

Best Practices for developing a Data Management Plan

Faisal M. Fadlelmola^{1*}, Sumir Panji^{2*}, Nicki Tiffin³, Judit Kumuthini⁴, Nicola Mulder²

¹ Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Khartoum, Sudan 11115; ² Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa 7925; ³ Centre for Infectious Diseases Research in Africa and Computational Biology Division, Department of Integrative Biomedical Science, University of Cape Town, South Africa; ⁴ Centre for Proteomic and Genomic Research, Cape Town, South Africa 7925

*Equal contributors. Email: faisal.mohamed@hotmail.com

1

Background

Africa faces a huge burden of diseases caused by infectious agents, emerging viral diseases such as the recent Ebola outbreak [1]. In addition, the prevalence of non-communicable diseases in Africa is on the rise, in part due to increasing urbanization. Modern human populations originated in Africa, and as a consequence, Africans have extremely high genetic diversity [2], which may lead to huge variations in disease susceptibility or resistance, drug metabolism and prevalence of adverse side effects [3]. Therefore, genomic information could potentially have a greater impact on the prevention, diagnosis and treatment of diseases in Africa than in many other parts of the world. African scientists are continuously making efforts to perform large-scale genome-sequencing studies that are focused on specific diseases. The H3Africa project, funded by NIH and The Wellcome Trust, is already supporting several studies involving collaborative centers within the continent. Collectively, these studies could generate genome sequence data from 50,000 to 75,000 Africans [4].

During any research process you get to handle a substantial amount of data. In order to keep track of this data, organize and structure it well, a Data Management Plan (DMP) is essential. A DMP is a document you develop at the start of your research project. It tells the story of your research data. Benefits of writing a data management plan: 1) Comply with funder mandates; 2) Make research transparent (for instance for reproduction and validation purposes); 3) Enhance data FAIRness (findable, accessible, interoperable, reusable); Protect data subjects; 4) Comply with the General Data Protection Regulation (GDPR).

2

Objectives

The aim of this poster presentation is to provide H3Africa Principal Investigators and data managers with some insights on how to develop, expand, and mature their data management plans. Using best practices through all stages of working with data will ensure the accessibility and longevity of the data. The presentation also stressed on the relationship of the research life cycle to the data life cycle.

Methods

A DMP basically consists of 5 parts:

1. The **types of data to be collected or produced during the project**, and the **processes or methodology including the informed consent model under which data are collected**;
2. The **formats for the data and the standards to be used**;
3. The **availability of the data**, including information about **ways in which the data will be accessed**, and whether there are **any issues related to privacy and/or intellectual property**;
4. The **guidelines, procedures, or policies for data reuse and/or redistribution**, attribution, as well as for the creation of derivatives from the data and the eventual destruction of the data and data custodianship;
5. The **measures that will be taken to help ensure the long-term preservation of, and oversight of access to, the data** - including possible mention of factors such as **format migration** and **who will be responsible for managing the data** for the duration of the project.

3

Access to research data

A data management plan is a document that tells how a researcher will collect, document, describe, share, and preserve the data that will be generated as part of a project (see Figure 1).

