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 tools1. 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 with 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) Guidelines, (2) Capacity Building and Awareness and (3) Infrastructure. The development of a **Guidelines on Open Data Sharing in Research** is one such commitment which is in line with one of the strategies in the **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 THE GUIDELINES

There remain some concerns and challenges in Open Science, and the way they are being accessed needs to be addressed. Among the issues raised include the potential misuse of the data, which could result in 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 clarifies research data ownership, recognises the role of data stewards and data curators, provide rewards and incentives to data contributors as well as establish rigorous security and privacy standards for data sharing practices.

More importantly, for MOSP to be successful, each initiative must be organised holistically, integrated and coherent with the overall goal, understood across all levels especially by researchers, and involves everyone's participation. This is important to develop 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 provide 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 AND OPEN ACCESS

For the purpose of the 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 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 be subscribed and be subjected to the FAIR principle.

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 reuse quality, valuable, high integrity and responsible data, fuelling scientific progress to its fullest potential.

---

2 There are other main acceptable definitions of Open Science globally and are as follows:

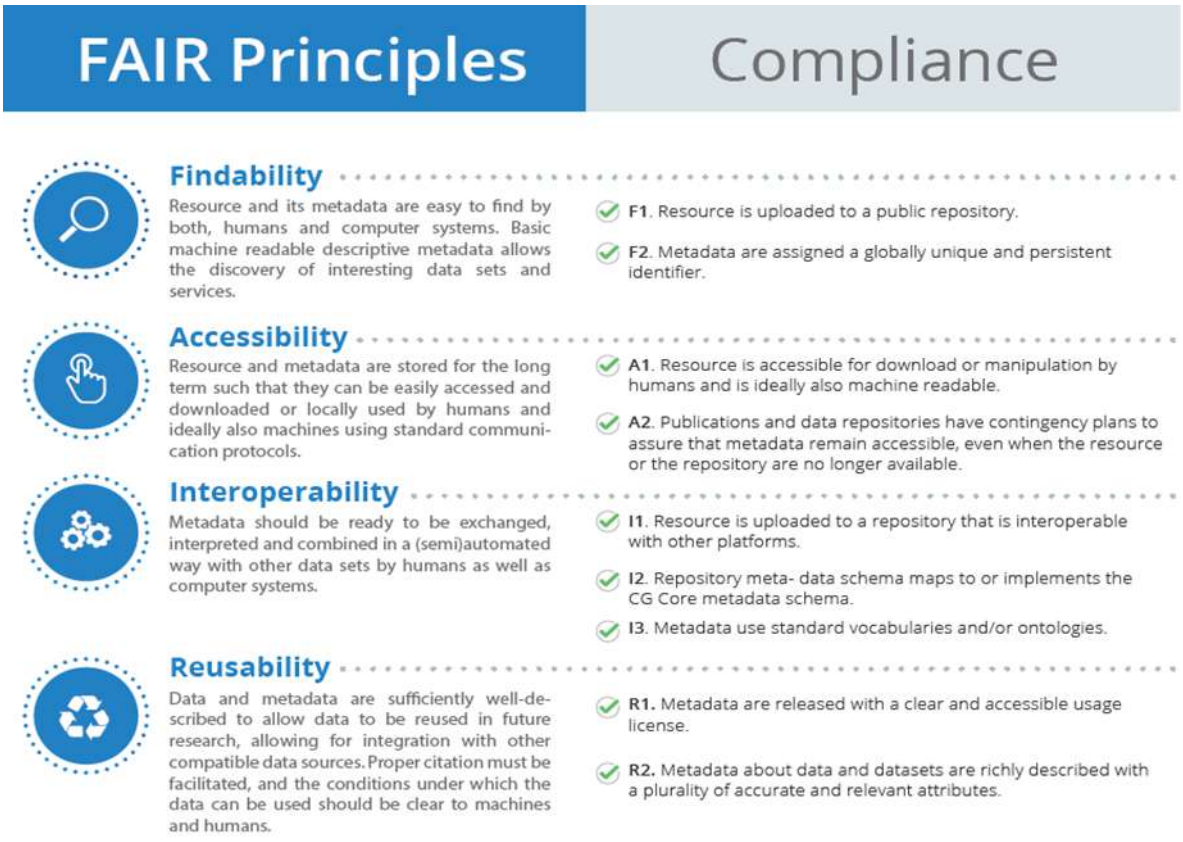
a) OECD2: "...efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as means for accelerating 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.".

- **Findable** means that data and metadata are easily found 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 THE GUIDELINES

a) The 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 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. The 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 are deposited at institutional repositories or data publishing partners' repositories and all metadata are permanently archived in the Malaysia Open Science Platform.

1.7. APPLICATION OF GUIDELINES

This Guidelines shall apply to:

(a) all staff, 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 the data life cycle- before, during and after. The 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 organisations 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 realisation 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 prerequisite in research grant agreements or spelt out in national strategies or institutional policy.

b) **Incentive mechanisms:** incentivising those that are 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 to "open access publishing or the release of data sets.". The incentive can also come in non-financial form such as awarding recognition or official acknowledgment (or even involving career advancement) to researchers and academicians who are strong advocates for 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 on open access to scientific research 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 needed to adopt together with the established rigorous security and privacy standards they need to conform or adhere to as part of the **mandatory rules** since open access does not mean "open to all" or that everything needs to be opened or disclosed.

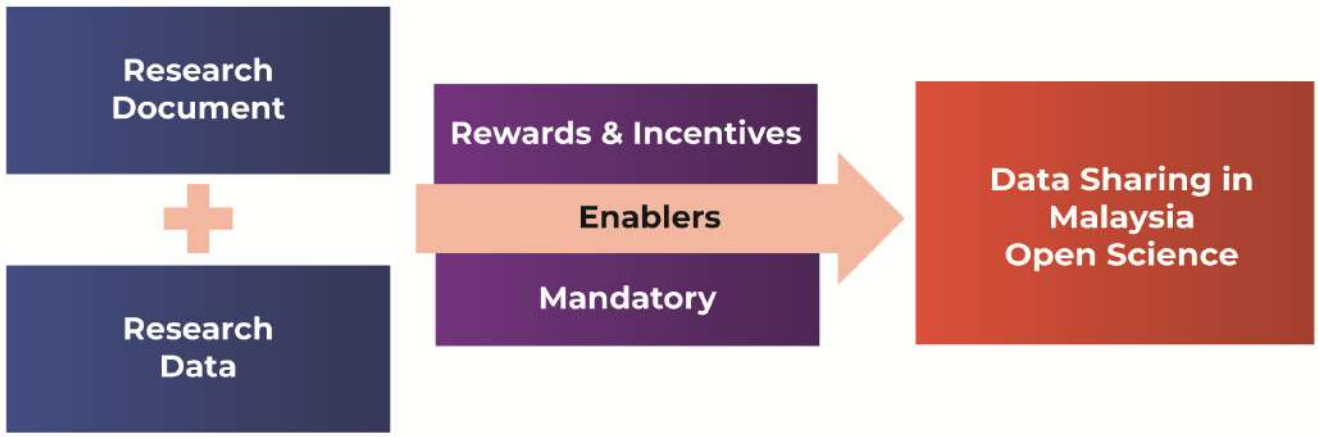


Figure 1.1. Data Sharing Framework in Open Science