CHAPTER 3

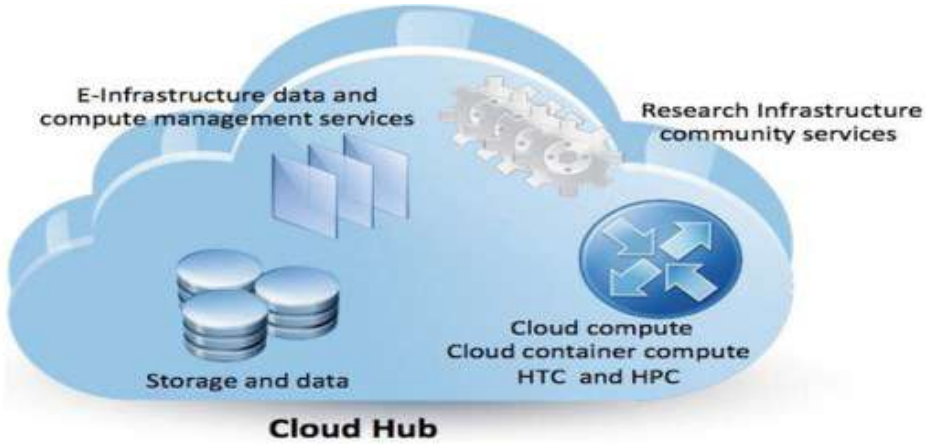
INFRASTRUCTURE ON OPEN SCIENCE IN PUBLIC FUNDED RESEARCH

3.1. BACKGROUND

Open Science can be implemented through online collaborative platforms that will link geographically-dispersed researchers or other stakeholders or interested parties. These platforms will enable them to cooperate seamlessly on their research, sharing research objects as well and ideas and experiences. Generally, in the form of online services, collaborative platforms provide a virtual environment that concurrently links to multiple people and allows them to work on the same task. These can range from extensive virtual research environments (VREs) that facilitate sharing and collaboration via web forums and wikis and, collaborative document hosting. The infrastructure of the European Open Science Cloud (EOSC) can be seen in Box 3.1. The infrastructure comprises research data, processing services, virtual laboratories and tools, that relies on a federated system of data and storage facilities.

In Malaysia, it is common for higher learning institutes to own a platform to deposit and share scholarly publications such as journal papers and students' theses. Based on the Landscape on Open Science in Malaysia, there are a few institutions that has advanced to develop their own research data repository such as Universiti Putra Malaysia (UPM) who has established an in-built research data repository, which at the moment is still at a project-based level, but will transition to the institutional level in the near future. Universiti Teknologi Malaysia (UTM) has set up the prototype and has identified potential users for a pilot test. Universiti Malaya (UM) also developed a prototype for a research data repository.

BOX 3.1. INFRASTRUCTURE ON OPEN SCIENCE



Source: ISSI Scientific Report Series Vol.15

[9] Virtual research environments have been defined as "innovative, dynamic, and ubiquitous research supporting environments where scattered scientists can seamlessly access data, software, and processing resources managed by diverse systems in separate administration domains through their browser" (Candela, Castelli and Pagano, 2013).

3.2. GUIDELINES FOR DEVELOPMENT OF INSTITUTIONAL REPOSITORY

When it comes to open data sharing through Open Science in research, researchers fear that their data will be misused for unethical purposes or they might get scooped or misinterpreted. Another concern relates to commercial entities who make use of freely available data with no strings attached. Therefore, a **robust policy is a must to secure and establish a trustworthy platform**. It is important to identify if an institution has established its own data sharing policy. For example, the University of Malaya is developing the UM research data management policy and Universiti Putra Malaysia (UPM) has also drafted a data sharing policy and is now waiting for endorsement from the top management.

In addition, an Open Science platform must have **a process that controls and oversees data usage** (e.g. who uses the data, for what purpose the data is used, and approvals from relevant authorities to access the data), and must ensure that data sharing practices **meet the FAIR (Findable, Accessible, Interoperable, Reusable) principle** and all data contributors and data users must be clear on their **responsibilities** and understand **ethical rules** when using the platform. Hence, it is of importance that fundamental technical elements are embedded in the platform to secure the deposited data.

An ideal feature of a trusted data sharing platform is having a **good data request handling process** and that the re-used data must be properly cited and acknowledge by the data originator. The cost and infrastructure setup involved should be planned to ensure **sufficient storage capacity** to deposit raw research data. The future expansion for the increasing capacity of stored data should be considered.

3.3. INFRASTRUCTURE DEVELOPMENT

3.3.1. Types of storage options or solutions

There are various storage solutions available but these solutions must be compared against two criteria:

a. The value of the data and its potential for reuse.

b. The types of components which give value to data, such as its discoverability, curation and whether the storage is reliable, large and sustainable.

Researchers incline towards using individual or project data storage (e.g. USB, hard drive on individual laptop, local drives etc.) as it is a simple, convenient and quick solution to store data. However, this option reduces the potential for data reuse, as well as the discoverability, reliability and sustainability of the data, thus reducing the value of the data itself. Institutions should encourage researchers to share their raw research data on institutional repositories to ensure that the stored data is reliable, well-curated and identifiable with appropriate metadata. The repository should be able to support the submission of raw research data at any stages of the research cycle, either the initial data, working data or final data stages. Researchers should plan at the start of a project how they will store data, and to outline its budget. Such planning must be documented in a data management plan, as described in Section 5.4.1 of Chapter 5.

[10] https://www.ands.org.au/guides/data-storage

3.3.2. Interaction between storage solutions and with metadata stores

An example of Malaysia Open Science Platform (MOSP) as indicated in **Figure 3.1**. enables researchers to easily store, discover, access and share their data for better research impacts. Storage of raw research data are held in institutional repositories and are made discoverable to users using the MOSP Portal. The MOSP portal is the central getaway to Malaysia's research data. The MOSP Portal forms a registry or a catalogue that harvests metadata from multiple institutional, agency-based and domain-specific repositories. The harvested metadata are archived and stored in the MOSP Central Portal. Using interoperable metadata standards and Application Programming Interface (API), the MOSP Portal can be integrated with multiple types of repositories. This includes institutional repositories and domain-specific repositories. The portal will be an extendable and flexible platform for data sharing. Integration of the MOSP Portal with existing repository systems must be complemented with applicable risk management mechanisms to identify potential risks associated with integration and interoperability issues during implementation.