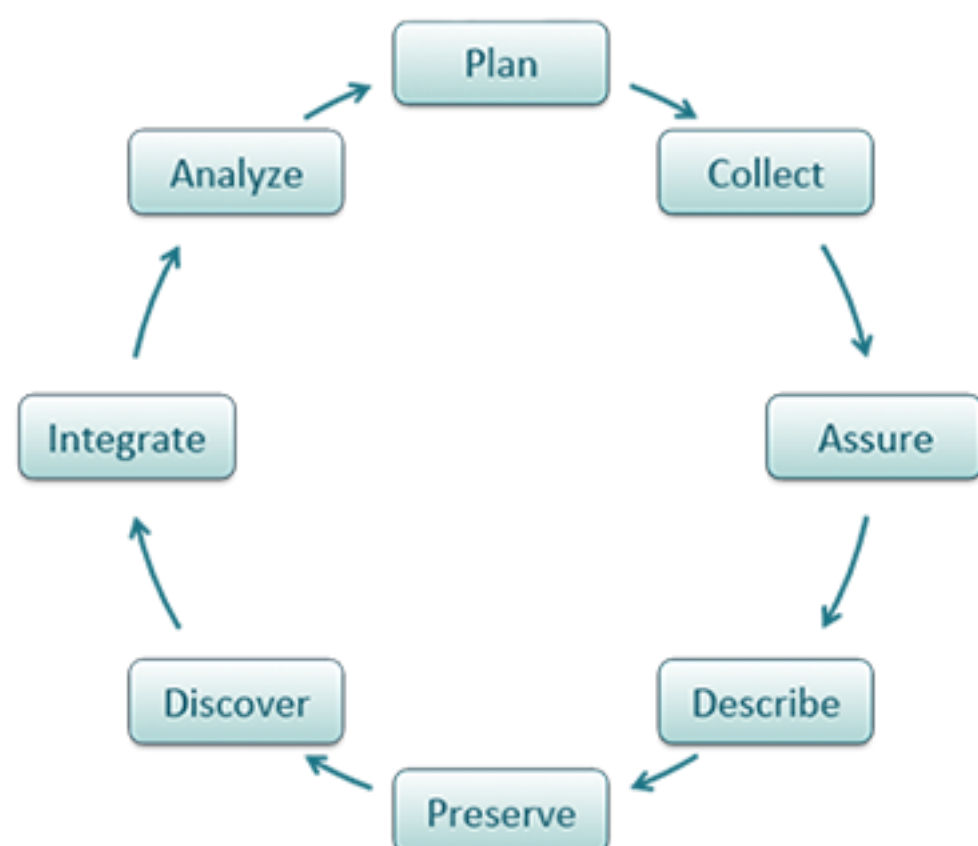


Figure 1: DataONE diagram of the data life cycle. Credit: [DataONE's "Best Practices"](#) [5].

Main Elements of DMP

In principle, projects participating in a research consortium are required to deposit the research data preferably into a research data repository. Projects must then take measures, to the extent possible, to enable for third parties to access, mine, exploit, reproduce and disseminate, free of charge, this research data. The guidance for DMPs calls for clarifications and analysis regarding the main elements of the data management policy within a project. The respective template identifies in brief the following five categories [6]:

1. **Data set reference and name**: an identifier for the data set; use of a standard identification mechanism to make the data and the associated software easily discoverable, readily located and identifiable.
2. **Data set description**: details describing the produced and/or collected data and associated software and accounting for their usability, documentation, reuse, assessment and integration (i.e., origin, nature, volume, usefulness, documentation/publications, similar data, etc.).
3. **Standards and metadata**: related standards employed or metadata prepared, including information about interoperability that allows for data exchange and compliance with related software or applications.
4. **Data sharing**: procedures and mechanisms enabling data access and sharing, including details about the type or repositories, modalities in which data are accessible, scope and licensing framework.
5. **Archiving and preservation (including storage and backup)**: procedures for long-term preservation of the data including details about storage, backup, potential associated costs, related metadata and documentation, etc.

