NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Services and Challenges in Managing the Institutional Research Data Repository

Dr Amy CHOU
Ms Yuyun W. ISHAK

NTU Library
Office of Information, Knowledge & Library Services

In collaboration with:

COAR Confederation of Open Access Repositories

COAR-Asia OA

22 Jun 2021

COAR Confederation of Open Access Repositories

https://www.coar-repositories.org/asia-oa/

ANNUAL MEETINGS

October 25-27, 2021

Asia OA Annual Meeting – Virtual

More details coming soon!

Asia OA 2020 – Seoul, South Korea

October 19th, 2020 | 0 Comments

Asia OA 2019 Meeting in Dhaka, Bangladesh

Asia OA Summit – Positioning Asian in the Global Movement of Open Science

December 19th, 2017 | 0 Comments

RECENT WEBINARS

- 27 Apr 2021: Balancing good practices with inclusivity: COAR Community Framework for Good Practices in Repositories
- 18 Dec 2020: Building Library Capabilities for Research Data Management Services
- 10 Dec 2020: Open Access Policy development process in India

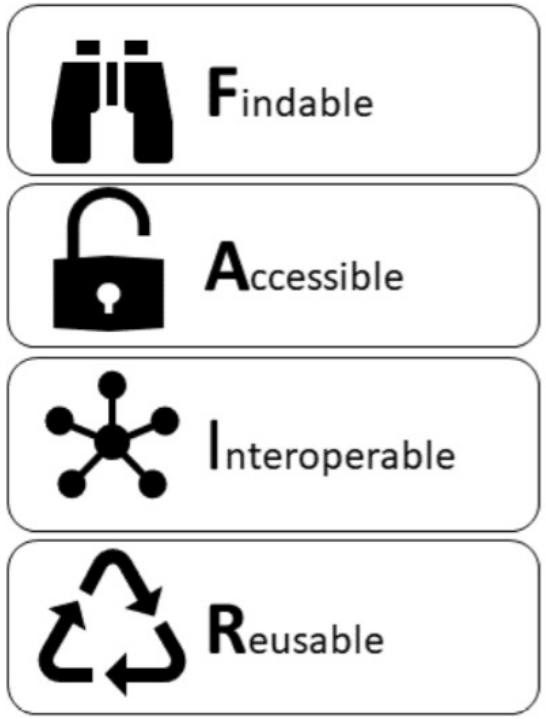
Agenda

- What value-added services can repository managers provide to help ensure that research data are FAIR (Findable, Accessible, Interoperable and Reusable)?
- What are some of the common challenges faced by data repository managers?

Goals of institutional research data repositories

- Showcase the research data outputs of the associated entities or institutions
- Make research data available for sharing and reuse for designated communities

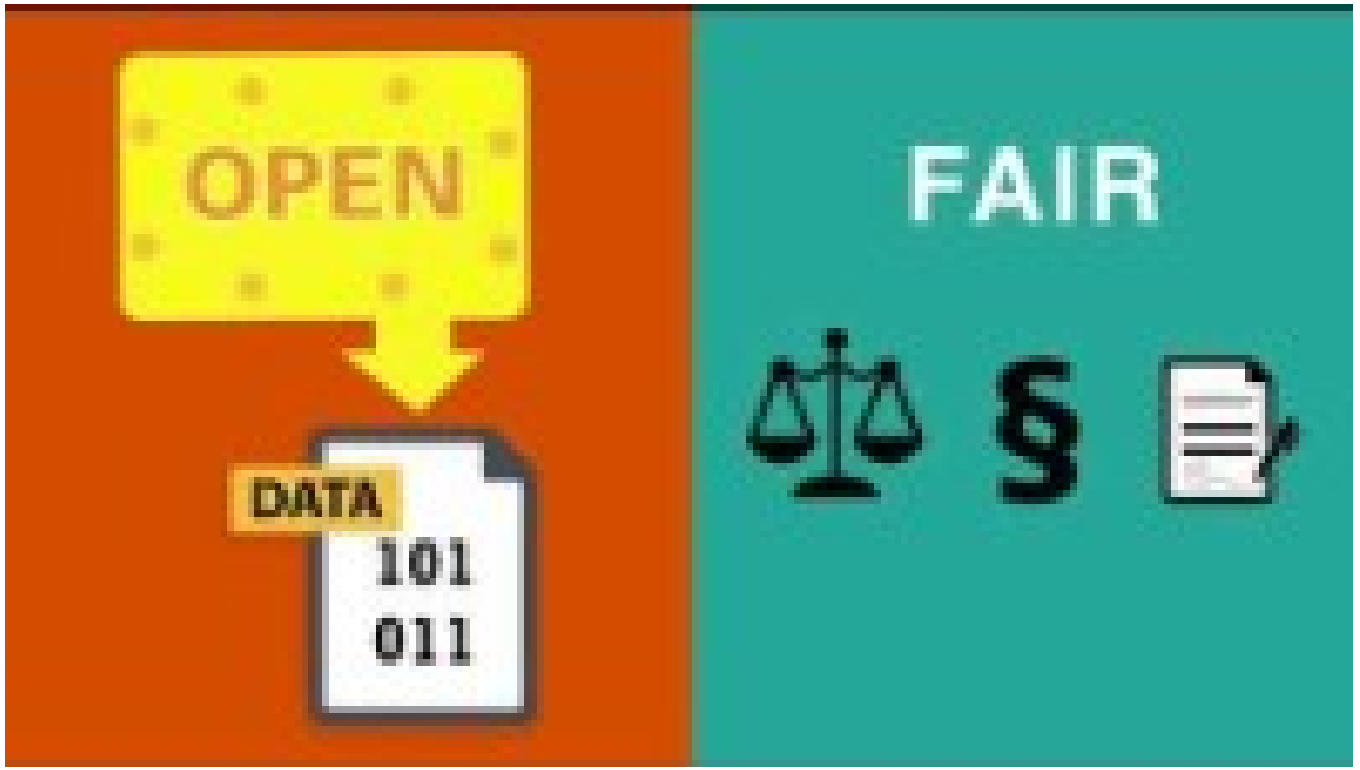
FAIR principles help repositories to achieve their goals



