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

Open SCIENCE
Malaysia Open Science Platform (MOSP)

AKADEMI SAINS MALAYSIA

TRAINING OF TRAINERS
on Data Stewardship for Open Science 01/2020

TRAINING OF TRAINERS on Data Stewardship for Open Science 01/2020

Open Science is an initiative to make research output such as data and publications more transparent and accessible, 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.

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 is a two-year pilot project and was launched on 7 November 2019. The initial task for the Malaysia Open Science Alliance is to look into three main areas, which are (1) Policy and Guidelines, (2) Capacity Building and Awareness and (3) Infrastructure.

A key feature in steering the MOSP initiative is professionalizing the role of data stewardship. Data stewards must understand Open Science skills and be equipped with knowledge and expertise in data stewardship for Open Science. Hence, this Training of Trainers (ToT) on Open Science Skills and Data Stewardship is organized to train future data stewards who will have the competencies and expertise in data stewardship for Open Science.

The learning approach for this programme is based on andragogy assumptions that identify a set of characteristics of adult learners who define their self-motivation depending on their personality and use the existing experience as a resource to complement and construct their knowledge.

Motivation and readiness to learn is based on the immediate tasks and responsibilities at hand. From these assumptions, the Data Stewardship for Open Science learners must be involved in the planning and evaluation of their instruction. The courses' objectives are outcome-focused, and the learning activities and learning process are relevant and will have an impact on the learner's daily responsibilities.

PROGRAM LEARNING OUTCOME

i. Learners will have broad awareness of the need, complexity, and challenges associated with open science, modern science communication and data stewardship.

ii. Learners will demonstrate fundamental, theoretical and technical understanding of the principles of data stewardship in open science.

iii. Learners perform effectively as an individual and in a group, at all levels within the organisation and with the public, with leadership, entrepreneurial and managerial skills.

PART ONE:
Open Science Skills

The first part of this training program is **Open Science Skills** course, which covers seven modules:

1. Fundamentals of Open Science,
2. Research Governance and Research Data Management,
3. Open Data,
4. Open Metrics and Analytics,
5. Open Access in Scholarly Communication,
6. Responsible Open Science: Ethics and Integrity, and
7. Citizen Science.

The overall learning hours for Open Science Skills course is **40 hours** and each participant is required to complete the learning hours for them to be able to progress to the next course, Data Stewardship.

PART TWO:
Data Stewardship

The second part of this training program is **Data Stewardship** course, which covers five modules:

1. Data Stewardship Core
2. Data Quality for Data Stewards
3. Data Governance for Data Stewards
4. Metadata Management for Data Stewards
5. Master Data Management for Data Stewards

The overall learning hours for Data Stewardship course is **80 hours** and each participant is required to complete the learning hours.

PART ONE:
Open Science Skills

MODULE 1:
Fundamentals of Open Science

5 hours

Synopsis of Module Content

This module introduces participants to the fundamentals of Open Science, encompassing the history, the concept, definitions and pillars. This module also provides a baseline understanding of FAIR principle in data stewardship and introduces students to various global Open Science initiatives including in Malaysia as well as benefits of Open Science in multiple perspectives. It will look at how broad or narrow the concept of Open Science can be defined, what the spectrum of Open Science practices is that researchers can engage with, and what potential motivations and barriers might make it easier or harder for researchers to actually practice Open Science. Through this module, participants will appreciate Open Science as an emerging trend within a broad spectrum of research fields that is changing the way research is currently being conducted.

Learning Outcomes

1. Understand the history, concept, definitions, and pillars of Open Science.
2. Become familiar of the FAIR principle and the needs associated with data stewardship in Open Science.
3. Gain insight into global Open Science initiatives and its implementation including in Malaysia.
4. Explore impacts, benefits and challenges of Open Science for researchers, research projects, and society more broadly.

Pre-requisite knowledge

There is no pre-requisite knowledge for this module. However, basic knowledge and some familiarities with Open Science is an added advantage.

MODULE 2:
Research Governance and Research Data Management

5 hours

Synopsis of Module Content

This module aims to introduce participants to the concept of the research lifecycle and research governance. This will include key components of research processes and exhibit the overview of data governance that encourages the participants to appreciate the steps involved before, during and after a research is being conducted. The module will also explain on importance of research data management practices that helps to make data organized, transparent, and protected. Using some examples of established research data management practices in other countries, course participants will be able to model research data management practices for researchers in their respective institutions. We will also discuss the role of Data Management Plan t o encourage good data management practices and data sharing following Findable, Accessible, Interoperable, and Reusable (FAIR) principle. This will involve getting course participants to be familiar with the research data management framework and be able to develop data management plan (DMP) in consultation with researchers in their institutions.

