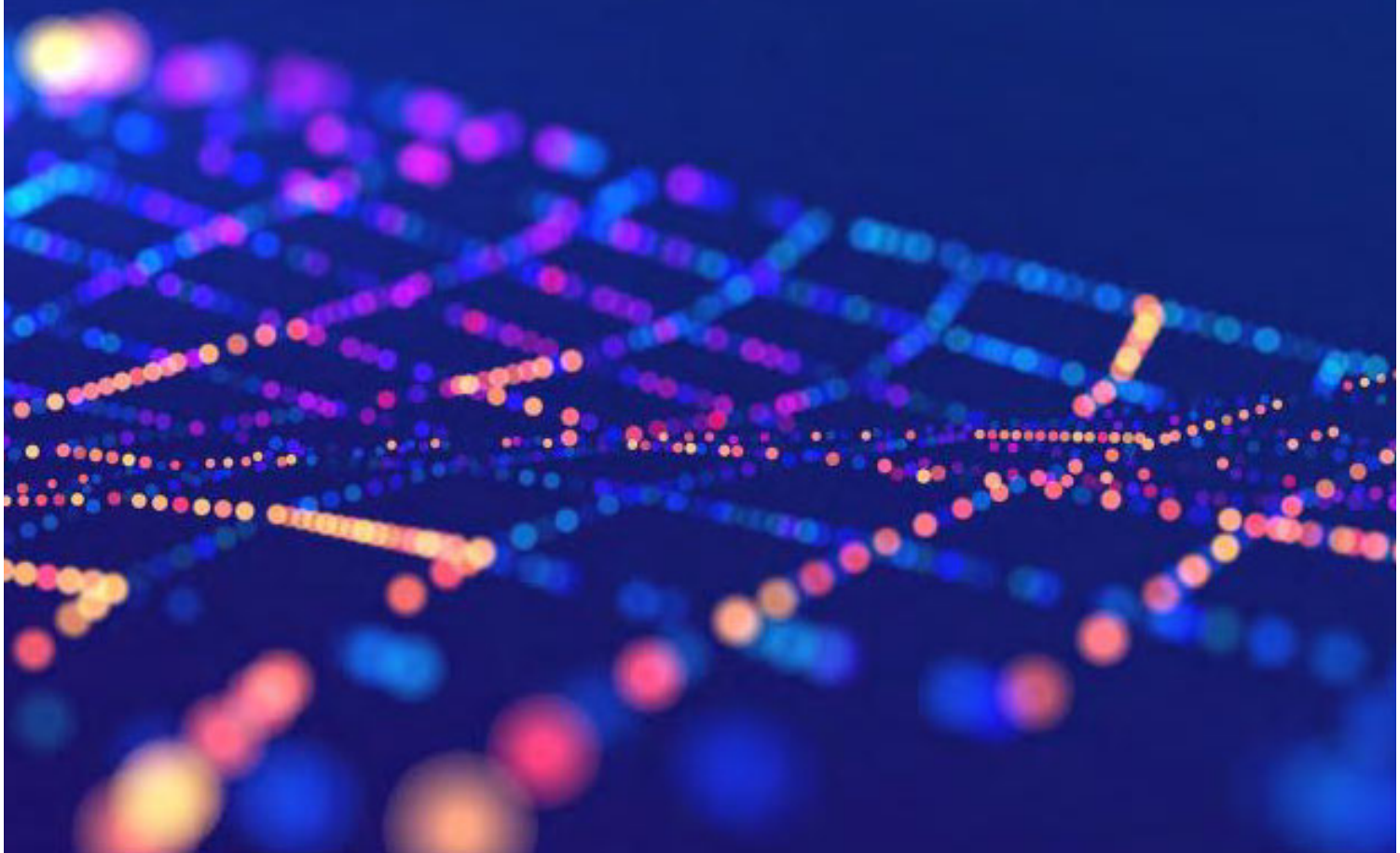
KEMENTERIAN SAINS, TEKNOLOGI DAN INOVASI
MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

NATIONAL GUIDELINES ON
OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH

Malaysia Open Science Platform (MOSP)



GUIDELINES

NATIONAL GUIDELINES ON OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH

2021

NATIONAL GUIDELINES ON OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH

This National Guidelines was prepared as a deliverable for the Malaysia Open Science Alliance Working Group on Guidelines.

Academy of Sciences Malaysia
Level 20, West Wing, MATRADE Tower
Jalan Sultan Haji Ahmad Shah off Jalan Tuanku Abdul Halim
50480 Kuala Lumpur, Malaysia

Perpustakaan Negara Malaysia Cataloguing-In-Publication Data
The Landscape of Open Science in Malaysia

ISBN application in process

CONTENTS

FOREWORD

The Malaysian Open Science Platform (MOSP) in collaboration with the Academy of Science (ASM) has been tasked by the Government in drafting a **National Guidelines on Open Science in Public Funded Research.** The mandate given was in line with the Open Science Initiative championed by MOSP. The proposal for the guidelines was drafted in collaboration and consultation with various stakeholders through a series of information and dialogue meetings. The views that emerged during these meetings were taken into account in the work on producing the proposal for national guidelines.

The National guidelines will help to ensure **scholarly publications and research data** resulting from **publicly funded research** to be openly and publicly available. Information and Communication Technologies (ICTs) as an enabler, has however made it easier and doable. As such openness in obtaining, processing, publishing and disseminating research information has become easily achievable due to the spread of ICTs and ICT-enabled services.

In addition, there are some socio-economic benefits and diverse opportunities to be derived from Open Science. Perhaps the most important reasons are the broad economic benefits and growth, both public and private. Scholarly publications and research data made available and accessible through Open Science have been shown to be economic force enhancers and multipliers, creating value many times over and providing much greater returns on public research investments. The generative or pro-creative effects as a result of Open Science are key in this regard.

Undoubtedly, Open Science will have an effect on society's social welfare. Not only will it meet society's expectations on appropriate management of Open Science assets and resources, it will also provide diverse reputational gains apart from incorporating ethical principles for accessing and using scholarly publications and research data. In the public research it can substantially reduce unproductive barriers to interdisciplinary, inter-institutional, and international research. Besides enabling data mining for knowledge discovery in a growing sea of big data, Open Science is essential for the verification of research results and in generating broad trust in them. It avoids many inefficiencies, such as the unnecessary duplication of research and the identification of erroneous results. Open Science will promote more research and new types of research. It also permits the legal interoperability of data when multiple sources of data are combined for new knowledge.

Finally, Open Science will help to improve governance. Public data made openly available through the public institutional portals will support improved decision-making and transparency in government and society. For a developing economy like Malaysia, Open Science will help to build freedom in society, and trust in governance and its many functions.

ACKNOWLEDGEMENT

We would like to extend our appreciation and thanks to all those that have participated in the study. We acknowledged with thanks the comments and suggestions imparted and guidance extended during the series of engagement with the Malaysia Open Science Alliance and its Working Group on Policy, the Academy of Science Management/ Secretariat, the management, the researchers and the librarians of the selected Malaysian public and private institutions of higher learning (IHL) and the public research institutions (PRIs).