- By both humans and machines
- Via authentication or authorisation where necessary
- Can be integrated with other data or with applications or workflows (for analysis, storage and processing)
- Metadata and data be well-described so that they can be replicated and/or combined in different settings

Source: https://www.go-fair.org/fair-principles/

What are FAIR principles



Source: https://www.youtube.com/watch?v=5OeCrQE3HhE

Why FAIR matters

- There is so much data available nowadays that machines are needed to discover and process at scale.
- The quality of the metadata and using standards is an important aspect in facilitating this.

Repository services to achieve FAIR data

- Repository platform features
- Data curation and user education

FAIR data and repository services

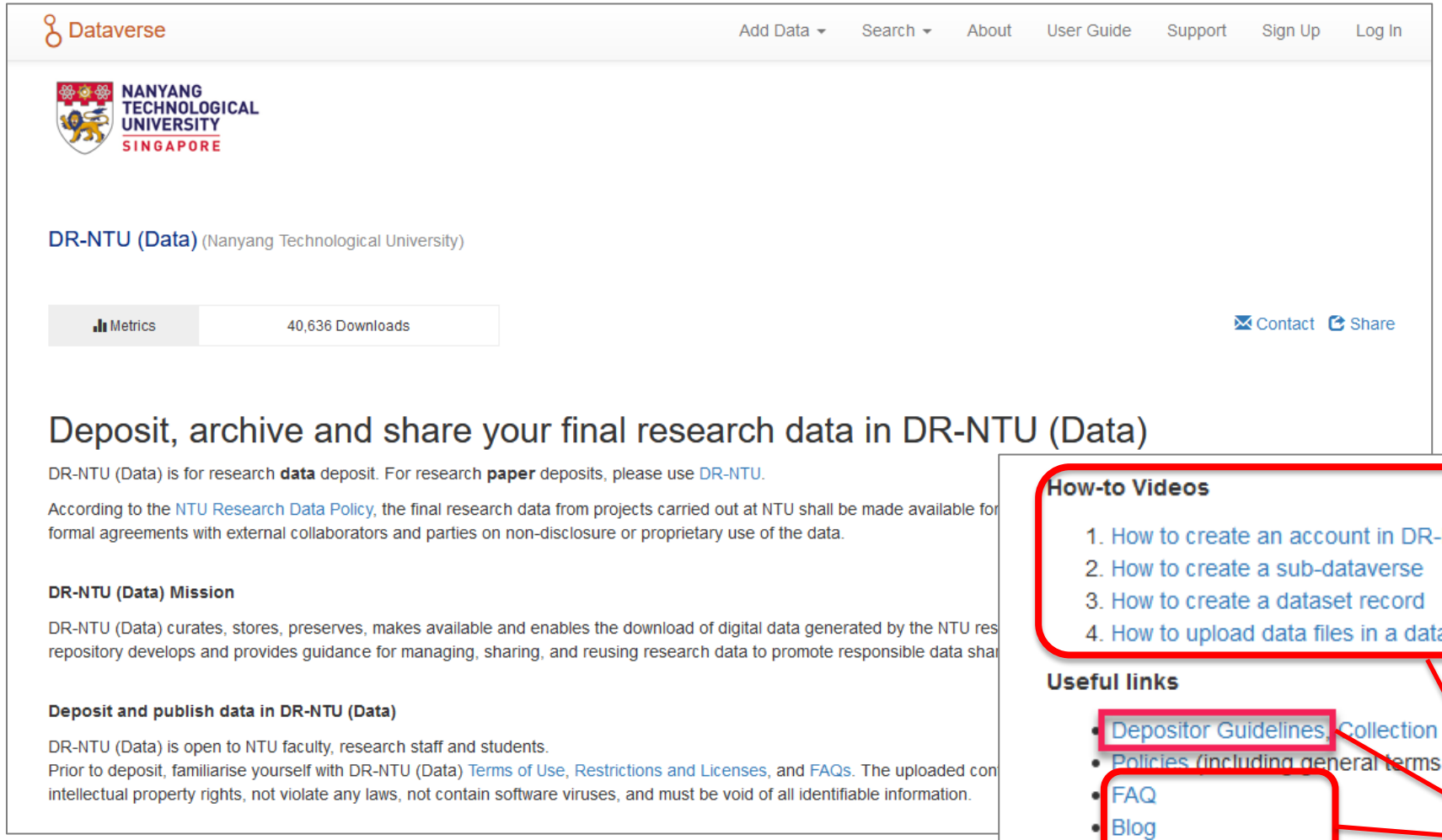
- **Findable**: Unique and persistent identifiers (e.g. DOI), rich metadata, indexed metadata, searchable platform
- **Accessible**: Metadata are retrievable by their identifier by open, authentication protocol, metadata are accessible even when data are no longer available (file restriction, deaccession)
- **Interoperable**: Metadata is accessible/applicable or uses controlled vocabulary, includes qualified references to other metadata
- **Reusable**: Rich metadata (e.g. data usage license), provenance (e.g. context, workflow), meet domain-relevant (subject-specific) community standards, open file formats, file versioning

