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**. describes the general routes practised and adopted.[3]. The Guidelines will adopt a similar route in the open data sharing process.

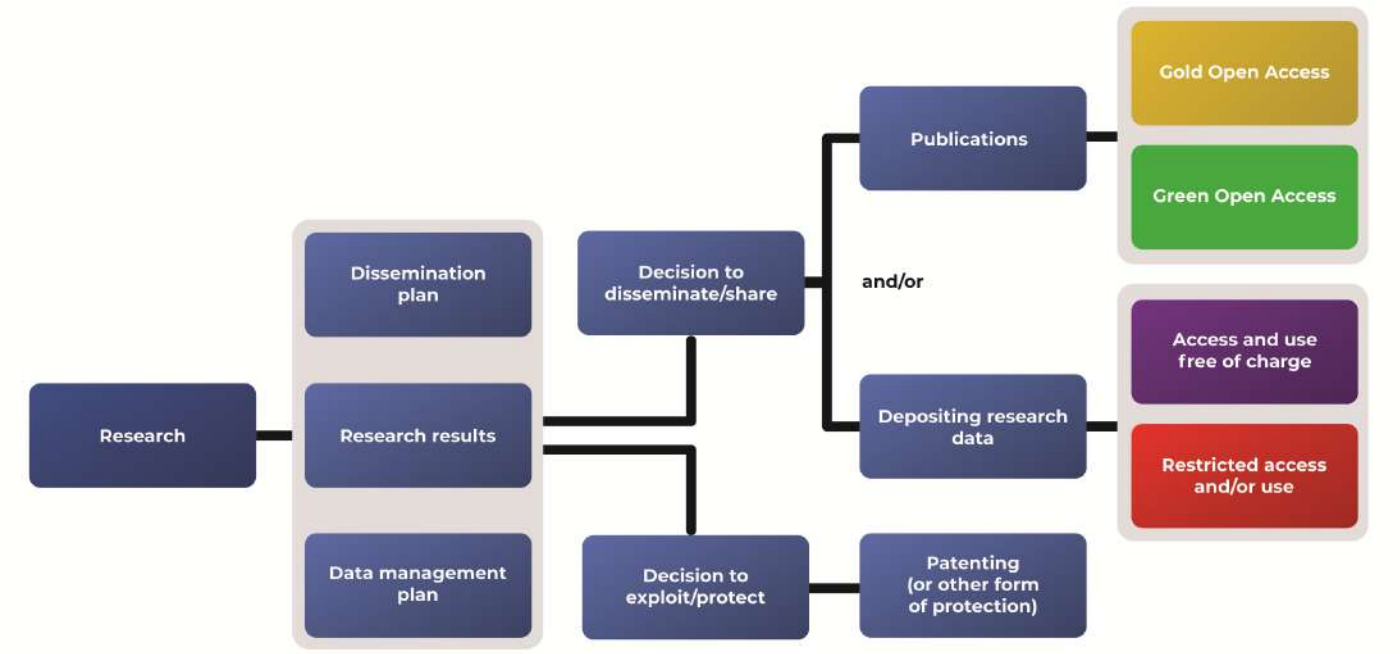

Figure 2.1. Flow process of Open Access to Scholarly Publications and Research Data
Source: Horizon Online Manual 2020, European Commission

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]:

[3] by many organisations 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 14th December 2006.

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

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

C. **Transparency** – means availability of data in a transparent way, preferably via the Internet. It includes research data that are easily findable and accessible via the Internet besides public research entities who 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 of national security (data intelligence, military activities etc), privacy and confidentiality (data on human subjects and other personal data that are subject to 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 is 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 to ensure good practices are 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 the safeguarding of data storage sets and equipment "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 are 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 to ensure 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 in 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 a repository and is dependent on the funder or publisher.

b. **Open access publishing/'gold' open access** - an author publishes an 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.

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").

---

[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.

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 licence 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 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 recognise 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 licence, such as Creative Commons (CC BY). In all cases, the licence applied should fulfil the requirements defined by the Berlin Declaration.

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 in the sense 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 belong 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.

2.4.1. Research Data Lifecycle

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

The data acquisition stage is where all the sensors, data streams, data repositories are connected for the use of researchers. Data processing involves 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. The data must undergo a curation process before it can be shared to others.