MALAYSIA OPEN SCIENCE ALLIANCE, WORKING GROUP ON POLICY AND SECRETARIAT/ASM MANAGEMENT

MALAYSIA OPEN SCIENCE ALLIANCE

1. Professor Dr Noorsaadah Abd Rahman FASc (Chairperson, UM)
2. Professor Datuk Dr Rohana Yusof FASc (ASM)
3. Academician Professor Emerita Datuk Dr Mazlan Othman FASc (ISC-ROAP)
4. Professor Dr Shaliza Ibrahim (UM)
5. Professor Dr Ismanizan Ismail (UKM)
6. Professor Dr Zulkifli Idrus (UPM)
7. Professor Dr Rosni Abdullah (USM)
8. Professor Dr Abrizah Abdullah (UM)
9. Puan Roshana Alma Mohd Ali (MOSTI)
10. Puan Noraziah Sharuddin (UTM)
11. Dr Liew Chee Sun (UM)
12. Dr Salina Abdul Aziz (MOH)
13. Encik Azih Yusof (MAMPU)
14. Puan Zaiha Mat Nor (MAMPU)
15. Encik Kamaruhzaman Mat Zin (MASTIC)

MALAYSIA OPEN SCIENCE ALLIANCE WORKING GROUP ON GUIDELINES

1. Professor Datuk Dr Rohana Yusof FASc (Chairperson, ASM)
2. Emeritus Professor Dato' Dr Rahmah Mohamed FASc (ASM)
3. Dr Khasiah Zakaria (UM)
4. Dr Sik Cheng Peng (UM)
5. Dr Nasuha Lee Abdullah (USM)
6. Dr Syed Norris Hikmi Syed Abdullah (UTM)
7. Puan Roshana Alma Ali (MOSTI)
8. Encik Kamaruhzaman Mat Zin (MASTIC)
9. Encik Kamarulzaman Kamaruddin (MOSTI)
10. Professor Dr Ismanizan Ismail (UKM)
11. Professor Dr Samsilah Roslan (UPM)

ACADEMY OF SCIENCES MALAYSIA - MALAYSIA OPEN SCIENCE PLATFORM

1. Puan Hazami Habib (Chief Executive Officer)
2. Dr Nurzatil Sharleeza Mat Jalaluddin (Former Project Manager)
3. Nurlizah Abu Bakar (Project manager)
4. Shyama Sundari Devi (MOSP Analyst)
5. Nur Hanisah Ismail (MOSP Analyst)
6. Nur Faizah Azhari (MOSP Anlayst)
7. Shaanthana Subramaniam (MOSP Analyst)
8. Muhamad Fariz Muhamad Suhaimi (MOSP Analyst)

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

ABBREVIATIONS AND ACRONYMS

APCs	Article Processing Charges
API	Application Programming Interface
ARDC	Australian Research Data Commons
BPCs	Book Processing Charges
DMP	Data Management Plan
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
DVD	Digital Versatile Disc
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable and Reusable
FOSTER	Facilitate Open Science Training for European Research
ICT	Information and Communications Technology
ID	Identification/Identity
IHLs	Institute of Higher Learnings
IP	Intellectual Property
ISC	International Science Council
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MOHE	Ministry of Higher Education
MOSP	Malaysia Open Science Platform
MOSTI	Ministry of Science, Technology, and Innovation
NCBI	National Centre for Biotechnology Information
NGO	Non-Government Organization
NPSTI	National Policy on Science, Technology and Innovation
OECD	Organisation for Economic Co-operation and Development
ORCID	Open Researcher and Contributor ID
PLOS	Public Library of Science
PIs	Principal Investigators
PRI	Public Research Institutes
RAND	RAND Corporation
RDM	Research Data Management
RI	Research Institutes
RU	Research University
SSH	Social Science and Humanities
STEM	Science, Technology, Engineering and Mathematics
STI	Science, Technology and Innovation
UKM	Universiti Kebangsaan Malaysia
UM	Universiti Malaya
UPM	Universiti Putra Malaysia
US	United States
USM	Universiti Sains Malaysia
UTM	Universiti Teknologi Malaysia
VRE	Virtual Research Environment

GLOSSARY

Article Processing Charges (APCs): fees that some scholarly publishers charge authors of academic papers to publish their work in open access.

Book Processing Charges (BPCs): fees charged by a publisher to make a book open access.

Confidential: highly restricted information due to the law such as Data Protection, policy, agreement or duty of confidence arising from the nature of relationship between the parties. Inappropriate disclosure of the information would be likely to cause serious damage or distress to individuals and/or constitute unfair/unlawful processing of "sensitive personal data" under the Data Protection Act; and/or seriously damage the government and institution interests and reputation; and/or significantly threaten the national security.

Copyrights: collection of legal rights that attach to an original work when it is created. Copyright allows the copyright owner to control certain acts to do with their work (e.g. copying) and to prevent others from using the protected material without permission.

Data Curator: responsible for organizing and integrating data collected from various sources. It involves publication, presentation, reuse and preservation of the data.

Data Custodian: Data owners are also data custodians who own the data storage facilities. A data custodian is an IT individual or organisation responsible for the IT infrastructure providing and protecting data in conformance with the policies and practices prescribed by data governance.

Data Governance: A cross-functional management programme that treats data as an organisation asset through collection of policies, standards, process, people and technology to achieve a set of goals.

Data Management Plan (DMP): a living 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.

Data Originators: Researchers who produce research data and who be credited for their work. Also known as data creator.

Data Owner: Institutions, which also are employer of researchers, or the research institutions receiving and administering the grants.

Digital Repository: an on-line archive for collecting, preserving and disseminating digital copies of the intellectual research outputs.

Data Sharing: Data can be shared at any time either publicly or privately among collaborators, while the proper documentation and code is open source to ensure that others can build on and benefit from.

Data Steward: protect the integrity and quality of data, adherence or compliance to standards and protocols, governance and advocacy. The role of data stewards complements with data curators in the aspects of both metadata management activities and data governance.

Data User: Individuals who re-use data and have responsibilities to acknowledge the sources of their data by citation or giving appropriate credits to data originators.

Errata: a list of errors and their corrections inserted, usually on a separate page or slip of paper, in a book or other publication. This is also referred to as corrigenda.

Embargo: the period during which a publication can be 'closed' while deposited in the repository (i.e. the publication is not openly available).

FAIR Data Principles: refer to a set of principles to make data Findable, Accessible, Interoperable and Reusable for scientific management, data stewardship and Open Science framework.

Gold Open Access: makes the final published version of an article freely available and permanently accessible for everyone, immediately after publication.

Green Open Access: also known as "self-archiving", it is "the practice of placing a version of an author's manuscript into a repository, making it freely accessible for everyone."[1] The version (pre-print or post-print) that can be deposited into a repository is dependent on the funder or publisher.

Metadata: means "data about data". Metadata are the descriptors used for describing, tracing, use and management of the deposited item. Metadata describes characteristics such as content, quality, format, location and contact information.

Open Access: its freely availability on the public internet, permitting any users to read, download, copy, distribute, print, search or link to the full texts of these articles, crawl them for indexing, pass them as data to software or use them for any other lawful purpose without financial, legal or technical barriers other than those inseparable from gaining access to the internet itself.

Open Data: is defined in essence, as data that can be freely used, re-used and redistributed by anyone. Besides commonly associated with Open Government Data, Open Data also refers to Open Business Data and Citizen Generated Data. The main criteria for Open data are complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary and license-free.

Open Peer Review: a scholarly review mechanism where both the identities of the reviewer and the author are known to one another during the review and publication process.

Pre-print: refers to the version of an academic paper which is submitted by an author for peer review.

Post-print: refers to final version of an academic paper before publication, incorporating the revisions made as a result of the peer review process or as accepted for publication if no changes were made.

Research: defined as any creative and systematically performed work with the goal of furthering knowledge.

Research data: any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats.

Research Data Lifecyle: consists of data acquisition, processing, analysis, curation, sharing and re-use. The data life cycle is divided into two domains i.e. private (green-colour coded) and public (blue-colour coded).

Research Data Management (RDM): concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.

Restricted data: data that is restricted or prohibited from disclosure. Restricted data would include confidential data. In some circumstances, access to sensitive data can be restricted, depending on whether there is any express prohibition or policy discouraging its disclosure.

Sensitive data: data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions.

[1] https://www.springer.com/gp/authors-editors/authorandreviewertutorials/open-access/what-is-open-access/10286522

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND

Innovative scientific research through public research funding has a crucial role in addressing national and global challenges. But this research is only meaningful if it is translated and depends a lot on how it is fostered. Fostering collaborative exchanges through Open Science between different scientific and research communities and other communities at large and assuring its widest dissemination in terms of speed and depth is also crucial too. In short, the exchange of ideas, knowledge and data emerging scientific research through Open Science is vital for country's progress and development in terms of knowledge creation, wealth creation and societal well-being. Thus, making scientific research and data, transparent and accessible to all is core.