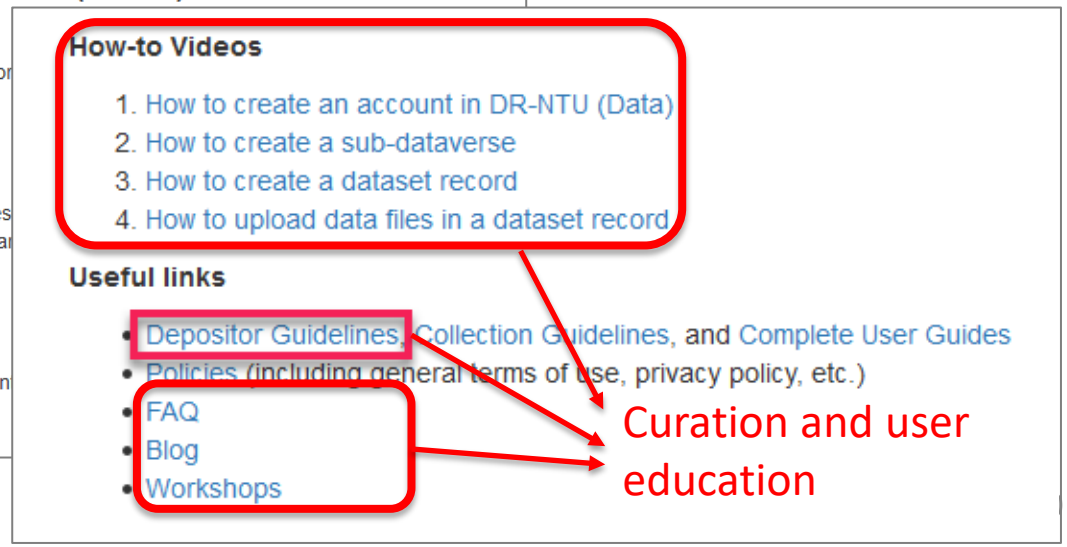
Repository platform

Curation and user education

Source: https://www.go-fair.org/fair-principles/; https://www.youtube.com/watch?v=DutWdCYZ45I

DR-NTU (Data) repository services





- Open-source software, Dataverse, developed at Harvard's Institute for Quantitative Social Science (IQSS)

Source: https://researchdata.ntu.edu.sg/

DR-NTU (Data) User Guides and Policies

Search this Guide | Search

- Home
- User Guides
  - Institutional Log In
  - Create Dataverse and Add Dataset
  - Edit Dataverse, Dataset, File Management
  - Supported Metadata and References
  - Finding and Using Data
  - Data Citation
  - Data Exploration Guide
  - Tabular Data File Ingest
- Policies
  - General Terms of Use
  - Privacy Policy
  - Community Norms
  - Data Usage Agreement (Sample)
  - DR-NTU (Data) API Terms of Use
- Collection Guidelines

General Information

What to deposit

What to deposit: (For more details, see Collection Guidelines)

- **Final, empirical data** (e.g. tabular files, text files, images, scripts) underlying your research carried out at NTU and **related data documentation**.

What not to deposit: (For more details, see Collection Guidelines)

- Sensitive research data*
- Data that might affect patent application
- Research data that has been deposited in other open data repositories
- Journal papers, conference proceedings, reports, or manuscripts (For these, deposit at DR-NTU)

What NOT to deposit

** If your data contains **sensitive, identifiable information**, please check the Sensitive data LibGuide for data sharing best practices. You are also strongly recommended to **notify** the NTU Data Librarians **prior to** depositing your anonymised dataset.*

What to do if there's sensitive information

When to deposit:

- Upon publication of the article or 12 months after project end date, whichever is earlier.

