Research Data Management and Archiving (RDMA) Policy

Table of Contents

1   Introduction

1.1   Preamble

icipe is a **knowledge-based** scientific research institution, and data is one of our most important resources for research and decision making. The objective of the Centre is to generate data as information global public goods (GPGs)[1] in pursuit of its Mission This Research Data Management and Archiving (RDMA) policy supports *icipe's* Mission to help alleviate poverty, ensure food security and improve the overall health status of peoples of people in the tropics by developing and extending management tools and strategies for harmful and useful arthropods while preserving the natural resource base through research and capacity building.

1.2   Policy Definitions

A **Project** is a bounded component of the total institutional research and development plan, which: (a) has a timeframe, from conception to closure, (b) consumes and/or generates data during its life, (c) often has different actors producing/using complementary data that as a whole meets some research objectives and (d) has an identified leader responsible for it.

1.2.1 Research Data include primary or 'raw' data (recorded information, resulting from original observations and activities of a study) and derived data (data that were originally 'raw' but were then converted to another form using some manual or automated process). Forms of Research Data include (but are not limited to): information in laboratory notebooks; genomic data; field notes and journals; survey questionnaires; photographs; geospatial; video and audio tapes; data in computer storage devices; statistical package output files; technical reports; journal articles and publications.

1.2.2 A **Data Catalogue** is a compilation of data descriptions (i.e. meta-data) organised such that Research Data can be discovered by browsing or keyword searching. It is a coherent presentation of data that anybody could use to find out, with reasonable effort, what data there is and whether it is of interest to them. This also include persons who look for data outside their discipline, supporting more interdisciplinary / cross-disciplinary work.

1.2.3 Research Data Management and Archiving (RDMA) involves the planning, development, implementation and administration of systems for the acquisition, storage retrieval and archiving of data. It is a process that begins with conceptualization and design of a research project, and continues through data capture and analysis to publication, data archiving and sharing with the broader scientific community and the public.

1.2.4 Project Team is the designer and implementer of research data collection. They are responsible for data quality and the maintenance and update of all meta-data information relating to their Research Data.

---

[1] Information global public goods (GPG's) are all the data collected from research, all the consolidations made of them, conclusions drawn from them (research findings), and the various representations of this (raw data in databases, statistical analyses, GIS products, publications, etc.). They are of critical importance to other researchers and decision-makers. In the modern information field Information GPG's must have the following attributes: searchable repository (electronic); global (cross-border); open and easily accessible to all.

2 RDMA Policy

The goal of this RDMA policy is to maximize the value of *icipe* data and improve scientific transparency and research integrity with the aim of sparking awareness of needs and benefits of good research data management and to offer clear rules for managing research data.

2.1 Policy Statement

The data collected in research projects provides invaluable research tool to researchers desiring to conduct investigations in similar fields/areas. *icipe* recognises the need to be able to extract information from data generated over time and in the long term, to conduct previously unanticipated analyses and also be able to perform meta-analysis on new and stored data for new research insights. Electronic journals are rewarding the large efforts that are often needed to document and avail quality re-usable data to the public. In this regard, the RDMA policy is *icipe*'s response to demands by researchers, partners, publishers and diverse stakeholders to provide the appropriate information infrastructure (policies, guidelines, ICT connectivity) to derive the most value from our research data.

2.2 Scope of the Research Data Management and Archiving (RDMA) Policy

This policy applies to all data collected for the purposes of research, including the metadata that describes it and information derived from it. The policy applies to numerical and non-numerical data and applies whether the data are stored in soft copy format (e.g. electronically, digitally) and/or in hard copy (e.g. paper, original photos).

This policy applies to all staff, visiting scientists, students, trainees, temporary staff, partners and consultants engaged by *icipe*.