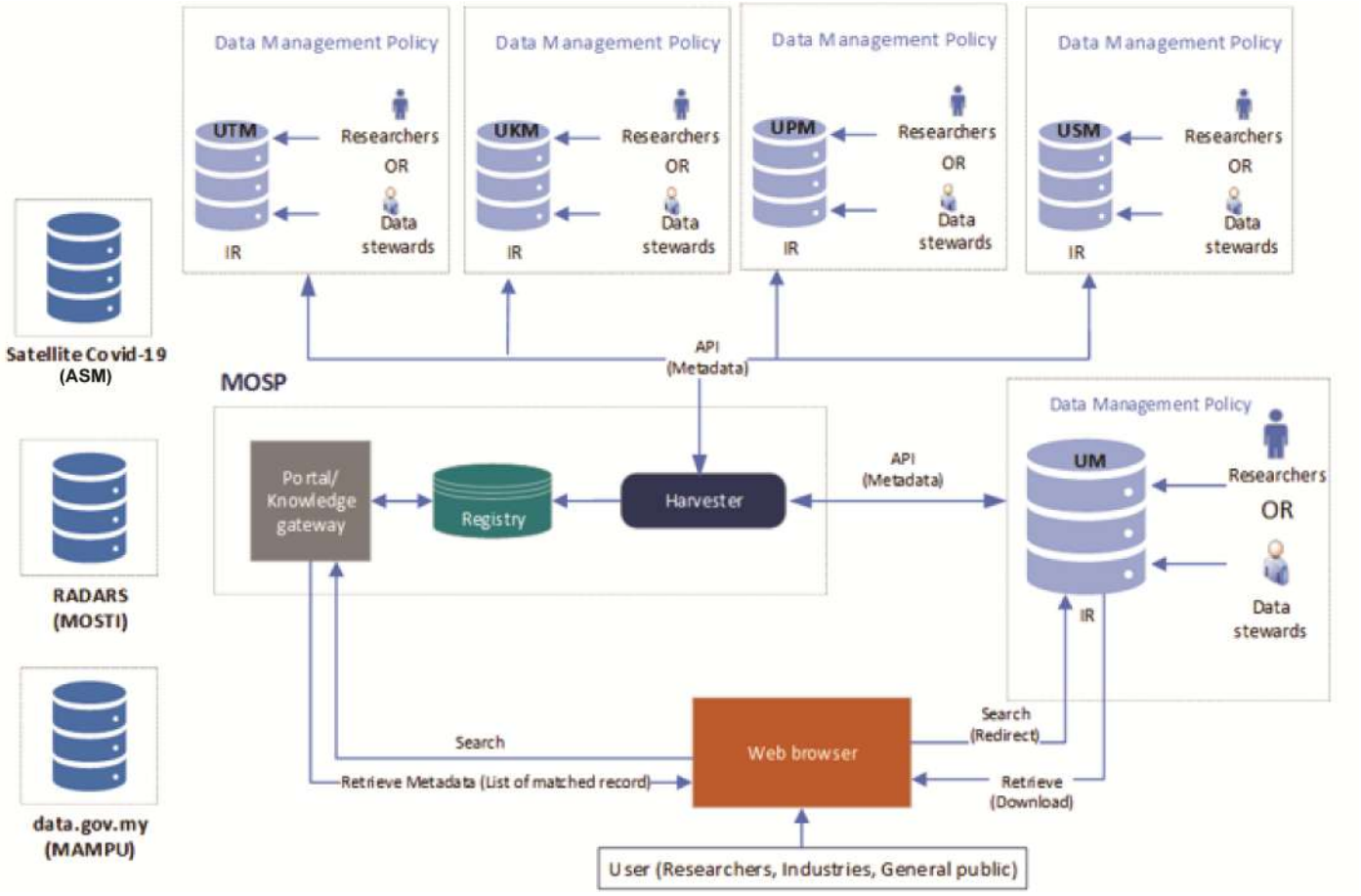


Figure 3.1: Architecture of the Malaysia Open Science Platform (MOSP) pilot project
Source: MOSP, 2020

3.3.3. Identifier[11]

A persistent identifier is any label that is used uniquely to guarantee that the deposited datasets can be managed and kept up to date over a defined time. An identifier will be assigned to each entry of research datasets by respective repositories. The IT team is responsible for keeping the system running while the data originator is responsible for providing up-to-date information about the entry that is being identified. The identifier serves as one of the metadata elements and will be captured by the MOSP Portal for each entry registered in the MOSP Portal.

3.3.4. Publishing and sharing sensitive data

Sensitive data is data that must be protected against unwanted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations.

Malaysia has strong regulations regarding personal (for example, Personal Data Protection Act 2010) and non-personal data. Examples of sensitive data are:

1. Personal data - Name, photographs, Identification Card (IC), bank details, medical records, bank details.

2. Confidential data - Physical or mental health or condition of a data subject, his political opinions, his religious beliefs, interview transcripts containing identifiable individuals' sensitive personal data such as drug dependence, research data/information/IP with significant commercial value/obligations.

3. Biological data – endangered or threatened species whose survival depends upon protection of their habitat location.

When handling and dealing with sensitive data, it is important that careful measures must be undertaken when collecting, processing, handling, and storing data throughout the research process. As such, appropriate permits, and informed consent must be sought before initiating the research process. Anonymisation of personal data should be taken into account to ensure that these data are non-identifiable when being deposited. The sensitivity of datasets must be identified and appropriate ways of handling these data must be written in a Data Management Plan.