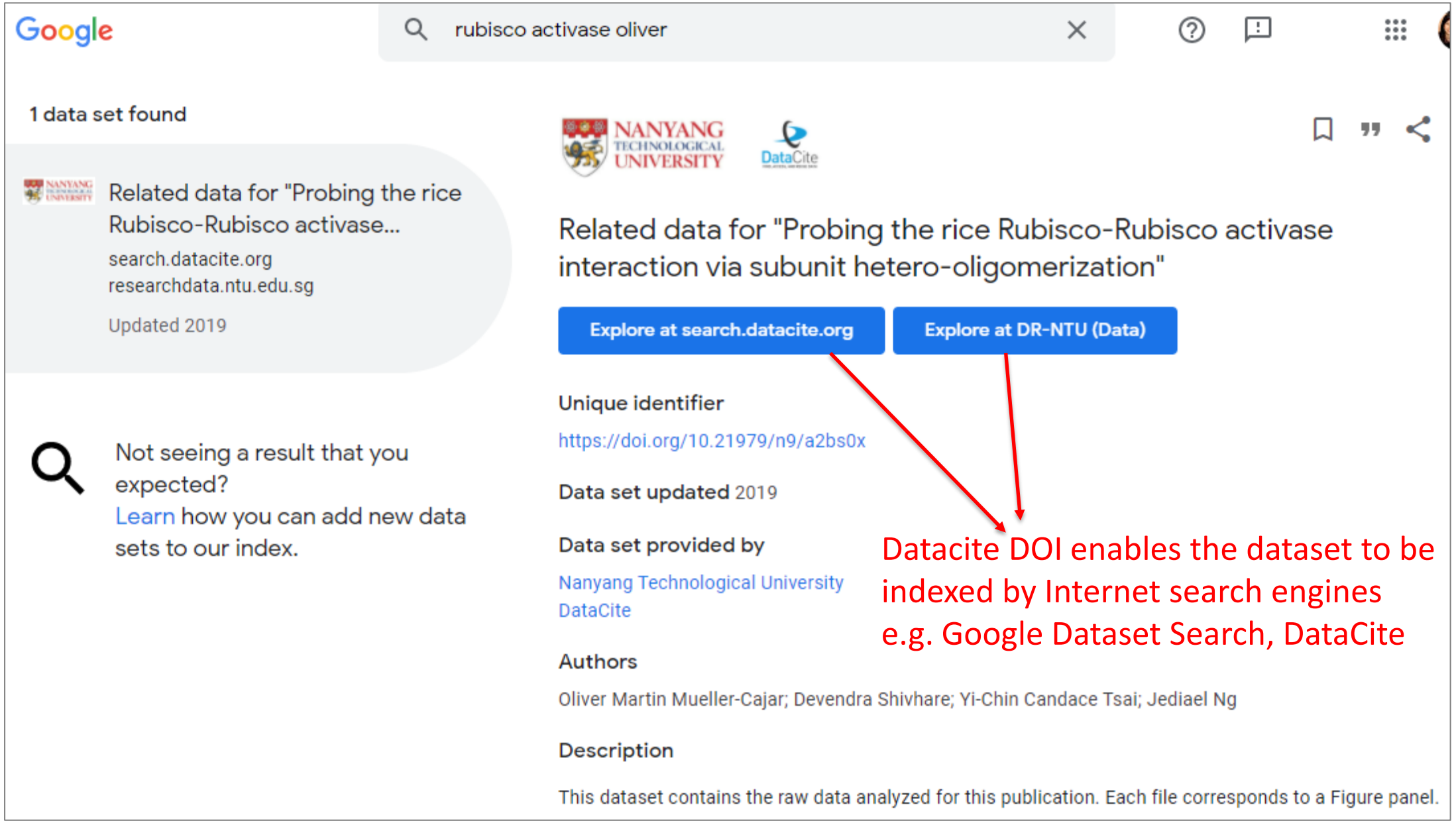
Prepare your data for deposit

How to prepare data for deposit

Prepare your **data files**:

- **File naming**:
  - Use meaningful file names. Avoid spaces, dots and special charaters. (See LibFAQ). For example:
    - [Project]-[Data type]-[Date in YYYY-MM-DDThh:mm:ss]_[Version number]
    - [Figure/Table number]-[Description]_[Version number]
- **File formats**:
  - DR-NTU (Data) accepts data in all formats.
  - However, we recommend that you save or convert your data into **recommended file formats** suited for long-term access and reuse. (For more

DR-NTU (Data) User Guides and Policies. https://libguides.ntu.edu.sg/drntudataguidespolicies/depositor



https://datasetsearch.research.google.com/search?query=rubisco%20activase%20oliver&docid=L2cvMTFqOWI2MTRtcw%3D%3D

DR-NTU (Data): Dataset example 1

Oliver Martin Mueller-Cajar (Nanyang Technological University)

DR-NTU (Data) > School of Biological Sciences (SBS) > Oliver Martin Mueller-Cajar >

Related data for "Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase"

Version 1.0

Mueller-Cajar, Oliver Martin; Tsai, Yi-Chin Candace; Ye, Fuzhou; Liew, Lynette; Di Liu; Bhushan, Shashi; Gao, Yong-Gui, 2019, "Related data for "Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase"" https://doi.org/10.21979/N9/K4IROM, DR-NTU (Data), V1, UNF:6:Jbn6MDwmQifQ9UkD9ghwKw== [fileUNF]

Cite Dataset Learn about Data Citation Standards.