To ensure that the *icipe* RDMA policy standards are mainstreamed into joint projects, the following guidance should be adhered to: (i) If project data management are the full responsibility of *icipe* then this policy applies in its entirety; (ii) If the project partner is responsible for certain deliverables in joint implemented project, then project contractual agreement must specify the principles, that will guide the data management strategy; (iii) In the absence of any principles, then this policy applies.
In all instances, this policy should be followed except where partners and collaborators provide similar policies.

3 RDMA Principles

This policy subscribes to **open science principles**. In line with *icipe's* mission, the research output, data and publications, should be openly accessible. Specifically, the RDMA policy calls for adherence to a set of global data policy guiding principles to make data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**). The principles emphasise the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention.

- All **Research Data** belongs to *icipe*, as per the organization's intellectual property (IP) policy. The responsibility of ownership is to maximize the value of the data to *icipe's* mission.

- In line with the FAIR principles, research data and related materials, both digital and non-digital, must be accompanied by proper metadata and documentation in a way that facilitates the verification,

replication and, if possible, reuse of the data. This documentation must also contain information about property rights and terms of use.

- As a Centre of excellence working to build the capacity of researchers, the policy advocates for Open Source tools to benefit researchers and students.

- For **transferability of skills**, especially regarding research software, this RDMA policy advocates for the use of **open source software** that does not limit users. This requires for tools used to process, visualize, store and make data accessible to be accessible to *icipe* staff and stakeholders.

- The main driver for the implementation of RDMA policy is a detailed **Data Management Plan (DMP)**, developed at all project implementation stage in collaboration with the Data Management, Modeling and Geo-Information (DMMG) unit. In this regard the policy stipulates that:
    o Each project should develop a DMP
    o Each student and researcher should have a DMP
    o The DMP should use the templates and tools developed by the DMMG unit in collaboration with project leader.
    o The DMP should be reviewed by project team and DMMG unit and constantly adapted to comply with the Centre's Standard Operating Procedures (SOPs).

4 RDMA Policy Implementation

4.1 The Model

A two-level workflow approach shall be adopted, with some responsibilities handled at project level and major tasks accomplished conjointly by the DMMG and ICT units (Figure 1). These two components will be supported by tools, platforms, guidance and trainings, and the whole system monitored for compliance and performance by an audit function.

The '**Project**' is central to the operation of this policy. As used here, a project may correspond to a grant, or it may be some other component of the organisational structure. For the purpose of this policy, there are two basic requirements:
1. Anything designated as a **Project** has a recognised individual as leader, who is responsible for all aspects of it.
2. Every data-generating activity falls into at least one **Project**.

All projects should have a prior DMP and all data arising from the **Project** implementation should be linked to documentation that describes how the data were generated (tracing back to project documents), pre-processed and processed and how they were formatted for archiving and storage. Advantage of ICT tools shall be deployed in tracking of data

4.2 Data Custody – Taking care of the interests of Data Originators & Project Team

4.2.1   No *icipe* scientist has exclusive rights to the use of the data in their custody. However, their expectation of recognition for their role in generating the data and subsequent analysis and publishing is respected. Project leaders and the DMMG unit will ensure that data originators are given the first opportunity to extract maximum value from the data. This is especially important for project data collected by graduate students who have not yet completed thesis or published manuscripts. If data are anticipated to have multiple uses, the access and rights of each user must be agreed at the start of the project. The data access right should be included into the DMP.