Open Science is an initiative to make research output such as data and publications more transparent and accessible. It is about extending the principles of openness to the whole research cycle based on cooperative work and new ways of diffusing knowledge through digital technologies and collaborative tools[1]. For those research outputs to be accessible and can be shared by "everyone", they should be properly managed and curated, meeting the principles of Findable, Accessible, Interoperable and Reusable (FAIR). With FAIR data, researchers are able to create, share and re-use quality, valuable, high integrity and responsible data, fueling scientific progress to its fullest potential.

Admittedly, Open Science is gaining worldwide consensus as more countries have introduced and implemented the initiative at the national and regional levels. In Malaysia, Open Science is introduced through the Malaysia Open Science Platform (MOSP), an initiative managed by the Academy of Sciences Malaysia through the Malaysia Open Science Alliance, and funded by the Ministry of Science, Technology and Innovation (MOSTI). MOSP aims to gather and consolidate Malaysia's research data which are valuable national assets in a platform that would enable accessibility and sharing of these research data in accordance to the **FAIR** principle. In short, open data sharing is the way while embracing the FAIR principles. This Platform represented as a strategic transformative initiative to strengthen Malaysia's STI collaborative ecosystem towards achieving Shared Prosperity Vision 2030 and addressing the United Nations Sustainable Development Goals.

In this regard, MOSP has embarked on a two-year pilot project, launched on 7 November 2019. It has tasked the Malaysia Open Science Alliance is to look into three main areas, which are (1) National Guidelines, (2) Capacity Building and Awareness and (3) Infrastructure. The development of a **National Guidelines on Open Data Sharing in Research** is one such commitment which is in line with one of the strategies in the **National Policy on Science, Technology and Innovation (NPSTI 2021-2030)**. The formulation of the Guidelines document has taken into account and consolidated inputs from all relevant stakeholders including researchers, top management universities, government agencies, libraries, research funder organizations, publishers, legal units, industries and research managers.

1.2. THE RATIONALE FOR A NATIONAL GUIDELINES

There remained some concerns and challenges in Open Science, and the way they being access to be addressed. Among the issues raised include the potential misuse of the data, which could result to

misunderstanding of the meaning and compromised quality of shared data either among researchers or the public. Other opinions highlighted the need to formulate a strong and robust National Policy and Guidelines on Open Science in Public Funded Research which unequivocally clarified research data ownership, recognized the role of data stewards and data curators, provided rewards and incentives to data contributors as well as established rigorous security and privacy standards for data sharing practices.

More importantly, for MOSP to be successful, each initiative must be organized holistically, integrated and coherent with the overall goal, understood across all levels especially by researchers, and involves everyone's participation. This is important to have a strong coherence with the mission, the vision and strategic thrusts outlined in the National Policy on Science, Technology and Innovation (NPSTI 2021-2030). The Ministry of Science, Technology and Innovation (MOSTI) is responsible for the promulgation and implementation of the Guidelines.

1.3. PURPOSE

The Guidelines provides the best practices for applying Open Science and achieving its fundamental goals. The main objectives of the Guidelines are to:

a) Provide better management of research data;
b) Assist researchers in Malaysia to deposit and retain research data files and datasets, publications and records and to contribute to scientific advancements through its availability for sharing; and
c) Ensure that research data generated in the conduct of research activities in all institutions are managed in a systematic and comprehensive manner to ensure quality, integrity, accountability, long term availability, appropriate sharing and compliance with the requirements of funding agencies.

1.4. DEFINITION OF OPEN SCIENCE[2] AND OPEN ACCESS

For the purpose of this Guidelines, the OECD's definition will be adopted as follows:

Open Science is defined as "efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration and fostering innovation."

Open Access is defined as the right to read, download and print as well as the right to copy, distribute, search, link, crawl and mine. However, the accessibility based on the rights must subscribed and be subjected to the FAIR principle.

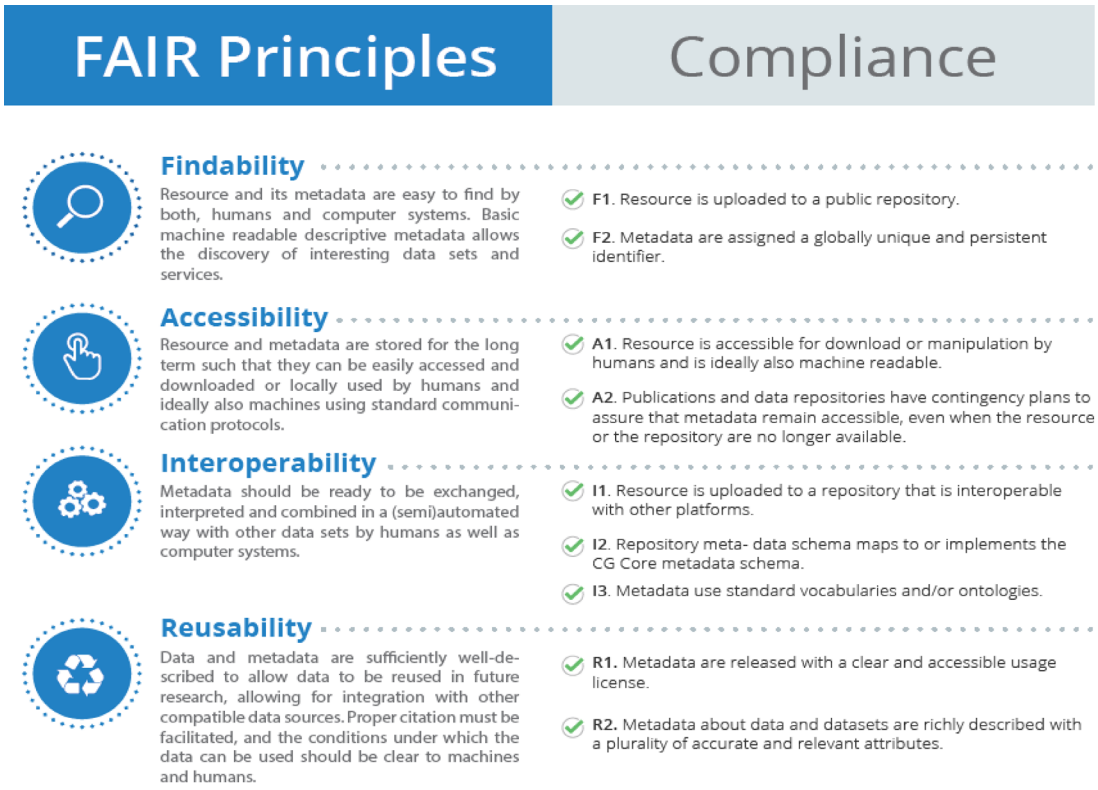
[2] There other main acceptable definitions of Open Science globally as follows:
a) OECD2: "…efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation."
b) FOSTER3: "…the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods."
c) RAND Corporation4: "Open Science refers to ongoing changes in the way research is conducted: for scientists themselves, through increasing the use of open access scientific publishing and open data, and for the public, through increasing their understanding of and participation in science … Open Science is one of three priority areas for European research, science and innovation policy."

1.5. ADOPTION OF FAIR PRINCIPLES

Open science should be principle-based but be adapted to local realities. Based on the above definitions, the principles of **Findable, Accessible, Interoperable and Reusable (FAIR)** will be adopted under this Guidelines. FAIR data means that data is not always open, but it should be as open as possible, and as closed as necessary. With FAIR data, researchers can create, share and re-use quality, valuable, high integrity and responsible data, fuelling scientific progress to its fullest potential.

- **Findable** means that data and metadata are easily to find by both humans and computers. Usually, this task is enabled by machine-readable persistent identifiers and metadata.
- **Accessible** means that data can be retrieved using the outline protocols by appropriate people, at an appropriate time and in an appropriate way. Data can be FAIR even if the data has various levels of accessibility, such as: (1) Data is completely private; (2) Data is accessible by a defined group of people; and (3) Data is accessible by everyone.
- **Interoperable** means that the terminology system, protocols, standards and formats built and employed for datasets that are stored in a platform can be used and can communicate with other tools or platforms.
- **Reusable** means that data is well-defined and can be used for different purposes and in different settings, and the legal use is regulated by different terms and conditions. Data can be FAIR even if the data has various levels of reusability, depending on the stipulated licensing terms (E.g. acknowledgement, access and methods of data re-use, charges, exemption use of personal, sensitive and restricted data and proprietary information).