[11] https://www.ands.org.au/guides/persistent-identifiers-expert
[12] https://www.openaire.eu/sensitive-data-guide and https://www.ands.org.au/guides/sensitivedata

CHAPTER 4

INCENTIVES AND MANDATORY RULES ON OPEN SCIENCE IN PUBLIC FUNDED RESEARCH

4.1. INCENTIVES

Some examples of incentives[13] to be adopted are:

a. **Policy incentives** – Appropriate national and institutional policies that detail out infrastructure, incentives for training of researchers and data stewards and recognising of support systems such as Research data management policy, people to develop DMP & Protection of Intellectual property).

b. **Infrastructure incentives** – From ICT provision, software, research funding and equipment.

c. **Personal incentives** – For the researchers. From getting flexibility in terms of working to career progression and recognition, to the extent that data sharing is being incorporated as a Key Performance Index (KPI).

d. **Funding Incentives** - Ensuring funds to cover open access publication costs and costs related to data management, stewardship and long-term preservation. In addition, allocating funds for Open Science activities such as (but not limited to) citizen science projects, mentoring/ training and awareness-raising activities, prizes to individual researchers for being a role model in practising Open Science in addition to those related to Open Access to scientific publications and research data. Introducing openness as a criterion in selection procedures for the awarding of grants (not limited to open access to publications and data, but also taking into consideration elements like contributing in open peer review processes, participating in citizen science projects).

e. **Social incentives** – Various infographics, leaflets, information and also videos to promote Open Science. Incentives for reimbursement are also provided (E.g.: If you buy resources or tools for your research to facilitate data sharing, the amount can be reimbursed).

4.1.1. Effective communication of incentives on data sharing

The incentives provided in Open Science must be adequately communicated in a manner that resonates well with Malaysian researchers. One way of doing it is by highlighting that Open Science could bring socioeconomic benefits from research works. As for now, incentives of data sharing are not communicated effectively. Those responsible in addressing the national incentives for MOSP, MOSTI, MOHE, or ASM would be the key entities to lead the initiative. Successful stories featuring researchers who have fostered industrial collaboration or with international players are great examples to convince the quadruple helix as a whole. Lastly, imposing Open Science practices as KPI for researchers/institutions can be considered as an incentive and it is important to have a proper policy implemented since researchers need to be given assurance on processes and protocol layers in Open Science, with data stewards will play an important role to assist researchers in managing the whole data lifecycle process.

[13] that an African Open Science Platform has adopted.

4.2. MANDATORY RULES

4.2.1. Compliance

The research entity/institution will take the grant holder's compliance with the Guidelines into account when assessing research performance and when future applications for funding are received from the grant holder. Reporting on compliance will be required both during and at the end of the funding periods for projects receiving support. In case of no compliance, the research entity/institution retains the right to reduce the grant amount at the payment of balance or afterwards.

4.2.2. Acknowledgement in all Publications

Recipients must acknowledge in all publications the name of the research entity /institution and identify the funding source using the project name, and/or acronym, and/or number in the standardise prescribed manner [provide the standardised acknowledgement here, or refer to the appropriate document/webpage where this is defined, e.g. Guidelines for Grant Applicants].

4.2.3. Provision of persistent address

Open Access for research outputs is demonstrated by providing a persistent address where the digital object can be accessed, read, downloaded.

4.2.4. Policy Review

An evidenced-based review of the policy implementation will take place [3 years] following its adoption and subsequent reviews will take place on biennial basis. After that, the policy will be reviewed and updated every [3 years].

CHAPTER 5

GOVERNANCE ON OPEN SCIENCE IN PUBLIC FUNDED RESEARCH

5.1. BACKGROUND

A sound institutional and regulatory framework is central to an effective and well-functioning STI system. Since matters pertaining to STI transcend all ministries and involve the participation of various stakeholders such as civil servants, industry, academia and the community, issues pertaining to coordination, collaboration and harmonisation assume importance. NPSTI 2021-2030 reinvigorates the nation's existing STI framework in order to enhance the execution of policies besides providing mechanisms to ensure commitment by all parties towards the development of STI in the country.

Hence, the STI governance must be enhanced to ensure effective implementation of policies and strategies with improved transparency and accountability in R,D,C such as in the Open Data Sharing initiative. As such, a Responsive STI Governance is identified as one of the Strategic Thrusts of the NPSTI 2021-2030. This Guidelines will ensure a more efficient and effective delivery system in Open Science with good STI governance.