4.2.2 When Project leaders decide that the institute's mission is best served by excluding a data originator from first publication they must (i) inform the data originator (ii) explain the reasons (iii) acknowledge the contributions of the data originator.
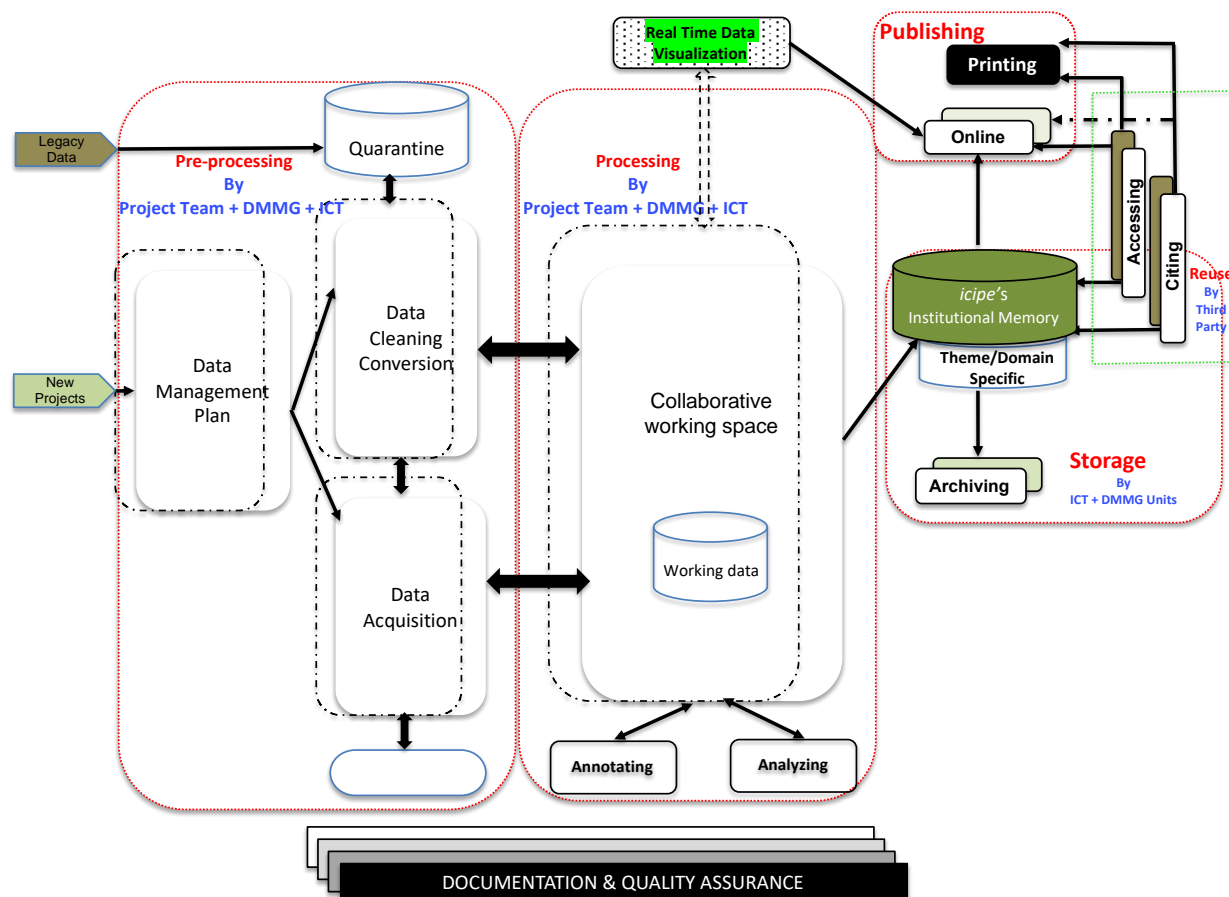
Figure 1. RDMA model – It displays 5 phases (Pre-processing, Processing, Publishing, Storage and Reuse) of the data workflow. Legacy data represent past data currently in people's storage devices, which need to be collated and temporally isolated (quarantine) for cleaning, formatting and digitization. For each new Project, the Pre-processing phase starts with the development of the DMP by the project team in collaboration with the DMMG unit, data collection uses data templates and digital tools obtained from DMMG and ICT units; the phase continues with data acquisition, data cleaning and data conversion. The data Processing phase includes the Annotation and Analysing of data; this is carried out in a collaborative working space by the project team, DMMG and ICT units. While the Processing of data is underway, the DMMG and ICT could apply analytics to create the "situation room" in which *icipe*'s work is captured, mapped and visualized in real time using graphs and numbers. Processed data are organized per themes and domains for Publishing and Storing. Archived data can be accessed and Reused by third parties.