---

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

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

Data stewards play a crucial role in guiding the data originators in managing their data to assure the data is in good quality and is 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 institutions or shared repositories by an independent party come into place. For example, NCBI has its gene bank where all the gene datasets are uploaded 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 the 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. The discovery service will then 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.

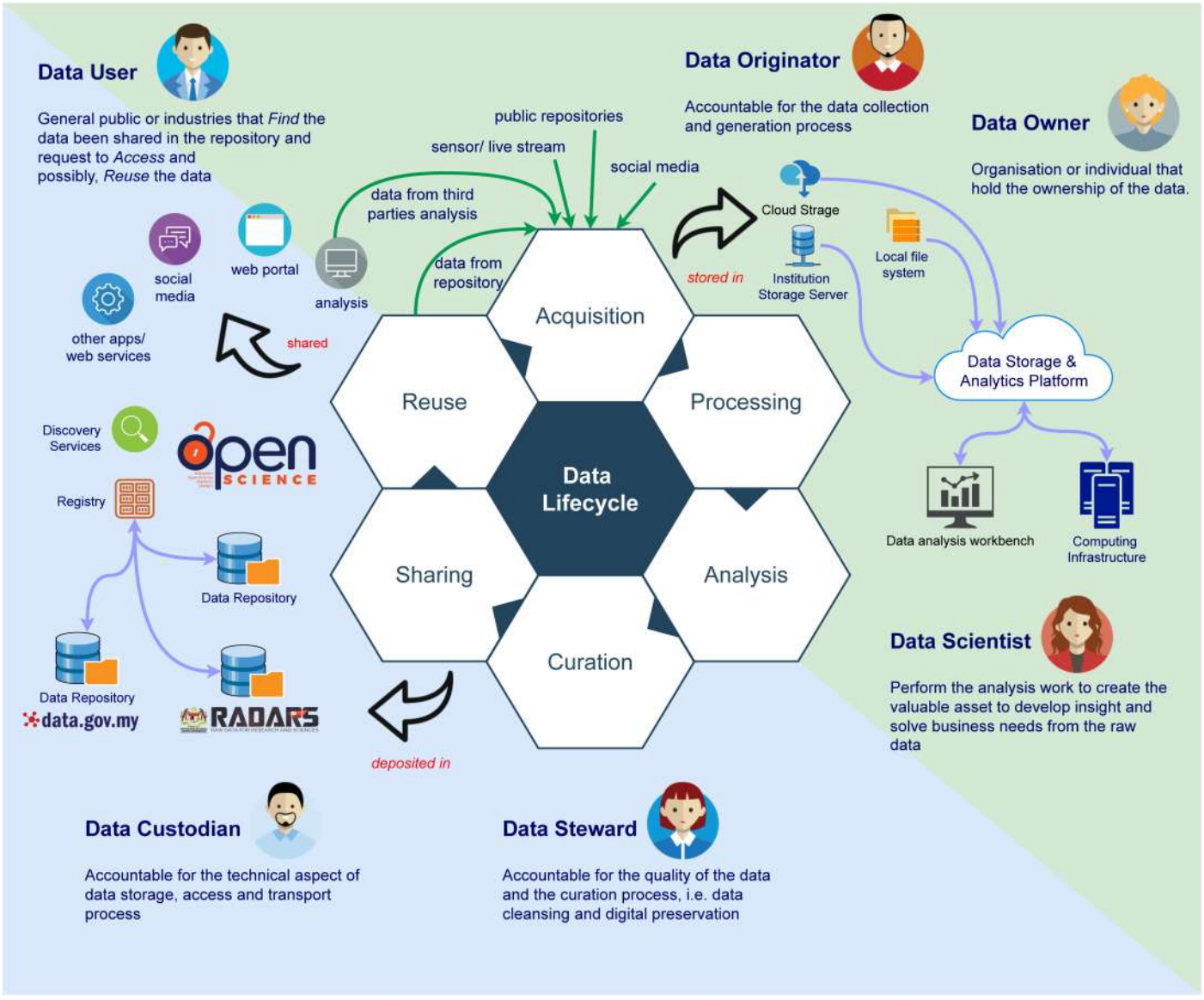

Figure 2.2. Research Data Lifecycle Source: MOSP, 2020

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 a 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 licences

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.