5.2. GOVERNANCE STRUCTURE

5.2.1. National Level

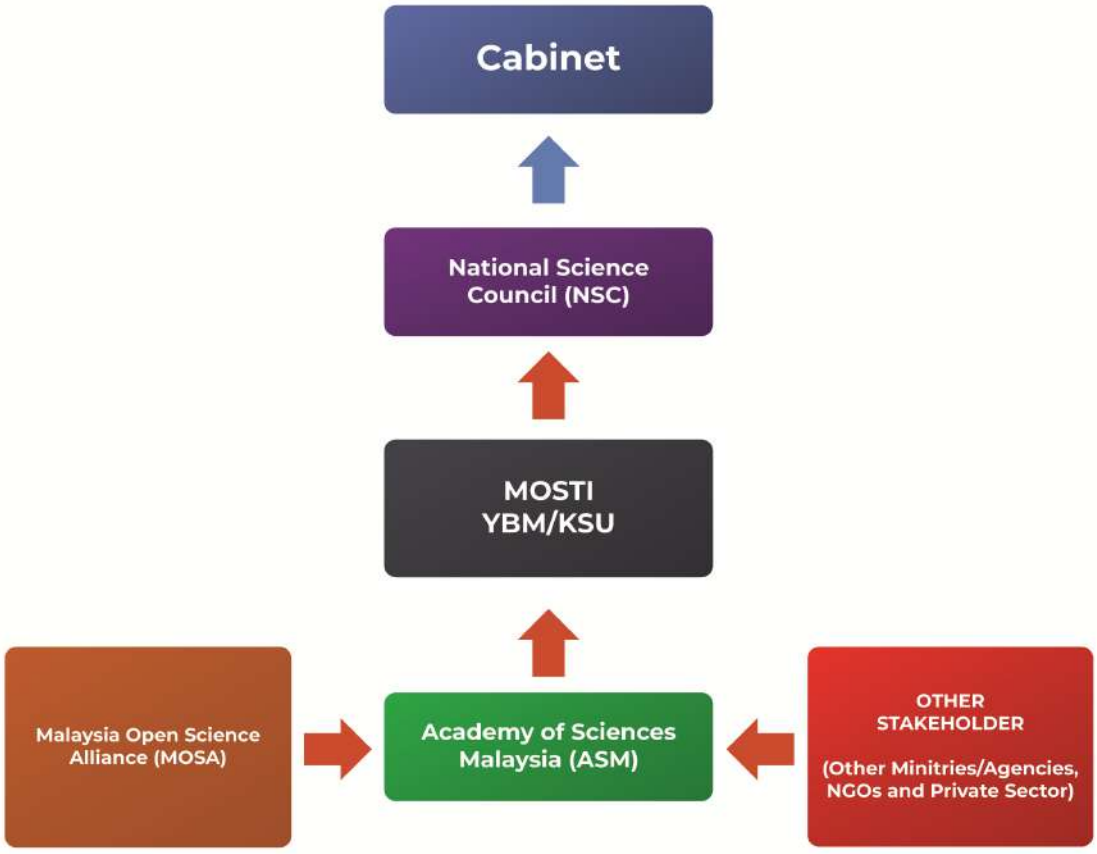
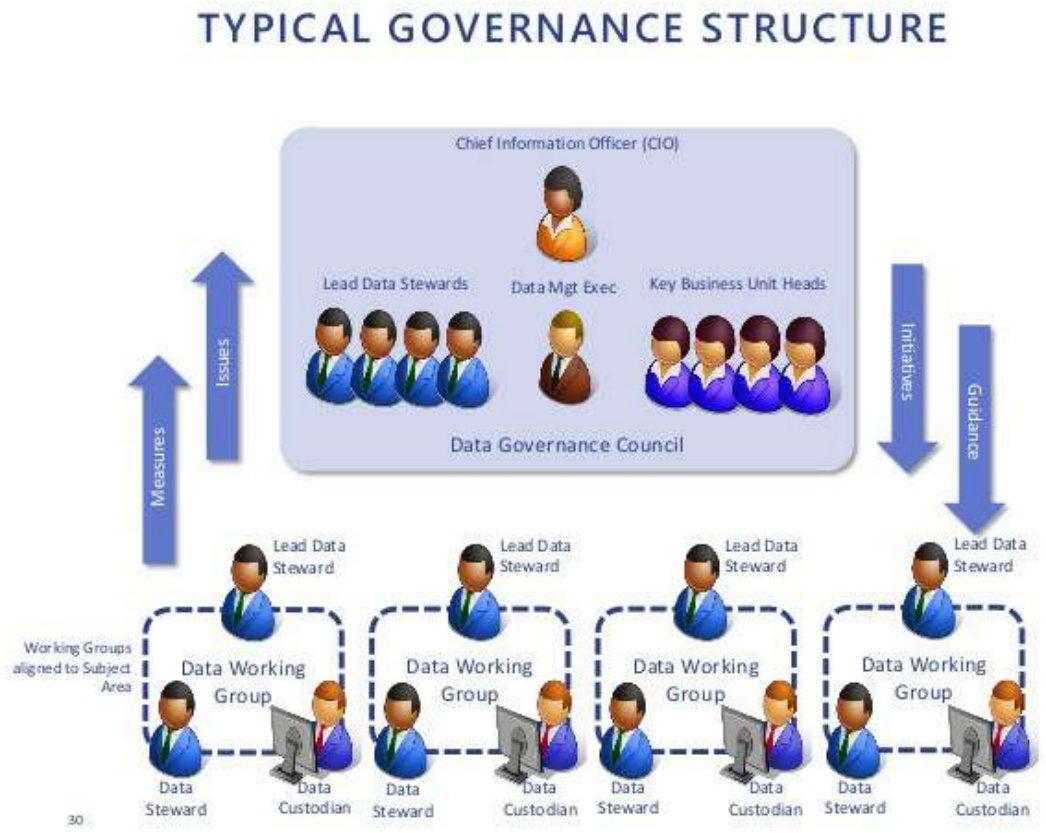


Figure 5.1. Governance of Open Science in Malaysia's Public Funded Research

5.2.2. Institutional Level

At this institutional level involves the governance of Open Science at institutions of higher learnings (IHLs) and the public research institutions (PRIs). Each institution will already have the governing body which has adopted the best international practices in Open Science Management. These institutions will continue to adopt and practise their good governance system and should always ensure they are in line with the national STI agenda, in particular the Open Science Initiative under the NPSTI 2021-2030.

BOX 5.1



Source: Christopher Bradley (2018)

For the purpose of this Guidelines, the recommended structure for the institution of higher learning and public research institution will harmonised as far as possible and will adopt the best practices in Open Science. Some variations are allowed as long as the main core structure exist and does not deviate from the original principle of Open Science. Examples of such governance structures for both kinds of institutions are illustrated below.

a. University or Institution of Higher Learning

Figure 5.2. Governance of Open Science in Malaysia's Higher Learning Institutes

b. Public Research Institute

Figure 5.3. Governance of Open Science in Malaysia's Public Research Institute

5.3. ROLES AND RESPONSIBILITIES

In relation to this Guidelines, the following entities are responsible for the following:

5.3.1. The Ministry of Science, Technology and Innovation (MOSTI):

(a) Familiar with and adhere to legislation, regulatory requirements, contractual obligations, ethical approvals, funding bodies' policies, and other licences and terms of use of research data.

(b) Ensure effective communication processes to manage research data and research records in accordance with the University, Research Institute or Government Entity policy and related Malaysia rules and regulations, or as otherwise determined by other statutory requirements, funding agency guidelines, or contractual arrangements with research partners by providing service support.

(c) Implement, coordinate, and review the execution of the Policy.

(d) Ensure that the Policy is updated on a regular cycle to take into account the latest funder requirements, and national research directives and guidelines.

(e) Provide training, support, advice and guidelines that promote a best-practice approach towards Open Data Sharing in Open Science and FAIR principle.

(f) Monitor compliance of researchers with this Policy and associated procedures.

(g) Acts as a "supra-national entity" that assumes the main role in defining international coordination, collaborations or co-operations in relation to agreements or guidelines in addressing open science issues at international fora.

5.3.2. University, Research Institution and Other Government Entity

(a) Familiar with and adhering to legislation, regulatory requirements, contractual obligations, ethical approvals, funding bodies' policies and other licences and terms of use of research data.