BOX 1.1. FAIR PRINCIPLES



Source: Horizon 2020, European Commission

1.6. ABOUT THIS NATIONAL GUIDELINES

a) The National Guidelines on Open Science in Public Funded Research sets out the directions in spurring Open Science movement in Malaysia. This document serves as a national guideline that defines types of published or documented research data including raw data that can be shared under specified conditions, harmonises definitions and terminologies, and outlines incentives for data sharing to promote the culture of openness where raw research data are shared among research from various disciplines in Malaysia, and to support good practice for raw research data sharing.

b) The National Guidelines also develops guidelines for the management of open data sharing and research data management plan that outlines how research data arising from the research project will be handled during and after the project is completed, by describing what data will be shared and/or made open, and how it will be curated and preserved to ensure that the raw research data is accessible beyond the life of the project. Malaysia Open Science Platform (MOSP) encourages all researchers to prepare a data management plan for publicly funded research projects to ensure that the raw research data generated by research projects be deposited at institutional repositories or data publishing partners repositories and all metadata permanently archived in the Malaysia Open Science Platform

1.7. APPLICATION OF NATIONAL GUIDELINES

This Guidelines shall apply to:

(a) all staffs, researchers, students and any other persons involved in the design, conduct, administration, or reporting of research performed at or under the auspices of Malaysian Universities, Research Institutes, and the Government Entities including consultants and visiting researchers.

(b) all research activities conducted in all universities, research institutes and government entities that received funding from the Malaysian government.

(c) all stages of data life cycle- before, during and after. This Guidelines will operate in conjunction with other related national and institutional policies and guidelines and the Malaysian Laws and government policies.

The Guidelines shall not apply to:

(a) Research activities conducted for third party organizations using private or international funding.

(b) Consultancy services conducted for third party organisations including work carried out using or by Malaysian Universities, Research Institutes and the Government Entities research facilities.

The National Guidelines on Open Science is prepared by referring to the existing Policy and Guidelines for Open Science which were developed in other countries, as well as based on the relevant laws, policies and regulations related to Open Science in Malaysia. Since the Guidelines is intended to first raise awareness and establish early adopters of Open Science, it is so designed to ensure the realization

of Open Science in the country at this stage. The Guidelines will be revised as and when necessary. In making Open Science though Open Access a reality the current Guidelines will adopt three kinds of measure: mandatory rules (sticks), incentive mechanisms (carrots), and "enablers (soft and hard infrastructure)" as follows:

a) **Mandatory rules:** compelling open data sharing in Open Science a compulsory requirement or a pre-requisite in research grant agreements or spelt out in national strategies or institutional policy.
b) **Incentive mechanisms:** incentivising those that involved or celebrating or promoting Open Data Sharing in Open Science in the form of a financial incentive such as to cover the cost related "open access publishing or the release of data sets." The incentive can also be in non-financial form such as awarding recognition or official acknowledgment (or even involved career advancement) to researchers and academicians who are strong advocates on open science.
c) **Enablers:** involves the development of soft and hard infrastructure such as building a trusted sharing platform on Open Science (that allows ease of sharing of scientific articles or research data), sets of skills and training etc.

Since Open Science is defined as a global movement to make **scientific research, data and dissemination accessible** to all levels of an inquiring society, amateur or professional, the focus of the Guidelines on Open Science will be an on open access to scientific 6research documents and research data including raw data in public funded research.

1.8. DATA SHARING FRAMEWORK IN OPEN SCIENCE

The framework is based on the 3 measures and 2 main areas as depicted in **Figure 1.1**. It will describe the **rewards and incentives** to research and data contributors as well as the best practices they need to adopt together with the established rigorous security and privacy standards they need to conform or adhered to as part of the **mandatory rules** since open access does not mean "open to all" or everything needs to be opened or disclosed.

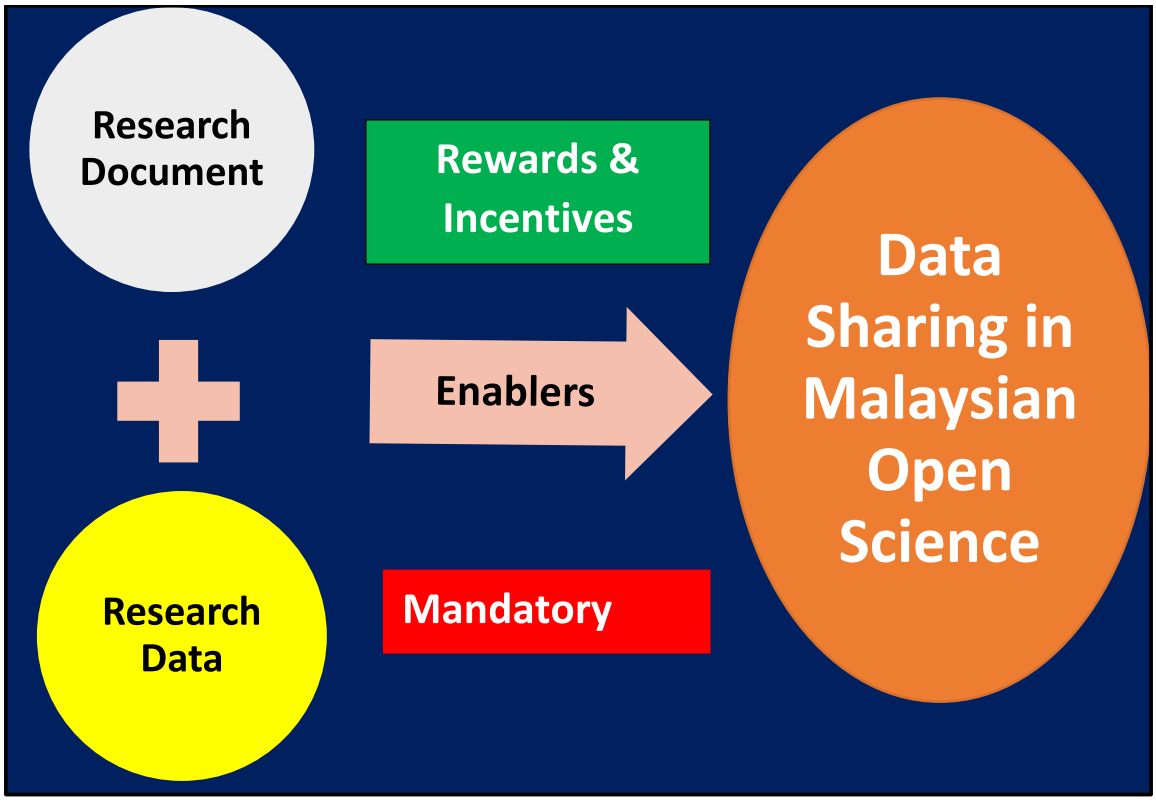


Figure 1.1. Data Sharing Framework in Open Science

CHAPTER 2

OPEN ACCESS TO SCHOLARLY PUBLICATIONS AND RESEARCH DATA

2.1. BACKGROUND

The Guidelines explain the rules on open data sharing through open access to scientific research documents and research data that beneficiaries have to follow in projects funded under government or public funds. **Figure 2.1**. described the general routes practised and adopted.[3] This Guidelines will adopt similar route in the open data sharing process.