Access Dataset

Edit Dataset

Link Dataset

Contact Owner | Share

Dataset Metrics

340 Downloads

Description	This dataset contains the biochemical data collected for this publication. One Excel file is provided for each Figure panel. The electron micrographs used for the 3D-reconstruction shown in Fig. 2 are also provided (*.mrc files).
Subject	Agricultural Sciences; Chemistry; Earth and Environmental Sciences; Medicine, Health and Life Sciences
Keyword	Rubisco activase, carbon fixation, AAA+ proteins
Related Publication	Tsai, Y. C. C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y. G., & Mueller-Cajar, O. (2020). Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. Proceedings of the National Academy of Sciences, 117(1), 381-387. doi: 10.1073/pnas.1911123117

Unique, persistent identifier assigned upon dataset creation

Metadata

Keywords, indexing terms

Source: https://doi.org/10.21979/N9/K4IROM

Subject	Agricultural Sciences; Chemistry; Earth and Environmental Sciences; Medicine, Health and Life Sciences
Keyword	Rubisco activase carbon fixation AAA+ proteins
Related Publication	Tsai, Y. C. C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y. G., & Mueller-Cajar, O. (2020). Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. Proceedings of the National Academy of Sciences, 117(1), 381-387. doi: 10.1073/pnas.1911123117 https://www.pnas.org/content/117/1/381
Grant Information	Nanyang Technological University: startup grant Ministry of Education (MOE): Tier 2 grants MOE2016-T2-2-088 Ministry of Education (MOE): Tier 2 grants MOE2015-T2-1-078 Ministry of Education (MOE): Tier 2 grants MOE2017-T2-2-089
Depositor	Mueller-Cajar, Oliver Martin
Deposit Date	2019-12-11
Kind of Data	Excel spreadsheets; *.mrc format files
Software	Microsoft Excel, Version: Office 365 Relion, Version: 3.0
Related Material	https://www.pnas.org/content/suppl/2019/12/17/1911123117.DCSupplemental
Related Datasets	The atomic coordinates and structure factors have been deposited in the Protein Data Bank, https://www.wwpdb.org/ (PDB ID code 6L1Q).; The electron microscopy (EM) density map of AfQ2O2 has been deposited in the Electron Microscopy Data Bank, https://www.ebi.ac.uk/pdbe/emdb/ (accession no. EMD-0789).

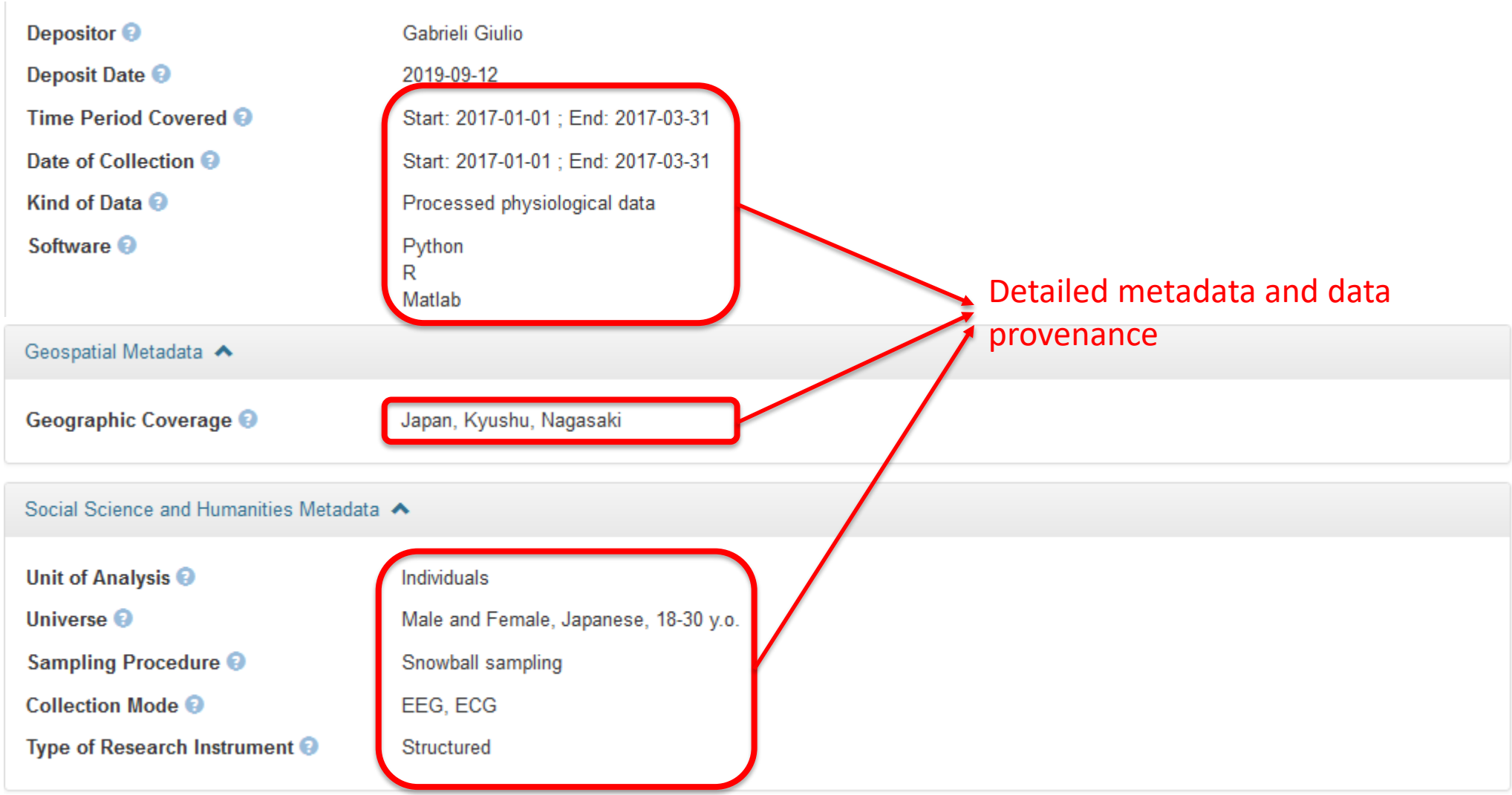
Industry accepted file formats and software