(b) Provide or secure approved IT infrastructure for the safe and secure storage of research.

(c) Ensure backup, archival and monitoring processes are in place to prevent loss of research data.

(d) Provide access to services and facilities for the storage, backup, registration, deposit, curation and archiving of research data.

(e) Provide technical support to maintain all systems (such as the University data repository and DMP submission system) required for compliance with the research data policy.

(f) Ensure that all research projects include a DMP and that it is attached to the relevant record in the Institutional Repository.

(g) Ensure that their Principal Investigators (PIs) adhere to their obligations as detailed in this policy.

(h) Task data stewardship to be responsible for data requirement, data definition and data quality.

5.3.3. Principal Investigators (PIs) And Researchers

(a) Familiar with and adhere to legislation, regulatory requirements, contractual obligations, ethical approvals, funding bodies' policies and other licences and terms of use that pertain to their research data.

(b) Have overall responsibility for the proper and effective management of research data generated during the research project, in accordance with the University, Research Institute or Government Entity policy and guidelines.

(c) Ensure research data are accurate, complete, authentic and reproducible.

(d) Keep clear and accurate records of the research methods and data sources, including any approvals granted, during and after the research process.

(e) Prepare a DMP and submit it online into the institutional repository. PIs and researchers shall provide an updated version whenever there are substantive changes to the research project.

f) Submit the research data to institutional repository no later than the first online publication of the article.

g) Deposit any data which is retained elsewhere at the Institutional Repository no later than two days after the data was deposited into the international data service or domain repository.

h) Ensure that formal agreements are reached with external collaborators and parties, if any, on the ownership, rights, use and sharing of research data arising from the research project before commencement of project.

i) Exclusive rights to reuse or publish research data should not be handed over to any external organisation without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

5.3.4. Data steward

Role:

a) Familiar with and adhere to legislation, regulatory requirements, contractual obligations, ethical approvals, funding bodies' policies and other licences and terms of use of research data.

b) Advise, support and train researchers on data life cycle and good data management practices, from initial planning to post-publication. This includes storing, managing and sharing research outputs such as data, images, models, programmes and codes.

c) Advise and educate researchers on the practices that support open science and reproducibility of research, ethical, policy and legal considerations during data collection, processing and dissemination.

Responsibilities:

a) Implement a data strategy and roadmap aligned with the Open Science goal.

b) Classify the most important data governance.

c) Manage the content of scholarly publications and research data metadata.

d) Prioritise the data quality projects and metrics, and build a data quality programme.

e) Partner with IT on infrastructure.

5.3.5. Data Curator

Role:

a) Provide metadata and ontological support for datasets.

b) Provide support and expertise to researchers engaged in research data management planning, data acquisition, data sharing and long-term data stewardship.

c) Support deposit and stewardship of datasets in digital repository platforms.

Responsibility:

a) Responsible for organising and integrating data collected from various sources, which involves annotation, cataloguing, indexing, publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation.

5.3.6. Open Science Manager

Roles:

a) Drive discussions across the organisation to develop and execute an implementation strategy for open science, especially research data and analysis preservation.

b) Ensure a smooth operation of open science tools and services, and participate in national or international open science collaborations.

c) Develop the organisation's strategic approach to plan for evidence research impact.

Responsibilities:

a) Connect People

Facilitate discussion among stakeholders in the research community, including the researchers, their collaborators, their funding agencies, the industries and the policy makers, to identify present and future open science needs and develop corresponding strategies to address them in a financially sustainable manner.

b) Keep Abreast

In partnership with the IT department and other teams inside or outside the organisation, operate existing data and analysis curation services, monitor their efficiency in serving the research community, expand their adoption and drive further improvements of these tools following the demands of the user community.

c) Strategise

Collaboration represents the interest of the organisation and delivers strategic input to governance bodies of open science collaborations.

d) Benchmark

To identify and develop impact case studies.

5.4. DATA MANAGEMENT IN OPEN SCIENCE

Data Management is defined as an administrative process by which the required data is acquired, validated, stored, protected, and processed, and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users (Business Dictionary).

Figure 5.4. Data Management in Open Science
Source: https://dmp.qut.edu.au/faq (2014)

5.4.1. Data Management Plan (DMP)

a. What is the Data Management Plan (DMP)?

a) A data management plan is a formal document that records how the research data arising from the research project will be handled during and after the project is completed, describing what data will be shared and/or made open, and how it will be curated and preserved.

b) Along the line in making research data Findable, Accessible, Interoperable and Reusable (FAIR), a DMP should comprehensively describe the types of research data that will be collected, processed, curated and preserved, and how these research data can be made, shared and available to be accessed and re-used in compliance with relevant laws and policies which regulate access to and use of data.

b. Why is the Data Management Plan is important?

a) Data Management Plan (DMP) is a key element for good data management. A fully developed DMP will help researchers plan how the research data will be treated throughout their research projects and research data before, during and after the completion of the research project. Ensuring a good DMP in place from the start of the project will improve efficiency, protection quality and access to the created research data, and prevent possible pitfalls such as loss of data, mismanagement in ownership of raw research data and privacy violation.

b) Creating an DMP at the start of a research project is required by Guidelines on Open Science in Public Funded Research. Compliance with the Guidelines is likely to be made compulsory by research funders, the Government, and institutions in the near future.

c) Data Management Plan is a living document that must be updated over the course of the project whenever significant changes are made, including among others, decisions to file for a patent, new data and changes in research project team composition. Once a research project has had its funding application approved and has begun, the Principal Investigator must submit the first version of a DMP as a deliverable of the research project within the first 6 months of the project.

c. What are key components of a Data Management Plan (DMP)?

a) The Guidelines provides an DMP template, detailing the main sections to be included in the DMP. The following DMP template is prepared as a guidance for local institutions to develop its own DMP template in the near future as tabulated in **Table 5.1**.

b) The DMP template aims to maximise the potential for access to and re-use of research data that takes into account the balance of openness and data sharing, proprietary and Intellectual Property Rights, privacy concerns, security risks and protection for certain types of scientific information according to stipulated legal provisions. This DMP template, therefore, recognizes possible opt-outs for research data sharing.

Data Management Plan Checklist

After answering each of the questions below, you should be able to easily create a data management plan.