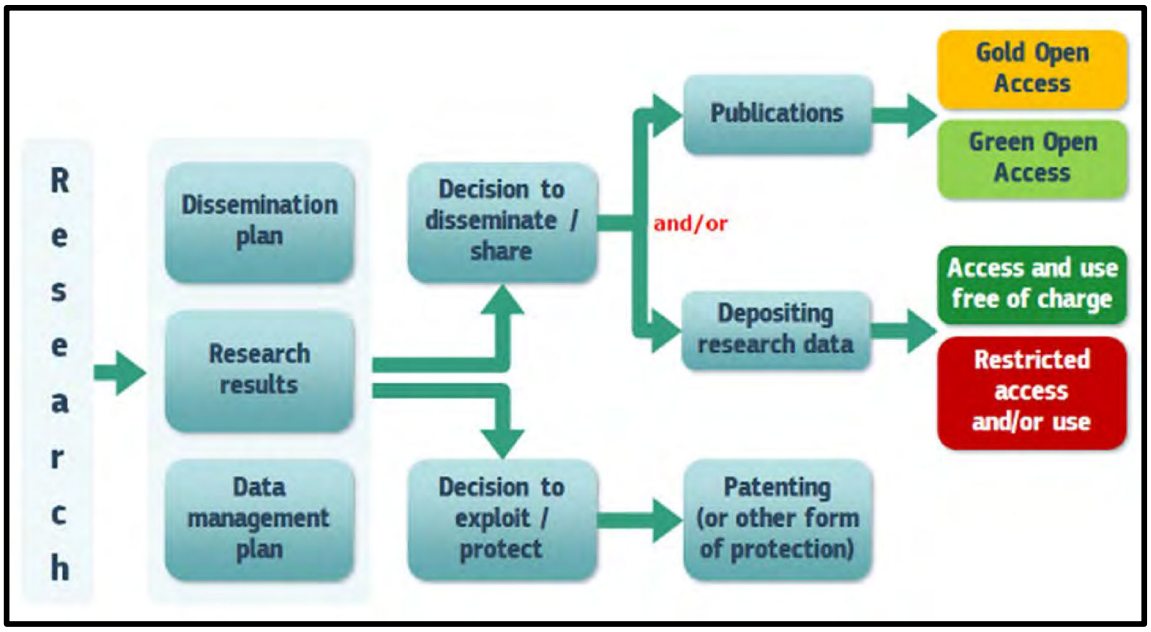


Figure 2.1. Flow process of Open Access to Scholarly Publications and Research Data
Source: Horizon Online Manual 2020, European Commission

2.2. PRINCIPLES

Though Open Science is based on 4 broad principles of FAIR, the Guidelines spelt out here are based on a set of specific principles adopted and adapted for this purpose. These principles are:[4]

A. **Openness** – implies access should be on equal terms, easy, timely, user-friendly and preferably Internet-based.

[3] by many organizations in particular the European Commission on open access to scientific research publications and research data.

[4] The principles and Guidelines are based on OECD document: "OECD Principles and Guidelines for Access to Research Data from Public Funding." In January 2004, 30 OECD countries including China, Israel, Russia and South Africa agreed to adopt a "Declaration on Access to Research Data from Public Funding." In recognising the significance of access to research data, OECD was asked "to develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding to be endorsed by the OECD Council at a later stage". In October 2006 the OECD's Committee for Scientific and Technological Policy approved the principles and guidelines and was endorsed by the OECD Council on 14 December 2006.

B. **Flexibility -** taking into account of the VUCA (vulnerability, uncertainty, complexity and ambiguity) local, regional and global events or situations on Open Science and tailor-made based on mutual "needs and offerings" of parties involved.

C. **Transparency –** means availability of data in a transparent way, preferably via Internet. It includes research data are easily findable and accessible via Internet besides public research entities are actively disseminating information on research data to individual researchers, academic associations, universities and other stakeholders.

D**. Legal and moral conformity** - Data access should conform to legal and moral rights of all stakeholders with some restriction to access of research data on the basis national security (data intelligence, military activities etc), privacy and confidentiality (data on human subjects and other personal data are subject to under national privacy laws), trade secrets and intellectual property rights, protection of rare, threatened or endangered species for protection and conservation and legal data under legal actions.

E. **Protection of intellectual property** - Data access should not violate copyright or of other intellectual property laws relevant to publicly funded research databases.

F. **Formal responsibility** – implies that access to data should be formalised in terms of institutional practices on data-related activities such as "authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving."

G. **Professionalism** relates to management of research data based on the professional standards and values embodied in the codes of conduct of the scientific communities involved such as the use of codes of conduct for professional scientists and their communities, mutual trust between relevant parties (researchers, institutions and other stakeholders).

H. **Interoperability** – one of the main principles of FAIR which relates to "technological and semantic interoperability" and considered as a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. Interoperability means that standards must be clearly described.

I. **Quality** involves "value and utility of research data" which is highly dependent on the quality of the data itself. Ensuring compliance to quality standards should be adhered to by data managers, and data collection organisations, if available, since "universal data quality standards" are occasionally not useful or realistic. However, measures should be taken of ensuring good practices adhered to (such as methods, techniques and instruments employed in data collection, dissemination and accessible) in "safeguarding quality and authenticity" in terms of origin of sources.

J. **Security** implies the guarantee of integrity and security of research data. Factors like completeness of data and absence of errors affect integrity. As for security, data protection through "intentional or unintentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols" should be given strategic priority apart from data storage sets and equipment be safeguarded "from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges."

K. **Efficiency** is one key goal in promoting data access and sharing "to improve the overall efficiency of publicly funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts." Some considerations be given to ensure its cost effectiveness in retaining data through cost-benefit assessments conducted periodically to ensure that the data sets with the greatest potential utility are preserved and made accessible.

L. **Accountability** implies that data access is subjected to "periodic evaluation by user groups, responsible institutions and research funding agencies" since such evaluations will assist to step up the support of open access among the relevant stakeholders.

M. **Sustainability** relates to long term retention to access of publicly funded research data. Given that most research projects, and the public funding provided are limited in terms of duration it can be very challenging of ensuring long term access to the data produced. The best the research funding agencies and research institutions could do in long term preservation of data is, at the outset of each new project to determine the most appropriate archival facilities for the data.

2.3. DATA SHARING THROUGH ACCESS OF SCHOLARLY PUBLICATION

What does Open Access to scholarly publications mean? Open Access to pre-processed and processed research results allows materials to be found or obtained via an internet search and be made available free of charge and free for further reuse. In providing accessibility, however, there are different routes of providing access to scientific articles, and they can be categorised as gold, green, diamond, and hybrid and bronze publishing. As for the term 'gold', it can be applied in different ways,[5] though in this Guidelines it is used as defined below:

> "1) the researcher publishes an article in an open access journal (gold);
>
> 2) the researcher publishes in a traditional subscription-based publication and thereafter a copy of the manuscript is published open access via a digital archive as soon as the publication permits this (green/self-archiving)
>
> 3) the researcher publishes the article in a traditional subscription-based publication and, for a fee, the article is made open access with immediate effect (hybrid). Researchers who publish their results in book form or in the form of an artistic work can also make available the material on an open access basis."[6]

2.3.1. Routes to Open Access to Scholarly Publications

There are 2 main routes to open access. They are:

a. **Self-archiving /'green' open access** – a version of an author's manuscript into a repository, making it freely accessible for everyone. The version can be deposited into repository and is dependent on the funder or publisher.

b. **Open access publishing /'gold' open access** - an author publishes article in an on-line open access journal. In this model, the payment of the publication costs is shifted away from the subscribing readers. The most common business model is based on on-off payments by authors. These costs, often referred to as Article Processing Charges (APCs) are usually borne by the researcher's university or research institute or the agency funding the research.

In other cases, the costs of open access publishing are covered by subsidies or other funding models.

[5] The EU Commission defines gold as "immediate open access that is provided by a publisher" a formulation that does not exclude hybrid publishing. http://ec.europa.eu/digital-agenda/en/open-access-scientific-information 27/11/2014.

[6] Proposal For National Guidelines For Open Access To Scientific Information, Vetenskapsrådets Rapporter 2015.

2.3.2. Open Access to Scholarly Publication and Scientific Records

1. The Guidelines requires that a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication of all peer reviewed publications produced as a result of research supported, either in entirety or in part by a Research Funder Organisation, is deposited in a suitable Open Access repository. Deposit should be made immediately upon acceptance for publication and the metadata made fully open, searchable and machine-readable from the time of deposit. This step also applies in the case of Open Access publishing ("Gold Open Access").