Qualified references (i.e. related material and datasets)

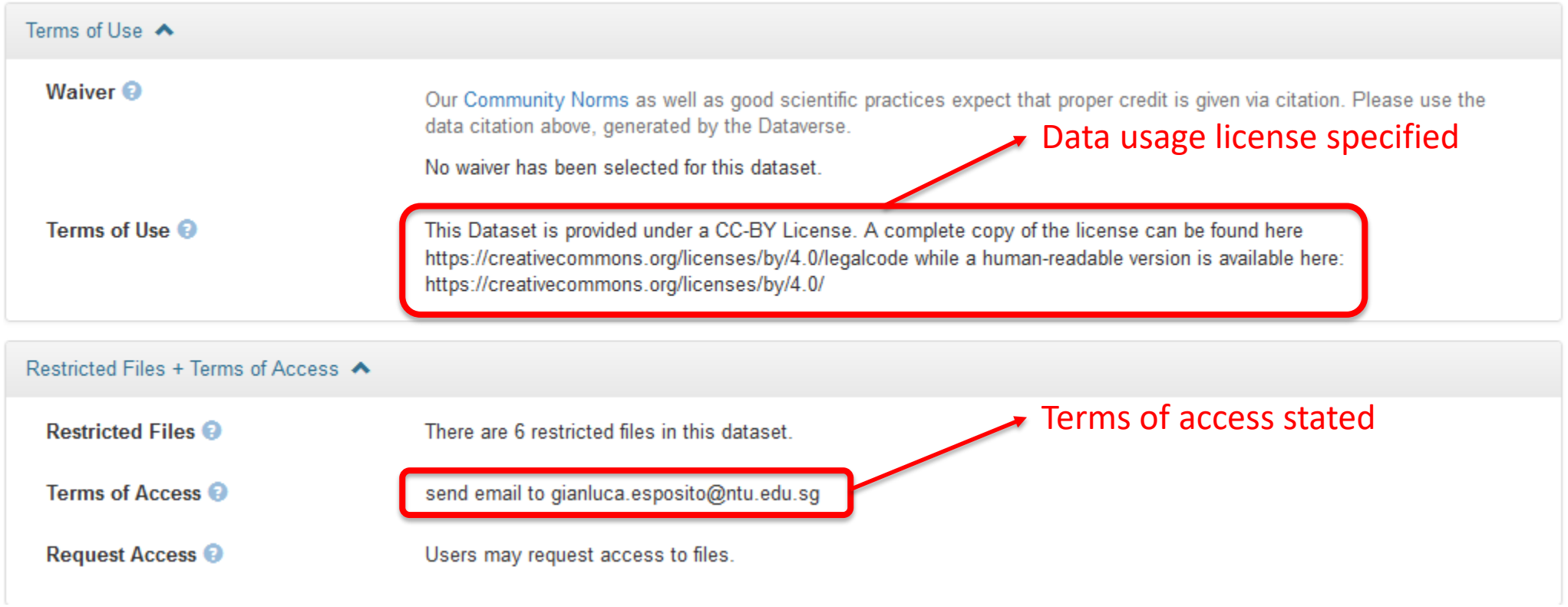
Source: https://doi.org/10.21979/N9/K4IROM

DR-NTU (Data): Dataset example 2

Source: https://doi.org/10.21979/N9/GTRLJJ



Source: https://doi.org/10.21979/N9/GTRLJJ



Source: https://doi.org/10.21979/N9/GTRLJJ

Data curation & user education

- User education via advisory and hands-on workshops
  - Choice of repository to supplement data deposits
  - Provision of metadata
  - Data file preparation (e.g. open file formats where applicable)
- Quarterly newsletter
- Yearly outreach/webinars during Open Access Week