Research and data life cycles

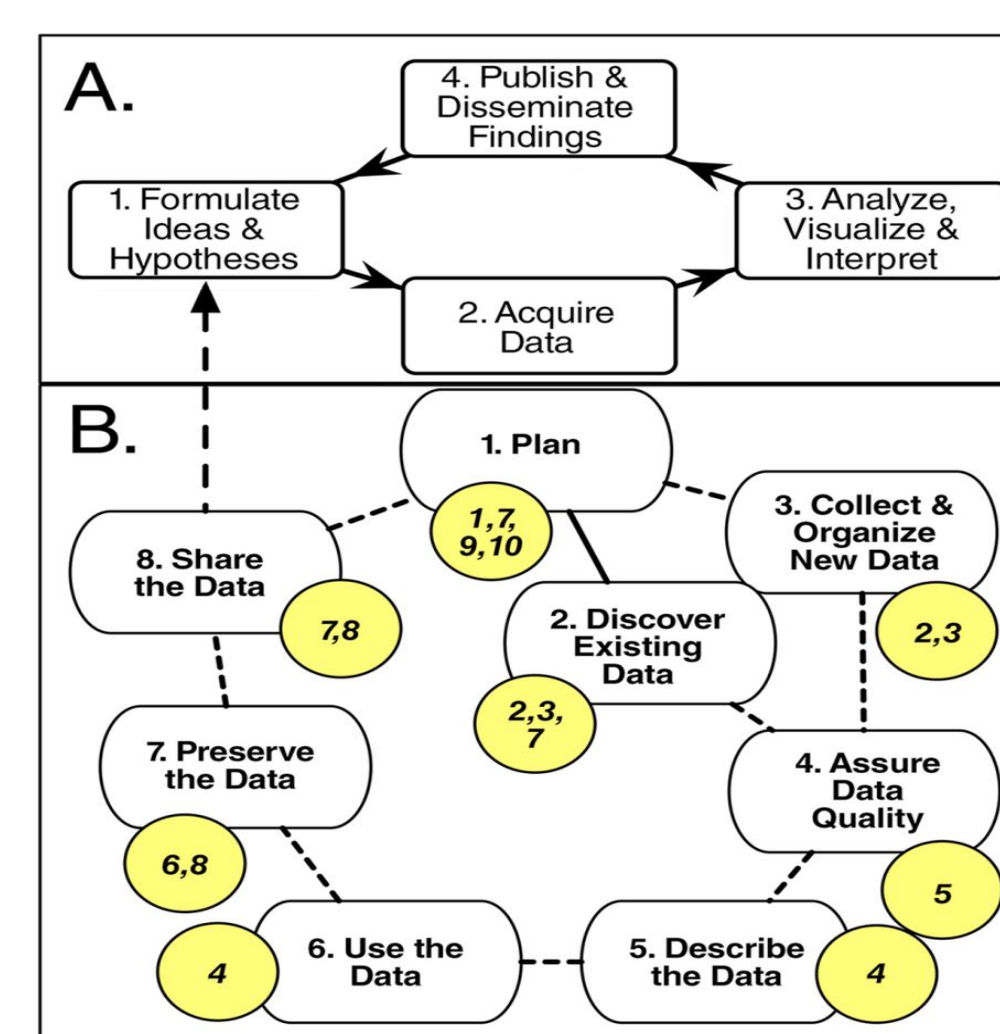


Figure 2: Relationship of the research life cycle (A) to the data life cycle (B); note: highlighted yellow circles refer to the rules (shown in Table 1 in more details) that are most closely linked to the steps of the data life cycle [7].

Why Manage Your Research Data?

Managing your research data will help you:

- Meet funding agency requirements;
 - Write more competitive grant applications;
 - Get credit for your data and increase its impact and visibility;
 - Encourage the discovery and use of your data to explore new research questions;
 - Improve your data's accuracy, completeness, and usability;
 - Ensure long-term preservation of data for future researchers;
 - Align/Comply with ethics and privacy policies.
- Add value to your research data.**

Table 1: Ten Simple Rules for Creating a DMP [7].

- Rule 1:** Determine the Research Sponsor Requirements
Rule 2: Identify the Data to Be Collected
Rule 3: Define How the Data Will Be Organized
Rule 4: Explain How the Data Will Be Documented
Rule 5: Describe How Data Quality Will Be Assured
Rule 6: Present a Sound Data Storage and Preservation Strategy
Rule 7: Define the Project's Data Policies
Rule 8: Describe How the Data Will Be Disseminated
Rule 9: Assign Roles and Responsibilities
Rule 10: Prepare a Realistic Budget

4

Best Practices in data management

- Writing a data management plan is increasingly seen as a key part of the academic research process.

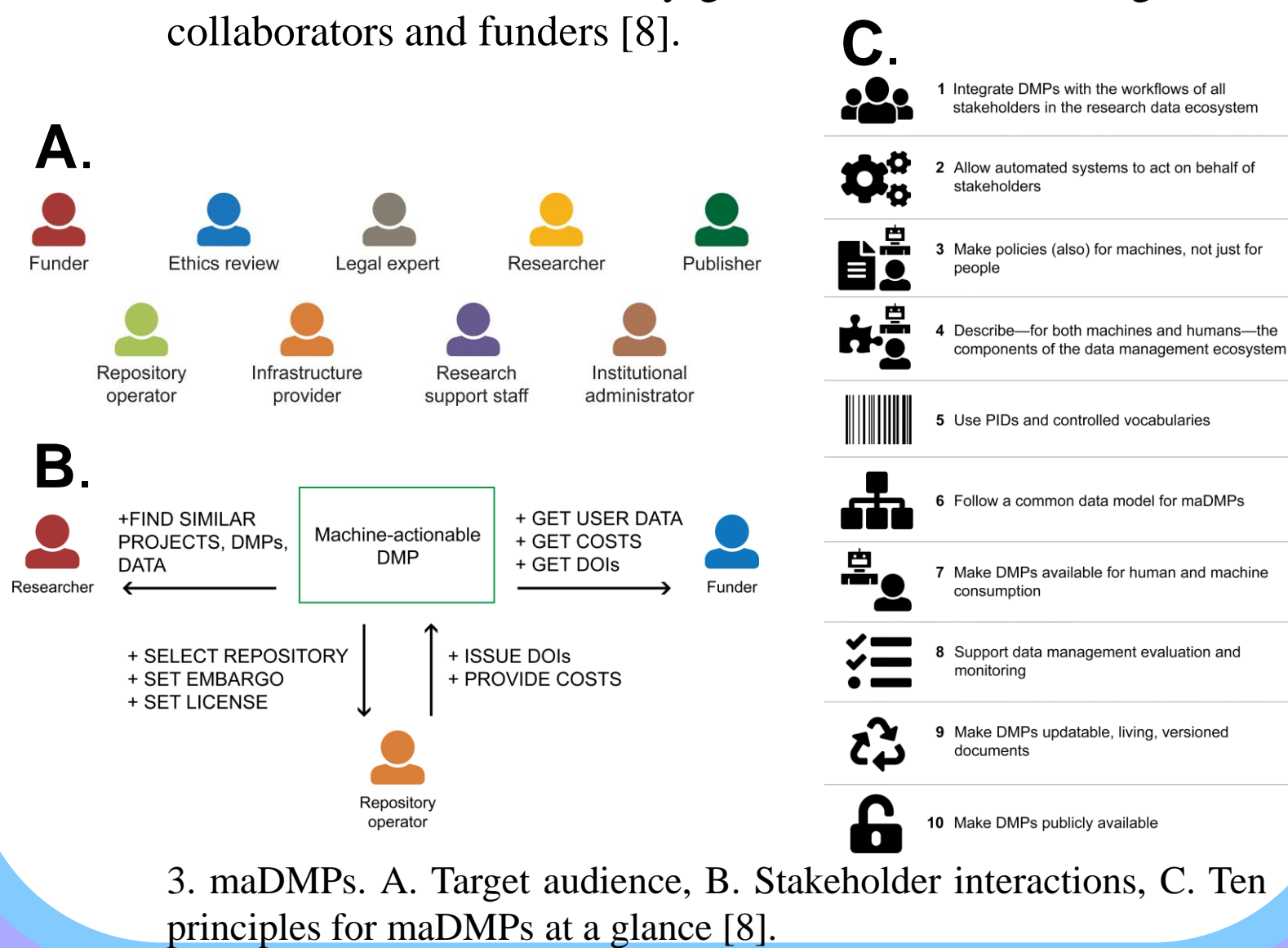
- The following practices are fundamental to effective data management and can be applied to all disciplines:

- **The Data Management Plan:** Adhere to the guidelines set by any funding agencies and institutions that are sponsoring the research.
- **Data Storage:** prevent data from being lost due to incompatibility.
- **Data Documentation:** Use metadata to record details about a study.
- **Ethical Issues:** How informed consent will be handled and how privacy will be protected.
- **Sharing Data:** Depositing of research data in an appropriate data archive is a much better option that helps ensure preservation and future discovery/use of the data.
- **Security:** To enforce the protection of confidential information, permissions, restrictions, and embargoes.
- **Responsibility:** Names of individuals responsible for data management in the research project.

5

Machine-actionable DMPs

Recently, Miksa et al. (2019) described 10 principles for machine-actionable DMPs (maDMPs), this will enable parts of the DMP to be automatically generated and shared, e.g., with collaborators and funders [8].



3. maDMPs. A. Target audience, B. Stakeholder interactions, C. Ten principles for maDMPs at a glance [8].

6

Conclusions

- A DMP is best prepared at the beginning of the research project. However, it is never too late to start mid-way during the research process - **better late than never**.
- A DMP is a **dynamic document** and should be reviewed regularly to update and improve it according to the actual needs as the project progresses.
- A DMP helps to ensure consistent practice in data handling among all project members regardless of future turnover.
- A data management plan should provide your project members, funders and others with an easy-to-follow road map that will guide and explain how data are treated throughout the life of the project and after the project is completed.

Literature cited

1. Tambo E. Need of surveillance response systems to combat Ebola outbreaks and other emerging infectious diseases in African countries. Infect Dis Poverty. 2014;3:1–7. doi: 10.1186/2049-9957-3-1.
2. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. Science. 2009;324:1035–1044. doi: 10.1126/science.1172257.
3. Bains RK. African variation at cytochrome P450 genes: evolutionary aspects and the implications for the treatment of infectious diseases. Evol Med Public Heal. 2013;118–34.
4. H3Africa Consortium. Research capacity. Enabling the genomic revolution in Africa. Science. 2014;20:1346–8.
5. The DataONE Best Practices. Available

- Online at: <https://www.dataone.org/best-practices>. Accessed Date: 31st March 2019.
6. Guidelines on FAIR Data Management in Horizon 2020. European Union. Available Online at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accessed Date: 31st March 2019.
 7. Michener WK (2015) Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
 8. Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>.

Useful DMPs Links

- NIH Data Sharing Policy: https://grants.nih.gov/grants/policy/data_sharing/
- WT Developing an outputs management plan: <https://wellcome.ac.uk/funding/guidance/developing-outputs-management-plan>
- NSF Data Sharing Policy: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>

Acknowledgement

Research reported in this study is supported by the National Institutes of Health Common Fund under grant number U24HG006941. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