2. In the case of "Green Open Access", the Guidelines requires that the full-text of all such publications be made available under a standard open license immediately where possible and in any case no later than 6-months after publication in Science, Technology, Engineering and Mathematics (STEM) or 12 months after publication in the Social Sciences and Humanities (SSH). If a journal's permitted embargo period is longer than these, authors should either negotiate with the publisher to retain the rights so as to comply with this policy, or find a journal that enables them to comply without the need for negotiation. The Guidelines encourages retaining of ownership of copyright and to licence to publishers only those rights necessary for publication. This is possible through the use of addenda to the publishing contract.

3. The Guidelines will recognize compliant journals as those that adhere to the above provisions. The hybrid model of publishing is not compliant with the above principles and related costs will not be considered as eligible.

4. The Guidelines will recognize Open Access publication fees such as Article Processing Charges (APCs) or Book Processing Charges (BPCs) as eligible costs according to the funding guidelines. For quality assurance purposes, eligible journal titles must be listed in standard directories like the Directory of Open Access Journals (DOAJ) or PubMed.

5. The Guidelines requires that funded publications must be made available under an open content license, such as Creative Commons (CC BY). In all cases, the license applied should fulfil the requirements defined by the Berlin Declaration[7].

6. While the dominant type of scientific publication is the journal article, grantees are strongly encouraged to provide Open Access to other types of publications such as monographs, book chapters, conference proceedings, grey literature, reports, etc.

2.4. OPEN ACCESS TO RESEARCH DATA

Open Science is commonly misunderstood that all data has to be open by default. However, there is a wide spectrum of data sharing practices from entirely making the data open, to restricted and closed access. In the cases of government funded projects, raw research data and datasets that belongs to public domains and are useful for collaborative solutions, such as COVID-19, must be made open and available, while also observing legal provisions for data sharing. For restricted or closed raw research datasets, the metadata should be published to indicate that a particular research has been done and the raw research datasets have been gathered and consolidated.

[7] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) https://openaccess.mpg.de/Berlin-Declaration.

2.4.1. Research Data Lifecycle

The research data lifecycle and its six stages (data acquisition, processing, analysis, curation, sharing and data reuse) involved in it is depicted by **Diagram 2.2**.

Data acquisition stage is where all the sensors, data streams, data repositories are connected for the use of researchers. Data processing involve computing platform or workbench for users to use or catalogue features that allow the researchers to choose a particular dataset and add to their data cut and send for data processing.[8] In the data analysis stage, the data scientist or analyst will draw a conclusion from the dataset. At all the three stages aforementioned, the project or the data is still active so the datasets will keep changing. These data must undergo curation process before it can be shared to others.

Data stewards play a crucial role in guiding the data originators to manage their data to assure the data is in good quality and preserved. From this stage onwards, the data is all set to be shared and reused by others. Here is where the data repositories or domain repositories whether owned by individual institution or shared repositories by an independent party come into place. For example, NCBI has its gene bank where all the gene datasets are uploaded here before paper publication. Each data repository allows the users to search the database, for instance government data can be accessed in 'data.gov.my' and all institutions data under MOSTI can be accessed in 'radars.mosti.gov.my'. However, the challenge appears when a user wants to search for a dataset across all these repositories under a same gateway. Hence, the moderator pointed out if we should have a registry to harvest the data from data repositories and put them into one place (data catalogue) and to be maintained, for instance by MOSP. Then, the discovery service will allow users to search for any data across all these repositories. At the final stage of the research data lifecycle, the data should be allowed to be reused.

[8] data cut feature in Edinburgh and whether to implement this data cut feature into MOSP is still questionable

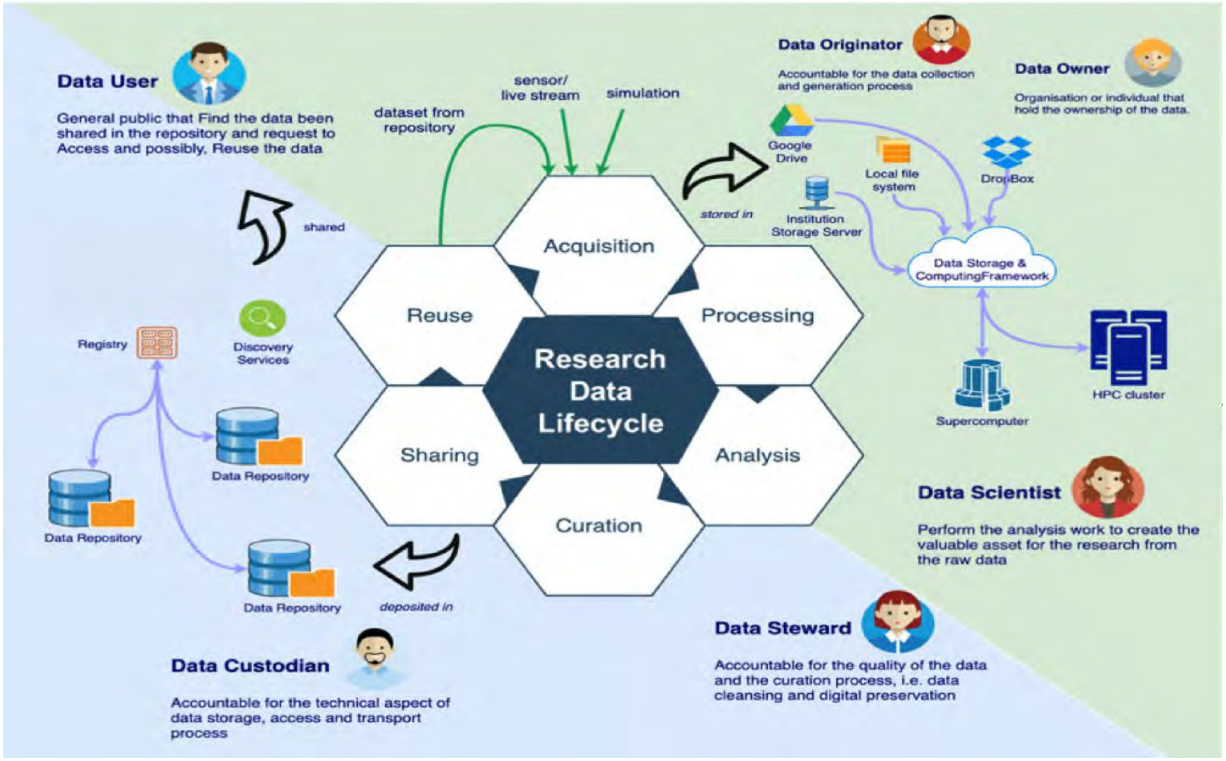


Figure 2.2. Research Data Lifecycle
Source: MOSP, 2020

2.4.2. Guidelines to Open Access to Research Data

1. Requires researchers to deposit the research data (which includes raw data) that were used and processed to yield the results that are published in scientific publications in institutional repositories. Research data should be assigned with persistent identifiers.

2. Requires that research data and services are handled according to FAIR principles (i.e. Findable, Accessible, Interoperable and Re-usable). Raw research data should also be traceable and whenever possible available for subsequent use.

3. The institutional repository follows the principle "as open as possible as closed as necessary". If data cannot be open due to legal, privacy or other concerns, this should be clearly explained. Metadata ensuring that data are findable should be provided in all instances.

4. Encourages the adoption of the MOSP requirements for monitoring of Open Science resources.

5. Requires researchers to submit an DMP to the appropriate service for every research activity they are involved in.

6. Requires researchers to define post-project usage rights through the assignment of appropriate licenses.

7. Requires that data are stored for a period as defined by the respective communities.

8. The minimum archive duration for research data is 10 years after the assignment of a persistent identifier. In the event that these records need to be deleted or destroyed after the expiration of the required archived duration or for legal and ethical reasons, such actions need to consider all legal and ethical perspectives.

9. All costs associated with the management of research data are considered eligible costs under the Guidelines. However, data management costs should be indicated or specified in the grant agreement application together with the data management plan created by the applicants.