Resources for FAIR data

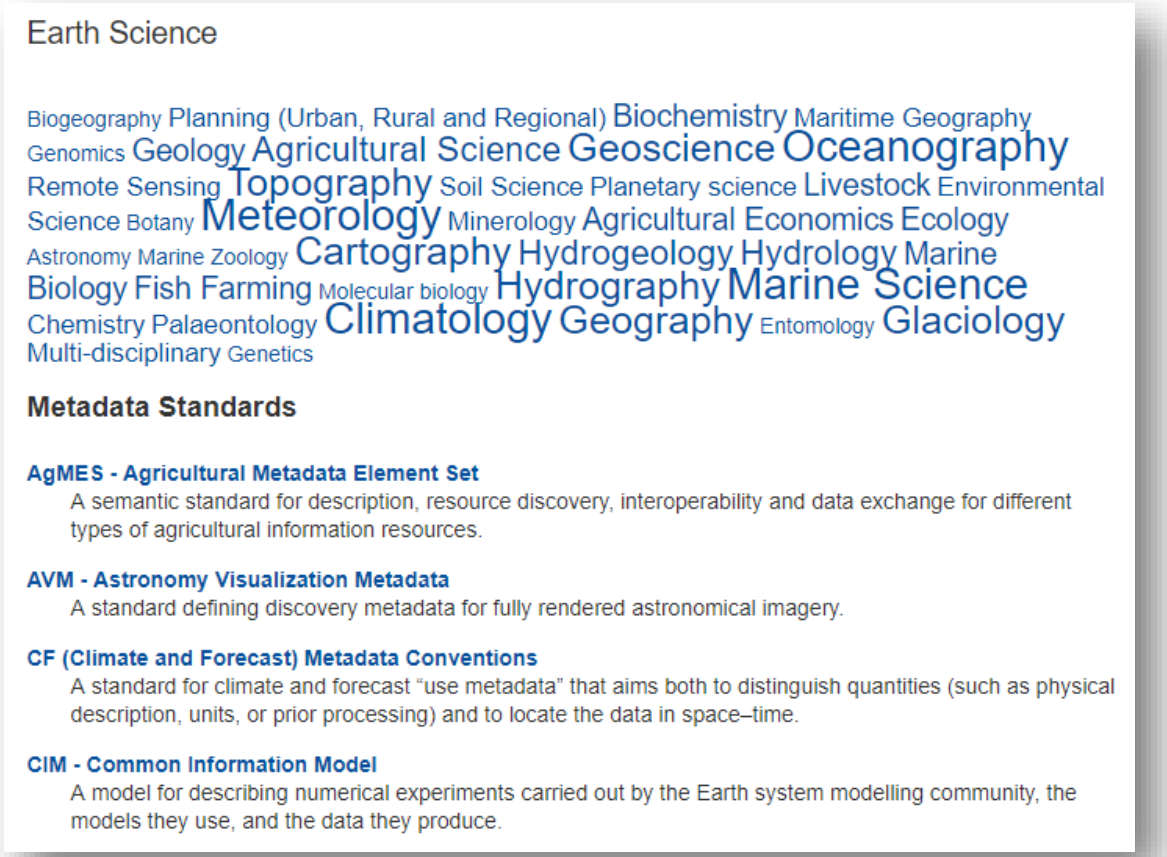
- Find FAIR Data tools:
  - FAIRifier and Metadata Editor (creating)
  - FAIR Data Point (publishing)
  - FAIR Search Engine (searching)
  - ORKA (annotation)
- FAIR data assessment tools:
  - ARDC FAIR Data self-assessment tool (before deposit)
  - F-UJI Automated FAIR data assessment tool (after deposit)
- FAIRsFAIR support programme for data repositories

Resources on metadata standards

- Several domain specific metadata schema have been established to describe data sets.
- Three examples of metadata schema are:
  - Dublin Core: a metadata schema aimed at resource discovery in **general terms**
  - The Data Documentation Initiative (DDI): a standard used in the **social sciences** to document survey and other observational data
  - The Encoded Archival Description (EAD): a standard for encoding descriptive information regarding **archival records**

Resources on metadata standards

- The Digital Curation Centre (DCC) offers a catalogue of disciplinary metadata standards.
- FAIRsharing.org provides a repository of disciplinary and data management metadata standards across the globe.



Resources on PIDs

Services that supply globally unique and persistent identifiers

- Identifiers.org provides resolvable identifiers in the form of URIs and CURIEs: http://identifiers.org
- Universally unique identifier: https://en.wikipedia.org/wiki/Universally_unique_identifier
- Persistent URLs: http://www.purlz.org
- Digital Object Identifier: http://www.doi.org
- Archival Resource Key: https://escholarship.org/uc/item/9p9863nc
- Research Resource Identifiers: https://scicrunch.org/resources
- Identifiers for funding organisations: https://www.crossref.org/services/funder-registry/
- Identifiers for the world's research organisations: https://www.grid.ac

Common challenges faced by data repository managers

Depositor

- Lack of incentive
- Not finding it important or necessary
- Not a priority
- Too tedious
- Lack of awareness
- Reluctant to share
- What will happen to my data?

Data

- Sensitive data
- Data that does not belong to you
- Data with commercialization potential (e.g. patent filing)
- Big data
- License for reuse
- Reusability
- Reproducibility

Curator/Data Manager

- Skill gap
- No manpower
- No funding
- No policy/mandate from the top
- No research data repository
- No requirement from local funders

Challenges and approaches

- Lack of incentive
- Not finding it important or necessary
- Not a priority
- Too tedious
- Lack of awareness
- Reluctant to share
- What will happen to my data?