Learning Outcomes

1. Appreciate research processes and research data lifecycle.
2. Model research data management practices for researchers in their institutions.
3. Develop data management plan in consultation with researchers in their institutions.

Pre-requisite knowledge

Completion of module Fundamentals of Open Science

MODULE 3:
Open Data

5 hours

Synopsis of Module Content

This module aims to provide an overview on the concept of Open Data; the level of openness and the open data model. Factors such as type of data to be shared and platforms for data sharing will also be explained. These factors will aid participants to identify ways to give the maximum impact from sharing data.

Open data and the sharing of raw research datasets are practiced within the FAIR principles where data are to be made Findable, Accessible, Interoperable and Reusable. The module looks closely at the degrees of openness and argue that data should be open and FAIR. Together with the open data model, it is more evident that sharing research data is our way forward for science.

A challenge for Open Data is getting buy-in from all stakeholders, who may understandably be concerned about data security and intellectual property rights. They may be skeptical about the benefits of data sharing. This module will provide participants an overview of relevant guidelines and legislations, mainly in Malaysia, that may have implications on data sharing practices and further, offer some insights into how to make data more open/ FAIR under existing guidelines and legislations.

Through this module, participants will be more familiar on what data can be shared, and where to share the data. Additionally, by identifying ways to get the maximum benefit from sharing research data, it will encourage research to be more transparent and accessible.

Learning Outcomes

1. Understand the main principles and ethical reasons for making data open or "FAIR"
2. Identify the degree of data openness and the open data model to be considered (to make data more open/FAIR)
3. Consider which data can be shared and the platform to share the data
4. Identify the ways to get the maximum impact from sharing research data

Pre-requisite knowledge

- Completion of Module Fundamentals of Open Science

MODULE 4:
Open Metrics and Analytics

6 hours

Synopsis of Module Content

This module covers the introduction of open scholarly metrics as an output and impact indicators whose data are open. The use 'open' in the sense of the Open definition and apply it to data, metrics, indicators, methods, software, and services. To achieve the objective of Open Science, scholarly metrics can be used to discover scholarly works and researchers, and to highlight ways of increasing their discoverability. Context, such as reader demographics, helps when evaluating scholarly works. Context can also aid stakeholders of the scientific enterprise, for example, to address the right target group in articles or to find the right publication outlet for the topic. Metadata and scholarly metrics are rich sources for the provision of context on works, publication venues, institutions and individuals. In addition, this module also incentivizes researchers to share scholarly works, promote achievements online and engage with audiences via other means than traditional scholarly media.

To develop the library services, the module will explain on how to join forces with stakeholders to provide services reusing existing resources, tools, methods & data and work with researchers to build awareness of benefits but educate about weaknesses of scholarly metrics. This module also focuses on the use of scholarly metrics in the context of research assessment and exploit rich network structures & implement bibliometric, Altmetrics and research impact methods to enable discovery of experts, potential collaborators and non-academic audiences.

Learning Outcomes

1. Describe the traditional and alternative tools used for measurement of scientific productivity and evaluation of research.
2. Discuss the transition from traditional citation-based indicators and platforms to altmetrics for research evaluation.
3. Analyse the impact, influence and cost effectiveness of scientific output using open platforms and visualisations.
4. Examine the innovation, creation and documentation of research metrics for Open Science.

Pre-requisite knowledge

- Completion of module Fundamentals of Open Science
- An interest and desire to learn about Metrics and Analytics
- Some general understanding of Data Analytics Principles

MODULE 5:
Open Access in Scholarly Communication

8 hours

Synopsis of Module Content

This module covers the ways in which open access scholarly communication is produced, disseminated, and evaluated in condition of open access models, complex cycle, tricky processes, different layers of hierarchy, and levels of participation. There are also topics regarding good practises in planning, processing, and archiving the scholarly materials on open access repositories. Furthermore, this course will touch on the issues with Open Access scholarly publishing from legal and ethical aspects.Overall, this course will emphasise the role of information professionals in open access scholarly communication in relation with the challenges and opportunities mentioned above.

Learning Outcomes

1. Explain the concept of Open Access and the need for Open Access in Scholarly Communication.
2. Differentiate the various routes of Open Access within a scholarly communication environment and the types of open access repositories and software.
3. Explain the issues with Open Access related to rights management, quality and trustworthiness.
4. Demonstrate the impact of Open Access within a scholarly communication environment

Pre-requisite knowledge

- Completion of module Fundamentals of Open Science
- Basics of scholarly publication
- Literacy in using paywall and open access digital tools, web services and online databases for scholarly publishing.