1 What type of data will be produced?

- ☐ How will data be collected?
- ☐ What would happen if the data were lost or became unusable later?
- ☐ How much data will your project produce, and at what growth rate? How often will it change?
- ☐ What is your data storage and backup strategy?

2 What standards will be used for documentation & metadata?

- ☐ What directory and file naming conventions will be used?
- ☐ What project and data identifiers will be assigned?
- ☐ Is there a community standard for data sharing or integration?

3 What steps will be taken to protect privacy, security, confidentiality, intellectual property, or other rights?

- ☐ Who controls the data (e.g. PI, student, lab, University, funder)?
- ☐ Are there any special privacy or security requirements to uphold (e.g. personal or high-security data)?
- ☐ Are there any embargo periods to follow?

4 If you allow others to reuse your data, how will the data be accessed and shared? How will it be archived for preservation and long-term access?

- ☐ Are there any data sharing requirements?
- ☐ How long should the data be retained?
- ☐ What file formats are you using? Are they proprietary, like .xls or .docx, or are they open, like .csv or .rtf? Are there any special tools and/or software needed to work with or view the data?
- ☐ Are there data repositories that are appropriate for your data?
- ☐ Who will maintain your data for the long term?

guides.lib.olemiss.edu/datamanagement

Table 5.1. Malaysia Open Science Platform FAIR Data Management Plan (DMP) Template

SECTIONS	KEY ITEMS TO BE INCLUDED IN A DMP
1. Data Summary	• Specify the objective of the data collection and its relation to the overall aim of the project.
	• Specify the types and formats of datasets that will be collected or generated.
	• Specify if existing data is being re-used (if any). If yes, has consent or licensed use been obtained from the data owner, or is the data open access or openly licensed?
	• Specify the origin (source) of the data.
	• Specify the expected size of the data (if known).
	• Specify to whom will the data be useful.
2. FAIR Data	
2.1 Making data findable	**Explain how research data can be made findable.**
	• Specify the standard identification mechanism of data. For example, the use of persistent and unique identifiers such as Digital Object Identifiers.
	• Specify naming conventions and versioning that will be used.
	• Specify search keywords that will be used.
	• Specify standards for metadata creation. In case metadata standards do not exist in your discipline, outline what type of metadata will be created and how.
2.2 Making data openly accessible	**Explain how research data can be made openly accessible.**
	• Specify which research data can be made openly available. If certain research data cannot be shared or can be shared but with certain restrictions applied, describe reasons for the restriction by clearly citing whether it is legal or contractual reasons.
	• Specify methods or software tools needed to access the data.
	• Specify if the documentation about the software is needed to access the data. If it is a yes, is it included?
	• Specify if it is possible to include the relevant software, such as via open source code.
	• Specify the location where the data and associated metadata will be stored and be made accessible. Preference should be given to certified repositories which support open access where possible.
	• Specify how access will be granted for research data that are restricted.
	• Specify how the identity of the person accessing the data will be ascertained.
2.3Making data interoperable	**Explain how research data can be made interoperable.**
	• Specify data, metadata vocabularies, standards or methodologies to facilitate interoperability.
	• Specify whether standard vocabulary will be used for all data types to allow inter-disciplinary interoperability.

2.4 Making data reusable	**Explain how and when the research data will be made available for further re-use:**
	• Specify plans for licensing arrangements for the use of shared research data to allow the widest reuse possible.
	• Specify when the research data will be made available for reuse. If applicable, specify plans for embargoing (i.e. a time delay that is applied to research data, codes and other materials before they can be made available, accessible and usable by others) as needed. Be sure to explain the reasons why the embargo period is necessary, whether it is due to research funding policy, patent and clinical trial reasons and for what period an embargo is needed.
	• Specify whether the shared research data is usable by third parties, especially after the end of the project. If legal and ethical restrictions on access and use of sensitive data is applicable, explain the reasons (Is it due to the use of human subjects, research data containing information with national security risks etc.)
	• Specify data quality assurance process.
	• Specify the length of time for which the data will remain reusable.
3. Allocation of resources	• Estimate the costs for making research data to be FAIR. Specify plans to cover the costs.
	• Specify roles and responsibilities for data management in the project.
	• Specify costs and potential value of long-term preservation.
4. Data security	• Specify the level of openness of data, whether it is confidential, restricted or public, and describe how to secure and dispose of the created data. • Address data recovery as well as secure storage and transfer of sensitive data.
5. Data storage and backup	• Specify data storage mechanism (i.e. where do you store the data during and after the research is conducted, and how long will the data be stored?)
	• Specify data backup mechanism (i.e. how many times will a backup on the data be done, will backup be performed on all data or just some of it?
6. Ownership	• Specify who will be responsible to collect the data.
	• Identify if there is a joint ownership between an organisation with an external body or organisation.
	• Identify if there are any contractual agreements that may affect copyright ownership.
	• Identify if the data is collected by an employee of the University in the course of their employment.
	• Identify if the data collected or compiled is in Malaysia or elsewhere.
	• List those involved in the creation of data.
7. Ethics	• Specify if ethics review and clearance is needed. If yes, describe action plans for ethics application.
	• Specify if informed consent for data sharing and re-use and long-term preservation will be sought when dealing with personal data.
8. Other	• Specify if there are any other procedures for data management that will be used (if any).

5.4.2. Metadata Management

a) Good metadata is key for research data access and re-use.

b) Metadata fields selected for the digital repository must match metadata standards including different naming schemes for domain-specific repositories.

c) All metadata for raw research data will be stored at the Malaysia Open Science Platform. Anyone may access the metadata free of charge.

d) The metadata may be re-used in any medium without prior permission from the data originator for not-for-profit purposes provided the persistent identifier or a link to the original metadata record is given.

e) Metadata fields selected for the digital repository must match unqualified Dublin Core metadata fields, as well as including publication and refereed status.

f) The list of metadata fields following Dublin Core are:

g) Title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights, audience, provenance, rights Holder, instructional method, accrual method, accrual periodicity and accrual policy.

5.4.3. Research Data Management (RDM) in Public Funded Research

a. Ownership

(a) Malaysian Universities, Research Institutes, and respective Government Entities own all research data produced by research projects and activities conducted at or under the auspices of the University, Research Institute and Government Entities.