Depositor

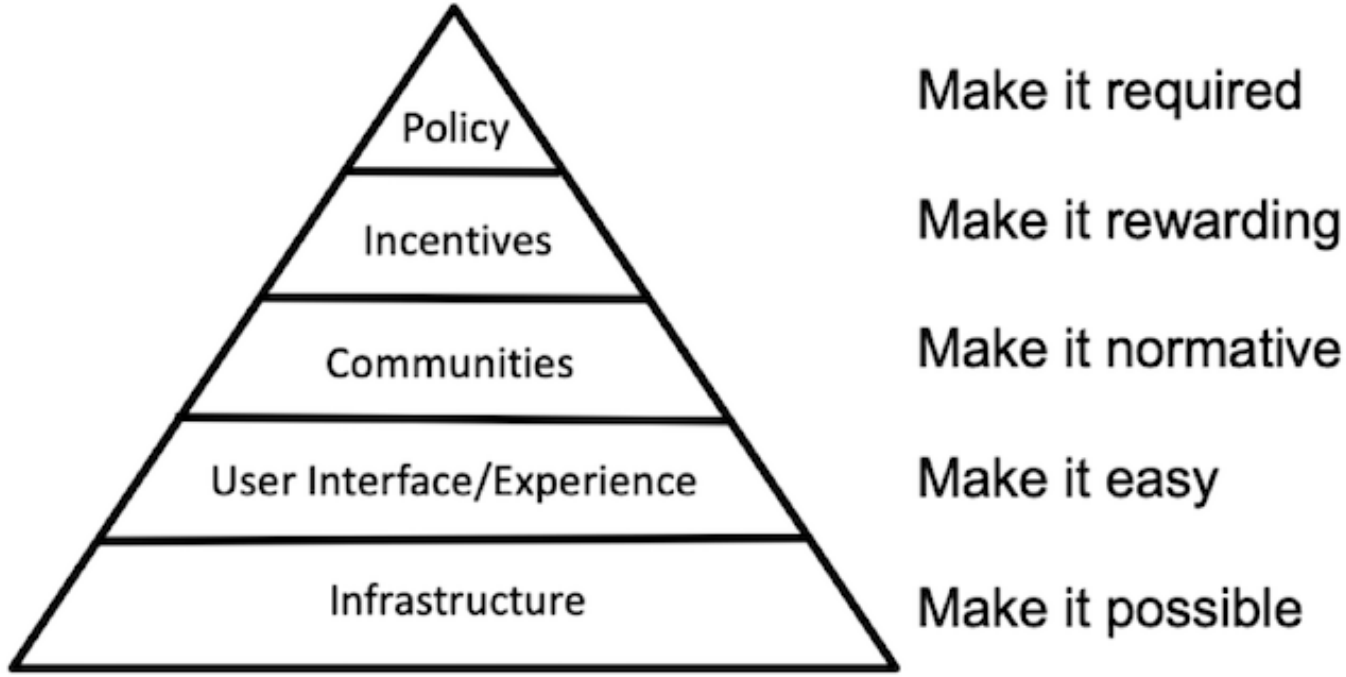
Approaches:

1. **Institutional** Policy/Mandate (e.g. NTU Research Data Policy) which stipulates researchers to deposit research data in repository.
2. Data sharing mandate from local **funders** (e.g. National Medical Research Council).
3. Requirement from **publisher/journal** with data sharing policies (e.g. Taylor and Francis, Wiley, Springer Nature).
4. Advocacy and outreach (e.g. online guides, workshops, elearning course, community of practice).
5. Repository certification (e.g. CoreTrustSeal).

Challenges and approaches

- Lack of incentive
- Not finding it important or necessary
- Not a priority
- Too tedious
- Lack of awareness
- Reluctant to share
- What will happen to my data?

Depositor



Source: Nosek, B. (2019). Strategy for Culture Change. https://www.cos.io/blog/strategy-for-culture-change

Challenges and approaches

- Sensitive data
- Data that does not belong to you
- Data with commercialization potential (e.g. patent filing)
- Big data
- License for reuse
- Reusability
- Reproducibility

Data

Approaches:

1. Closed/Restricted repository.
2. Embargo on datasets or data files.
3. Data citation metrics which may indicate data reusability.
4. Data curation (mediation).
5. Advocacy and outreach (e.g. depositor guidelines, workshops, elearning course, community of practice).

Challenges and approaches

- Skill gap
- No manpower
- No funding
- No policy/mandate from the top
- No research data repository
- No requirement from local funders

Curator/Data Manager

How would you overcome these challenges?

Add your idea in 1-3 words

Get management buy-in!

National repository

there are different levels of curation...

Training for library staff interes ted to learn about RDM

Cultural change

enhance collaboration

just do it :)

self learning

promote re3data.org, Zenodo

learn from pioneer like NTU :)

No research data repository - prom

ca recommend third-party repository to researchers

NEXT WEBINAR:

Title: New Roles and Capabilities of Academic Libraries in the Evolving Open Access Landscape

Date: 3 Aug 2021 (Tue)

TIME:

Beijing / Hong Kong / Kuala Lumpur / Singapore / Taipei
11:00 am – 12:30 pm

New Delhi
8:30 am – 10:00 am

Tokyo / Seoul
12:00 pm – 1:30 pm

Bangkok / Jakarta
10:00 am – 11:30 am

REGISTER AT:

http://bit.ly/sharingAug2021

The Zoom link will be emailed upon registration.

Thank you!