MODULE 6:
Responsible Open Science: Ethics and Integrity

6 hours

Synopsis of Module Content

Integrity is the basic moral attitude in Science. As research has become more complex, more collaborative, and more costly, issues concerning research integrity have become similarly complex, extensive, and important. The relationship between Open Science and ethical standards becomes increasingly complex and relevant in a context of data-intensive research. As Open Science (OS) aims to promote transparency and reproducibility of results by increasing and widening the diffusion of knowledge, it brings both opportunities and challenges with regard to research integrity.

Hence, in order to spread awareness among researchers and maximize the benefits of Open Science, this module aims to provide knowledge on responsible conduct of research in an Open Science environment. Data stewards will be trained to equip themselves with relevant knowledge on responsible Open Science, so they will be able to advice researchers, especially young researchers, on the ethical way of conducting a research.

The course participants will also be informed about education and training in Open Scholarship era as well as the responsible conduct of research in various focus areas, including authorship and publications, peer review, research data management, financial responsibilities and collaborative research.

Learning Outcomes

1. Examine the ethical, legal and social implications/challenges as well as the research integrity issues related to Open Science;
2. Understand the landscape of responsible conduct of research to promote research integrity;
3. Analyse the necessary elements to support the integration of research ethics and integrity as structural component of Open Science.

Pre-requisite knowledge

Completion of module Fundamentals of Open Science

MODULE 7:
Citizen Science

5 hours

Synopsis of Module Content

This module provides an appreciationof how Citizen Science changes the relationship and communication between science and society. It covers the history, the concept and the trends of Citizen Science, as well as interrelations between Citizen Science and Open Science. This course also offers insights into the importance and benefits of Citizen Science in many aspects including the opportunity to stimulate knowledge sharing between science and society, and the specific roles of society and their motivation in contributing towards advances in modern research across the world. Besides, this module also aims to showcase several international and national Citizen Science projects of various themes - from outer space to biodiversity, ecology, climate and down under the ocean, and highlight key sustainability issues for a Citizen Science-driven project to function (e.g. recruiting and maintain volunteers or securing research grants) and how the project management team can involve local people in a particular form of Citizen Science project (e.g. who is involved/ what types of people involved, what is their motivation to get involved, and how do they get involved). This module also provides an introduction to data management as it pertains to data stewardship in Citizen Science projects. As an overall, this course aims to spark interests and stimulate thoughts among course participants about their contribution as a citizen scientist.

Learning Outcomes

1. Understand the history, concept and trends of Citizen Science within the context of Open Science
2. Analyse the role of Citizen Science in digital era and participants motivation to get involved.
3. Identify Citizen Science projects and the importance of data management and IT platforms in Citizen Science projects.
4. Develop an evaluation plan in Citizen Science and public engagement activities

Pre-requisite knowledge

- Completion of module Fundamentals of Open Science.
- Basic knowledge, prior experience and some familiarities in public-driven project management is added advantage.

PART TWO:
Data Stewardship

MODULE 1:
Data Stewardship Core

0.4 credit (16 hours)

Synopsis of Module Content

This module builds a foundation of knowledge for data stewards, or information professionals who want to become data stewards. Data stewards are important leaders in a company's information management program and they are accountable for the data strategy, definition, requirements, and quality of the data. To be effective in their duties, data stewards must understand how the data is created, stored, manipulated, moved about, and used. This module covers fundamentals of data stewardship: who are the data stewards, what they do, what are their responsibilities, and what are the key principles and practices of data stewardship. It also provides foundational knowledge of information management.

Learning Outcomes

1. Explain the basic concepts, principles, and practices of data stewardship.
2. Describe the roles and responsibilities of data steward and how to perform as one.
3. Be acquainted with information management processes and disciplines.
4. Familiarize oneself with data quality, data governance, and metadata management.

Pre-requisite knowledge

- Completion of Part 1: Open Science skills
- Interest and desire to learn about data stewardship

MODULE 2:
Data Quality for Data Stewards

0.4 credit (16 hours)

Synopsis of Module Content

This module introduces participants to the fundamentals of data quality. Data quality is one of the core responsibilities of data stewards, each steward needs a foundation of concepts, principles, terminology and problems related to data quality, as well as the key processes and projects of data quality management. This module provides an overview of the field of data quality with the goal of building strong fundamental knowledge for data stewards. It covers topics ranging from data quality definitions and dimensions, common causes of data quality problems, data quality assessment, root cause analysis, to data quality management practices and methodologies as well as core data quality processes and projects.