CHAPTER 3

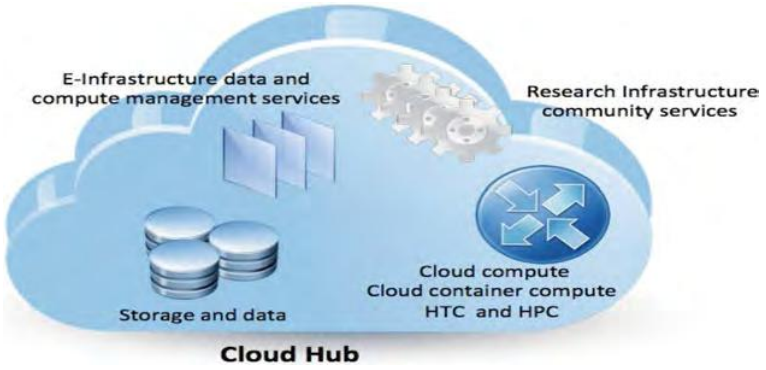
INFRASTRUCTURE ON OPEN SCIENCE IN PUBLIC FUNDED RESEARCH

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 link to multiple people and allowing them to work on the same task. These can range from extensive virtual research environments (VREs)[9] that facilitate sharing and collaboration via web forums and wikis, collaborative document hosting. The infrastructure of the European Open Science Cloud (EOSC) can be seen in Box 3.1. The infrastructure comprises of 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) that has established a in-built research data repository, which at the moment is still at a project-based level, but it 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 research data repository.

BOX 3.1. INFRASTRUCTURE ON OPEN SCIENCE



Source: ISSI Scientific Report Series Vol.15

Besides Open Data repositories, there are several efforts on-going to establish institutional research data repositories at these five research universities. In University of Malaya, the prototype of data repository is not ready yet since the policy is waiting for approval from the top management. However, in Universiti Teknologi Malaysia, the University has set up the prototype and has identified potential users for a pilot test. The Universiti Putra Malaysia has established an in-built research data repository, which at the moment is still at a project-based level, but it will transition to the institutional level in the near future.

[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 developed 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, University of Malaya is developing UM research data management policy and that Universiti Putra Malaysia (UPM) has also drafted 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 use the data, for what purposes the data is used, and approvals from relevant authorities to access data), 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, the importance of fundamental technical elements 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 the re-used data must be properly cited and acknowledge to 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[10]

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, and 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 research cycle, either the initial data, working data or final data stages. Researchers should plan at the start of a project for how they will store data, and to outline the budget of it. Such planning must be documented in a data management plan, and as described in Section 5.4.1 of Chapter 5.

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

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

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.

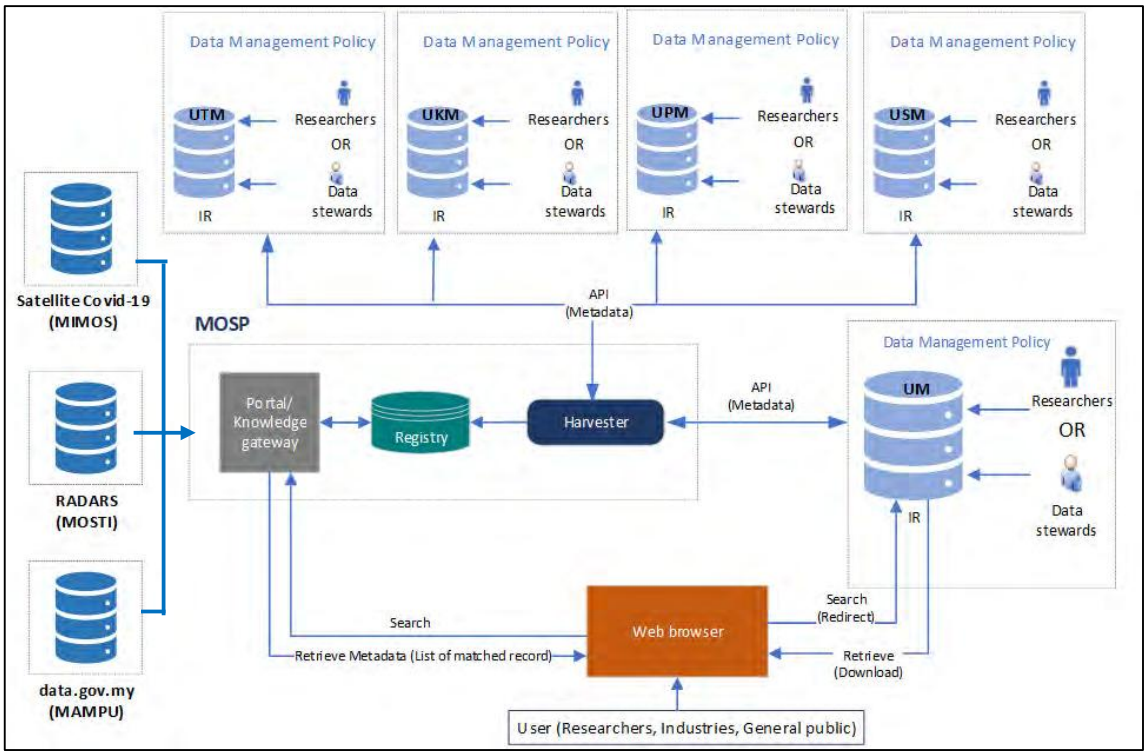


Figure 3.1: Architecture of the Malaysia Open Science Platform (MOSP) pilot project
Source: MOSP, 2020

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 to keep 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 metadata elements and will be captured by the MOSP Portal for each entry registered in the MOSP Portal.

[11]https://www.ands.org.au/guides/persistent-identifiers-expert

3.3.4. Publishing and sharing sensitive data[12]

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 which their 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. Anonymization 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.

[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 recognize 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 practicing 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 award 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 is 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. In terms of *who is responsible to address national incentives for MOSP,* MOSTI, MOHE or ASM, are 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 this 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 project name, and/or acronym, and/or number in the standardized prescribed manner [provide the standardized 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 [X years] following its adoption and subsequent reviews will take place on biennial basis. After that, the policy will be reviewed and updated every [X 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 servant, 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 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