(b) All new contractual agreements for joint projects, studentship agreements and any other type of collaborative agreements with external bodies must comply with the Guidelines.

(c) Exceptions:

i. Where the research funder retains ownership of the research data.

ii. In joint projects, the research data management plans (DMP) shall address the creation, management, confidentiality, retention, and publication of data both digital and non-digital.

(d) The PI and his/her designated researchers have their rights to use and publish research data arising from their project, unless specific terms of sponsorship, other agreements, institutional policies or other relevant national laws and policies supersede these rights.

b. Data Management Plan (DMP) in Public Funded Research

(a) All funded research must include a DMP that records how the research data arising from the research project will be handled during and after the project is completed, describing what data will be shared and/or made open, and how it will be curated and preserved.

(b) DMP must comply with relevant laws which regulate access to and use of data.

c. Deposition

(a) Research data is deposited in the University, Research Institute or Government Entity owned data repository which complies with the FAIR principle (DOI: 10.1038/sdata.2016.18) and will be linked to the MOSP.

(b) The MOSP Platform accepts research data used in establishing and validating research findings, pre-print and post-print materials, publications and reports.

[14] https://www.dcc.ac.uk/guidance/standards/metadata/list

d. Data sharing

(a) The research data shall be made available for sharing via the MOSP unless there are prior formal agreements with external collaborators, funding bodies and parties on nondisclosure or proprietary use of the data.

(b) In the following circumstances, several additional criteria to (a) must be applied:

i. Consent must be obtained from all data subjects for all human data collected and must be anonymised before being deposited and published. The consent form must indicate the use of the data, if it is to be published and reused, and the type of third-class party who may have access to the data.

ii. For human data collected from data subjects under the age of eighteen years, consent must be obtained from the parent, guardian or person who has parental responsibility for the data subject concerned.

iii. For data containing information intended for commercialisation, it must not be deposited until the patent has been filed.

iv. For data that concerns national security matters, it must receive clearance from an authorised body prior to deposition. Access to the data will be completely restricted.

e. Storage and Retention

(a) All research data shall be stored in locations or devices on the Institutional Repository.

(b) Reasonable steps shall be taken to ensure the security and integrity of all research data under retention.

(c) All research data related to a research project or an activity shall be retained not more than ten (10) years after publication or after the completion of the project or last access to the dataset, whichever is later. A longer period of retention may be specified by external research funders. Under both circumstances, the period of retention is subject to legal and regulatory requirements.

(d) Material can be withdrawn from the MOSP, if it is proven copyright violation or plagiarism or falsified research. Withdrawn material is not deleted per se but is removed from public view. Withdrawn items' identifiers are retained indefinitely.

e) If there are major changes to work in the MOSP, an updated version may be deposited as a separate item and can be linked to the first deposited material.

f) Errata and corrigenda lists may be included with the original record if required. If necessary, an updated version may be deposited.

g) In the event of the MOSP being closed, the information stored will be transferred to respective institutions.

f. Data reuse

(a) If the data that will be reused is licensed, the conditions of that licence regarding data sharing i.e. redistribution, must be followed.

(b) If redistribution is permitted, the data may be shared and must be attributed to the originator of the research data.

(c) Data stewards will be able to guide researchers on the information on licensing terms.

(d) If data is re-used, the original author is not implicated with the consequences from the activity if the original author is not involved.

(e) Metadata may be re-used in any medium without prior permission from the data originator for not-for-profit purposes provided the persistent identifier or a link to the original metadata record is given. Anyone may access the metadata free of charge.

g. Disposal

Beyond the period of retention specified here, all research data must be disposed of. Any destruction of the research record, either whole or part shall follow the Guidelines.

CHAPTER 6

IMPLEMENTATION OPEN SCIENCE IN PUBLIC FUNDED RESEARCH

6.1. IMPLEMENTATION AUTHORITY

6.1.1. National Level

The Ministry of Science, Technology and Innovation (MOSTI) will be the implementation authority at the national level.

6.1.2. Institutional Level

At the institutional level, the Public Research Institutes (PRI) and Institutions of Higher Learning (IHL) will respectively be implementing Open Science Guidelines in their entities. For PRI, the Deputy Director General will be responsible for overseeing and managing the operation of the Guidelines while in IHL, this responsibility is undertaken by the respective Deputy Vice Chancellor (Research).

6.2. IMPLEMENTATION REQUIREMENTS

The fundamental principle underlying the Guidelines as mandated is that the results of research carried out with public funds shall be openly available for both other researchers and the interested public and companies. This is based on the accepted arguments that open access to scientific information is good to create better conditions for scientific research and for society at large. In addition to its research value, there is also a clear emphasis on the social benefits of open science.

In implementing the Guidelines on Open Science in Funded Public Research the followings are required:

a) Malaysia Open Science Alliance is responsible to ensure that the guiding principles in the Guidelines on Open Science in Public Funded Research is observed and implemented holistically across all levels of relevant stakeholders. Malaysia Open Science Alliance, as well as Institute of Higher Learnings, Government Research Institutes and Non-government Research Institutes may use assessment indicators for implementation of Malaysia Open Science in the form of "Checklist for Implementation of Raw Research Data Repositories at Institute of Higher Learnings, Government Research Institutes and Non-government Research Institutes" and "Checklist for Researchers Readiness to Share Raw Research Data" as in **Appendix 6.1 and Appendix 6.2**, respectively.

b) This Guidelines on Open Science in Public Funded Research is subjected to reviews and amendments at any time, as it deems necessary in the interest of technological changes, applications, procedures, legislations and societal benefits.

c) All beneficiaries of publicly funded research activities must follow this Guidelines on Open Science in Public Funded Research to achieve an effective and efficient implementation of Malaysia Open Science Platform that will empower planning and management of research, development, commercialisation and innovation.

