuse** data
  - ▶ allows **data impact** to be tracked
  - ▶ gives proper **recognition** to data producers
- ▶ Examples:

Creator (Publication Year): Title. Publisher. Identifier

Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo.

<http://dx.doi.org/10.1594/PANGAEA.726855>

Creator, Publication Year, "Title", Identifier, Publisher, Version, [Universal Numerical Fingerprint]

Campbell, Cameron Dougall; Lee, James; Dong, Hao, 2015, "Longitudinal Links to Construct the Korean Multi-Generational Panel Dataset – Tansung (KMGPD-TS) from the Tansung Household Registers", <http://dx.doi.org/10.14711/dataset/IVIDZV>, DataSpace@HKUST, V1 [UNF:6:dSzr9Zfpl4QhmtlZRtXBRg==]

# Data Discovery and sharing

- ▶ Adding DOIs to finalized data sets
  - ▶ Makes data publications more acceptable for CVs
  - ▶ Creates more recognition for data sharing
  - ▶ Provides a means of tracking publications
- ▶ Creating your own ORCID iD for publishing
  - ▶ ORCID iD is a unique researcher identifier for distinguishing your data/article publications from those of others with similar names
  - ▶ You can establish clear ownership of your scholarly output
  - ▶ 124 HKUST staff and 349 students have already registered their ORCID iDs
  - ▶ You can use your HKUST login to easily create yours in a few clicks via <http://repository.ust.hk/orcid/>



# Information on data management

<http://library.ust.hk/sc/>

The image shows a screenshot of the Scholarly Communications @ HKUST Library website. The header features a logo with the letters 'SC' inside a colorful, overlapping geometric shape, followed by the text 'Scholarly Communications @ HKUST Library'. Below the header is a navigation menu with items: Home, Author Tips, Data Management, Open Access, Research Repositories, Seminars, and About Us. The 'Data Management' menu is expanded, showing a list of links: Why Research Data Management, Data Management Plans, Data Repositories & Discovery Tools, Standards, Guidelines, Best Practices & Training, Data Citation, DataSpace@HKUST, and Funder Requirements & Publisher Policies on Data. Four red callout boxes with white text point to these links: 'DMP' points to 'Data Management Plans', 'Data repositories' points to 'Data Repositories & Discovery Tools', 'Data Citation' points to 'Data Citation', and 'Funder requirement & publisher policies' points to 'Funder Requirements & Publisher Policies on Data'. The main content area includes a book cover for 'Sino-US Energy Triangles' and a news snippet about Springer Nature's ORCID milestone.