Learning Outcomes

1. Understand key concepts, principles, terminology and problems related to data quality.
2. Understand data quality management processes and projects.
3. Explain the importance of data quality assessment and how it can be applied to research projects.
4. Understand the principles of root causes analysis and data quality monitoring.
5. Apply ways to ensure data quality in the integration process

Pre-requisite knowledge

- An interest and desire to learn about data quality.
- Fundamental knowledge in database management system

MODULE 3:
Data Governance for Data Stewards

0.4 credit (16 hours)

Synopsis of Module Content
This module introduces participants to the meaning of data governance and why data governance is important. In addition, data stewards will learn on who is the drivers of data governance that will lead and guide the direction of data governance. This will lead them to understand the fundamental principles of selecting data and setting goals and then the standards, policies, processes, people and technology involve in data governance. This module also focuses on managing and measuring data governance such as measuring data governance effectiveness. In addition, designing data governance will be discussed in this module including the component of governance, the governance team and the different between data management and data governance in information management functions. Sustaining metrics is one of the important elements enclosed in this module including data metrics for big data and data in cloud. The presence of big data and cloud applications will bring some challenges to data governance in creating the policy and standard of external data, therefore this module will explain on how data governance acts as an emerging solution for this modern data sources.

Learning Outcomes
1. Understand the importance of data governance and the data that should be governed.
2. Explain the basic concepts, principles and practices of data governance.
3. Understand where and how to start a data governance programme.
4. Understand the processes, people and tools that enable effective data governance.
5. Explore techniques for measuring success of a data governance programme.

Pre-requisite knowledge
- An interest and desire to learn about data governance.
- Some general understanding of data management principles.

MODULE 4:
Metadata Management for Data Stewards

0.4 credit (16 hours)

Synopsis of Module Content

This module provides foundation knowledge about metadata management anddata modeling techniques for effective information management. Understanding data meaning, constraints and relationships, as well as using the essential tools to collect, record, and organize such information is essential to the success in data stewardship. To maximize the benefits of metadata management in data stewardship, participants will learn the core elements of describing data covers describing the, meaning, constraint and relationships of data; and the purposes of metadata as regards to classification, description, guidance, and control. This module also covers the necessary elements of creating data models and their process; and understand data naming, data definitions, and data structures, as well as the basics of entity-relationship and dimensional data modeling. The module highlights the importance of metadata in providing sufficient information and standardized descriptions to give the impact to the researchers in term of managing their research data and its improvement in visibility and research impact.

Learning Outcomes

1. Understand metadata and its place in the research lifecycle
2. Understand the core elements of describing data in terms of metadata types, fields, standards and its function
3. Identify the primary purposes of metadata as regards to classification, description, guidance, control and its relationship with data documentation
4. Analyse the necessary elements of creating data models and their process.

Pre-requisite knowledge

- Interest and desire to learn on metadata management
- Fundamental knowledge on metadata management

MODULE 5:

Module 5 : Master Data Management for Data Stewards

0.4 credit (16 hours)

Synopsis of Module Content

This module provides an overview of the field of master data management (MDM) with the goal of building strong fundamental knowledge. MDM is complex and challenging, but it causes good results in future when done well. The complexities of managing identities, managing hierarchies, and resolving conflicts among disparate data sources make MDM an ambitious undertaking. Add to these complexities the multi-faceted nature of MDM - with human, organizational, architectural, and technological implications - and it becomes clear that knowledge is an essential component of MDM success.

The module introduces participants to the core concepts, principles, terminology and methodology of master data management (MDM). Specifically, it covers the following topics: Sources and types of master data; Architectural options for MDM implementation; MDM challenges and best practices; Principles and techniques of data parsing, standardization, matching, and deduplication; and Finding and using external reference data This module also provides overviews and an understanding of the challenges of global data, ways to overcome them.

Learning Outcomes

1. Describe what Master Data is, why and how it must be managed,
2. Identify the styles and architectures used for Master Data Management projects.
3. Familiarize with the challenges and best practices in a different phases of Master Data Management projects.
4. Be acquainted with data parsing, standardization, data matching, and de-duplication.
5. Understand the challenges of working with global data and ways to overcome these challenges

Pre-requisite knowledge

- An interest and desire to learn about Master Data
- Some general understanding of data management principles

Check out our website for latest updates:
https://www.akademisains.gov.my/mosp/

Follow us on:

MOSP_Malaysia
Malaysia Open Science Platform
MOSP_Malaysia

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

03- 62036033 ext 172

Shyama Sundari Devi shyama@akademisains.gov.my /
Nur Hanisah Ismail hanisah@akademisains.gov.my