CONTACT INFORMATION

For queries relating to this document please contact:

Guidelines Owner:	Academy of Sciences Malaysia (ASM)
All Enquiries Contact:	Academy of Sciences Malaysia (ASM) Level 20, West Wing, MATRADE Tower, Jln Sultan Hj Ahmad Shah, Off Jln Tuanku Abd Halim, 50480 Kuala Lumpur.
Telephone:	03-6203 0633
Email address:	admin@akademisains.gov.my

BIBLIOGRAPHY

(ARDC), A. R. D. C., 2018. Research Data Rights Management Guide. [Online].

(CRUK), C. R. U., 2009. Data Sharing Guidelines. [Online]
Available at: https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

(MOSP), M. O. S. P., 2020. Infographic: New Career Path Acknowledging Open Science Practices. s.l.:s.n.

(MOSP), M. O. S. P., 2020. Infographic: Research Data Lifecycle. s.l.:s.n.

(NSF), N. S. F., 2011. Data Sharing Policy. [Online].

(UKRI), U. R. a. I., n.d. Common Principles on Data Policy. [Online].

Bradley, C., 2013. Implementing Effective Data Governance. [Online]
Available at: https://www.slideshare.net/inforacer/impdata-gover

Brous, P., Janssen, M. & Krans, R., 2020. Data Governance as Success Factor for Data Science. s.l.:Springer.

Candela, L., Castelli, D. & Pagano, P., 2013. Virtual Research Environments: An Overview and a Research Agenda. s.l., Data Science Journal.

Council, S. R., 2015. Proposal for national guidelines for open access to scientific information, Stockholm: s.n.

Ferrari, T., Scardaci, D. & Andreozzi, S., 2018. The Open Science Commons for the European Research Area. In: Earth Observation Open Science and Innovation. s.l.:Springer.

Group, T. C. W., 2016. Concordat on Open Research Data. [Online]
Available at: https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf

Guidelines, M. O. S. A. W. G. o., 2020. First Stakeholder Engagement Rapporteur Report, s.l.: s.n.

Guidelines, W. G. o., 2020. Minutes of 3rd Meeting. s.l.:s.n.

Ladley, J., 2019. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. 2nd ed. s.l.:Elsevier.

Mathieu, P.-P. & Christoph, A., 2018. Earth Observation Open Science and Innovation. s.l.:Springer.

OECD, 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding, s.l.: OECD Publishing.

Springer, n.d. What is Open Access. [Online]
Available at: https://www.springer.com/gp/authors-editors/authorandreviewertutorials/open-access/what-is-open-access/10286522

APPENDICES

CHAPTER 6

Appendix 6.1. Checklist for Implementation of Raw Research Data Repositories at Institutes of Higher Learnings, Government Research Institutes and Non-government Research Institutes

Number	Activity	Status (/)
1	Establishment of a new or existing governance structure to develop planning, execute tasks and activities according to the implementation planning related to Open Science.	
2	Development and implementation of Open Science-related policies such as Research Data Management Policy.	
3	Establishment of institutional repositories or identification of data publishing partners repositories for raw research data deposition.	
4	Identification of raw research datasets that can be deposited on institutional repositories or data publishing partners' repositories.,	
5	Ensuring that the respective repositories are supported with a backup system, scalable and interoperable to Malaysia Open Science Platform (MOSP) Portal.	
6	Ensuring that the data classification for the data repository systems are determined.	
7	Ensuring that the data repository systems are established with a strong and robust security system, possibly with ISO 27001.	
8	Formal training for data stewards and their active participations in Open Science activities including to assist researchers to develop a Data Management Plan.	
9	Formal training for Principal Investigators and researchers about key practices for Open Science and preparation to write a Data Management Plan.	
10	Formal education for undergraduate, graduate and postgraduate students to equip them with essential knowledge and skills about principles and best practices for Open Science.	
11	Preparation of a Data Management Plan is made compulsory for research grant beneficiaries at institutional levels.	
12	Recognition of Open Science and data sharing practices in the form of academic rewards systems, such as for security tenure and career promotion assessment.	
13	Effective communication about Open Science and its incentives across all levels.	
14	Review and evaluate implementation and address emerging gaps.	

APPENDIX 6.2. Checklist for Researchers Readiness to Share Raw Research Data

Reference: https://guides.library.uq.edu.au/c.php?g=500758&p=3429166

Number	Activity	Status (/)
Before depositing raw research datasets		
1	A researcher creates an ORCID identifier before proceeding to publish raw research data.	
2	In the case of publishing raw research data that underpins a publication, a researcher must ensure that the datasets contain all the information needed to support the conclusions and quantitative statements as evidenced in the publication.	
3	Make decisions on types of access to the raw research data. Although ideally, raw research data should be made shareable, conditional or restricted access are also available under the institutional repository.	
4	Identify if the files saved with file extensions are acceptable to the institutional repository. If it is not, identify alternative file extensions that are acceptable.	
5	Consider converting the files into a non-proprietary form to increase access and use of your data.	
6	Prepare and upload supplementary documentation of raw research data that better explain the data to decrease misinterpretation of the data.	
7	Prepare and upload a Data Management Plan that tags along with the raw research datasets. Consult a data steward when developing the document.	
8	Specify if there is an embargo period applicable to the raw research data.	
9	Nominate a data steward or contact person for the raw research datasets in case the originator or creator of the raw research data cannot be contacted.	
Legal, ethical and commercial considerations		
11	Identify if the created raw research datasets arise from the institution that houses the repository. This also includes projects conducted with other institutions.	
12	Verify if the funding agreement permits the raw research datasets to be published at an institutional repository.	
13	Verify if there is any third-party data included in the files? If yes, has it been consented by the third-party for it to be deposited and published?	
14	Verify if the raw research data has confidentiality implications upon sharing it on an institutional repository. Special actions must be undertaken to treat such datasets, including anonymisation.	
15	Verify if the raw research data may concern a pending patent. Special actions must be undertaken to treat such datasets. Clarify with a technology transfer office if the data is commercially exploitable.	

Think Science.
Celebrate Technology.
Inspire Innovation.

ACADEMY OF SCIENCES MALAYSIA
Level 20, West Wing, MATRADE Tower, Jalan Sultan Haji Ahmad Shah
off Jalan Tuanku Abdul Halim, 50480 Kuala Lumpur
T: +6 (03) 6203 0633 / F: +6 (03) 6203 0634
www.akademisains.gov.my

ISBN 978-983-2915-63-8