**DMP**

**Data repositories**

**Data Citation**

**Funder requirement & publisher policies**



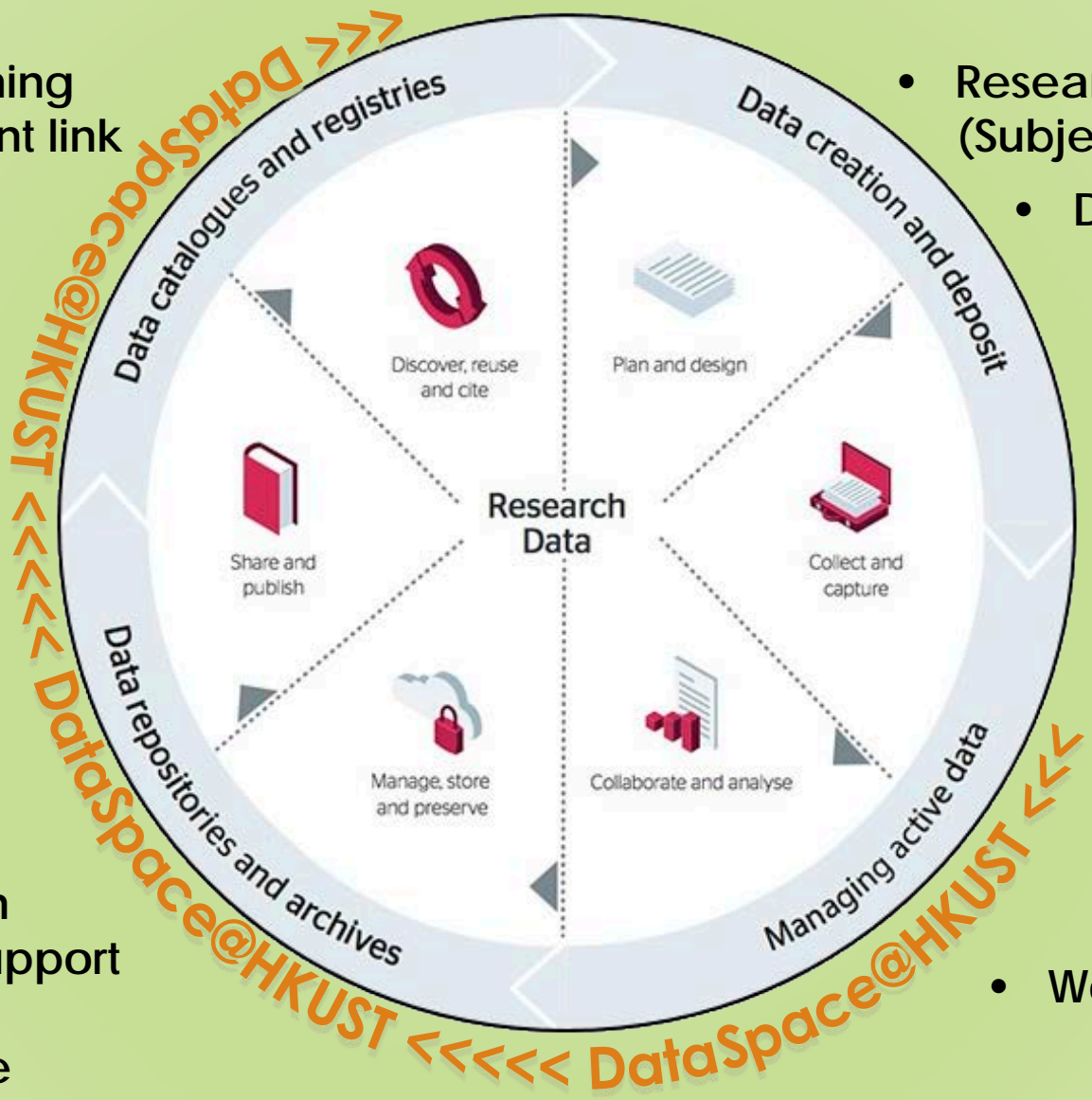
# How Library services support data management

- Data publishing
- DOI persistent link
- ORCID iD

- Research consultation (Subject librarians)
- DMP

- Data citation
- Metadata support service
- Data storage

- Workspace





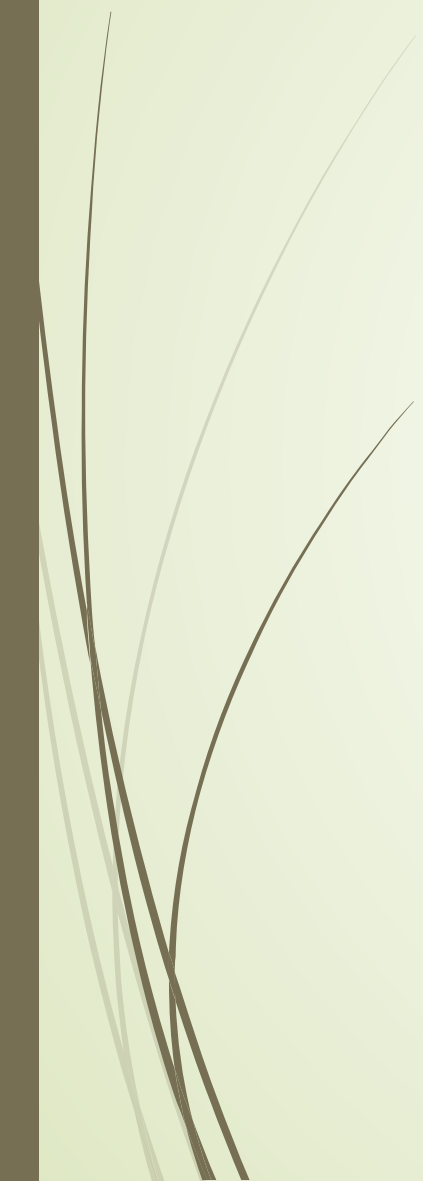
Contact our Subject Librarians  
at

<http://library.ust.hk/serv/contact.html>



# Demonstration of DataSpace@HKUST

*Mr. K.T. LAM*  
*Associate University Librarian*  
*Head, Systems & Digital Services*  
*Chief designer of DataSpace@HKUST*





**Q & A**