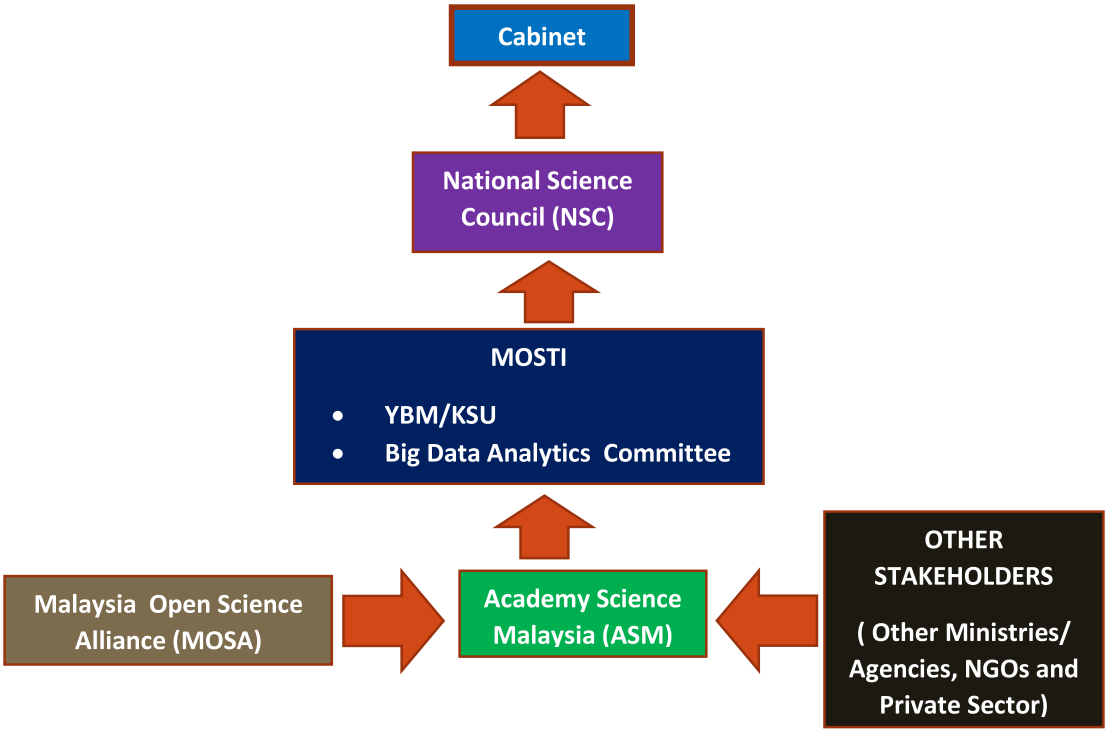
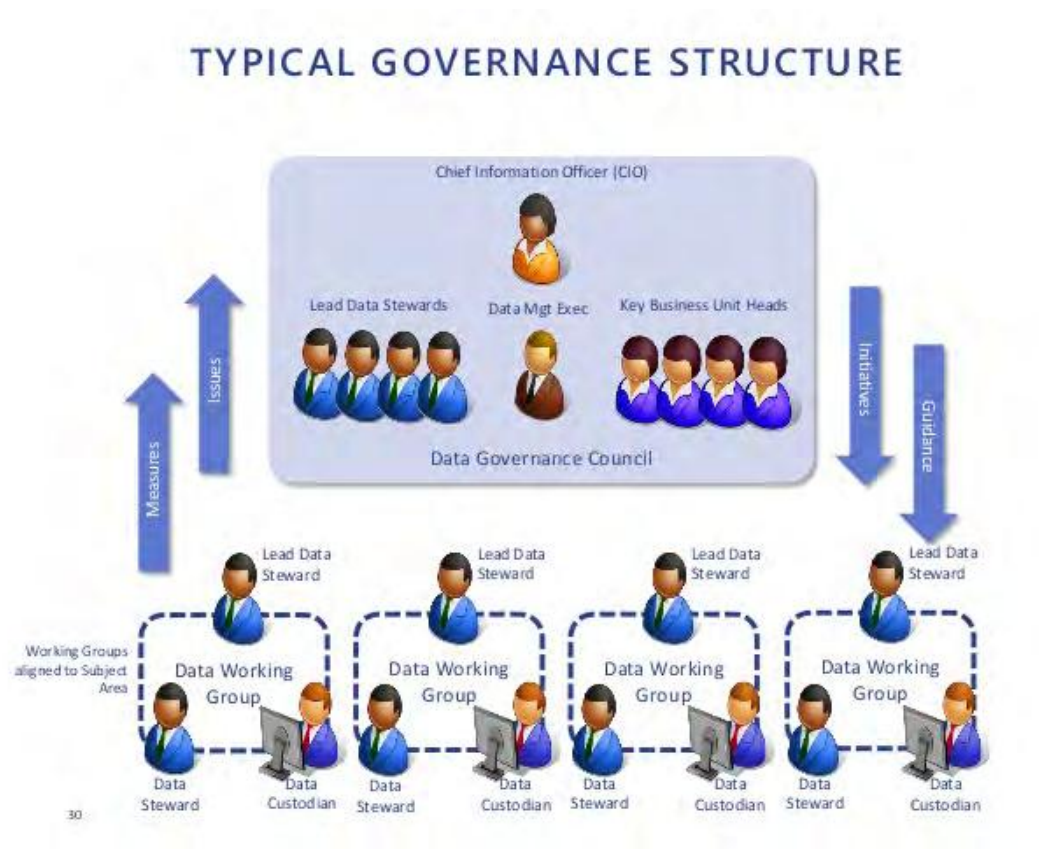


Figure 5.1. Governance of Open Science in Malaysian Public Funded Research

5.2.2. Institutional Level

At this institutional level it involves the governance of Open Science at institution of higher learnings (IHLs) and the public research institutions (PRIs). Each institution will already have the governing body which have adopted the best international practices in Open Science Management. These institutions will allow to continue adopting and practising 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 be as far as possible harmonised 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 kind of institutions are illustrated below.

a. University or Institution of Higher Learning

Figure 5.2. Governance of Open Science in Malaysian Higher Learning Institutes

b. Public Research Institute

Figure 5.3. Governance of Open Science in Malaysian 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 licenses 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 assumed the main role in defining international co-ordination, 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 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 an 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 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 to Open Science goal.
b) Classify the most important data governance.
c) Manage the content of scholarly publications and research data metadata.
d) Prioritize 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 organizing 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 organization 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 organization's strategic approach to plan for and 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 organization, 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 Represent the interest of the organization and deliver 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).

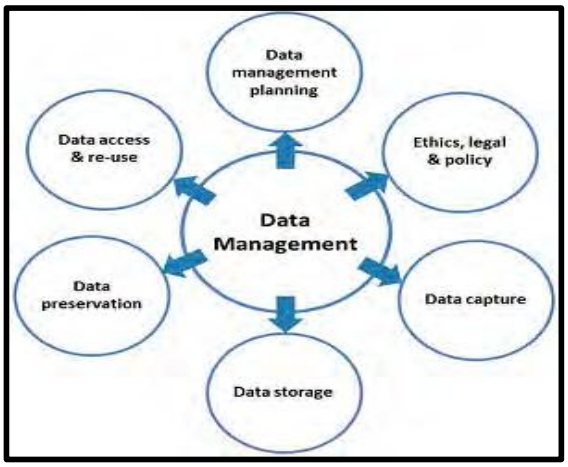


Figure 5.4. Data Management in Open Science
Source: https://dmp.qut.edu.au/faq (2014)

5.4.1. Data Management Plan (DMP)

a. What is 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 to make research data Findable, Accessible, Interoperable and Reusable (FAIR), an 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 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 to 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 National Guidelines on Open Science in Public Funded Research. Compliance with the National 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 National 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 maximize the potential for access to and re-use of research data that takes into account the balance 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.

Box 5.2.

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 AN 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 has 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 restriction 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 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 will the identity of the person accessing the data be ascertained
2.3. Making 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 re-usable	**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 widest reuse possible.
	• Specify when the research data will be made available for re-use. 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) is 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 subject, 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 re-usable
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 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, does backup will 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 organization with an external body or organization
	• 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 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 any other procedures for data management that will be used (if any)

5.4.2. Metadata Management[14]

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 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 are 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:
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.

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

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 this 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 own 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.

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 subject for all human data collected and must be anonymized 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 the access to the data.

ii. For human data collected from data subject 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 commercialization, it must be not be deposited until the patent has been filed.

iv. For data that concerns with 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 funder. 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 are 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 re-use

(a) If the data that will be reused is licensed, the conditions of that license 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 are given. Anyone may access the metadata free of charge.

g. Disposal

(a) Beyond the period of retention specified here, all research data must be disposed. Any destruction of the research record, either whole or part shall follow the National 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 (IHKL) will respectively be implementing Open Science Guidelines in their entities. For PRI, the Deputy Director General will be responsible to oversee and manage 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 the 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 National 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 this National 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 National 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 National 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, commercialization and innovation.

CONTACT INFORMATION

For queries relating to this document please contact:

Guidelines Owner	
All Enquiries Contact:	
Telephone:	
Email address:	

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 Guidelinesfor 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 Institute 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 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 backup system, scalable and interoperable to Malaysia Open Science Platform (MOSP) Portal	
6	Ensuring that the data classification for the data repository systems for which they have responsibility 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 develop Data Management Plan	
9	Formal training for Principal Investigators and researchers about key practices for Open Science and preparation to write 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 system, such as for security tenure and career promotion assessment	
13	Effective communication about Open Science and its incentives across all levels	
14	Review and evaluation to fill emerging gaps during implementation	

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 identifer before proceeds to publishing 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 acceptable to 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 an embargo period is 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 permit 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 be 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 with a pending patent. Special actions must be undertaken to treat such datasets. Clarify with a technology transfer office if the data is commercially exploitable.	

AKADEMI SAINS MALAYSIA

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