4.3 Data Sharing

Data ownership must be stipulated in grants/partners agreements. Projects leaders and partners should jointly agree on modalities of data sharing prior to commencing work. Usually data ownership is captured under the IP section of Project Grant/Subgrant agreement.

4.4 Data Archiving

For the purpose of referencing, verification, replication and reuse, research data and related materials must be moved to storage for long-term retention after the research project. Within one month of the end of the project, and following the FAIR principles, the data should be archived for as longer as defined in agreements, unless legal or contractual regulations demand another term.

4.5 **Third Party Data**

Data protection and data security will be handled in line with the Centre's Data Protection Policy.
Data provided by a third party for use by *icipe* researcher are assumed to be owned by this third party, unless otherwise stipulated.
Data owned by a third party and requiring permission to reproduce will be marked as such in all *icipe* data products. Public access to jointly, or third party, owned data is dependent on the details of any grant/partnership agreement or a memorandum of understanding (MOU).

4.6 **Confidential Access**

Data must remain 'open as possible but as closed as necessary'. Project leaders in collaboration with a Data Management Specialist (DMS) at DMMG will choose one of the many available options for sharing data. Alternatives are providing only summary data; providing anonymised data; 'fuzzing' data (in case of Global Positioning System: GPS) and providing data with a material transfer agreement (MTA) that restricts further dissemination. DMS must ensure compliance with any national laws on personal or confidential data.

4.7 **Fraud**

It is a violation of the most fundamental aspect of the scientific research process to set forth measurements that have not, in fact been performed (fabrication) or to ignore or change relevant data that contradict the reported findings (falsification). Employees and any other person who may have become aware of data fraud must report the matter to the Director General, any member of the Senior Management Committee, the employee's immediate Supervisor , the Head of Human Resources or to any member of the Anti-Fraud, Harassment and Ethics Committee (AHEC) and/or through the completely anonymous offsite whistle blower service. All matters reported shall be investigated in line with laid down procedures to confirm the allegation prior to disciplinary action being taken.

4.8 **Roles and Responsibilities**

4.8.1 **All staff**, visiting scientists, students, trainees, temporary staff and consultants engaged by *icipe* should familiarize themselves with this Policy and the associated Research Ethics Policy, Intellectual Property Policy, Data Protection Policy and the respective project guidelines and procedures.

4.8.2 Management will provide support and leadership to ensure that the objectives of this policy is dispensed and ensure that projects will adequately include budget for data management during proposal development. In addition, management will ensure that skilled staff in the field of data management are recruited and provided with safe, secure and sustainable infrastructure and repositories to make data and information available, while respecting the rights of stakeholders in terms of confidentiality, intellectual property and data ownership.

4.8.3 The DMMG unit will provide guidance and guidelines relating to good research data management practice; support scientists and scholars in executing their responsibilities under this policy and will assist them in creating and implementing data management plans; develop SOPs and guidelines for managing data; promote the use of best practices for managing data. The unit will also establish capacity development programs on data management and the use of tools and platforms for processing, categorizing, classifying storing, archiving, publishing and retrieving of Research Data.

4.8.4 The ICT unit will provide leadership and strategic direction for the Centre-wide ICT infrastructure and services for data management; oversee the implementation and monitor the compliance with the policy including enforcing compliance to data policies, processes, standards and guidelines.

4.8.5 The **DMMG and ICT** units shall provide to projects tools and systems for electronic data collection, processing and archiving. Further, **DMMG and ITC units** are responsible for the security of the data.

4.8.6 The Heads of Themes will ensure that DMP for projects are developed and implemented in line with this policy and Standard Operating Procedures (SOPs); create awareness among staff and students of their responsibility and obligation in data management and help to identify trainings and skills required for good data management practice.

4.8.7 Project Leaders shall include cost and time implications of data storage and management in grant proposals; develop DMP that addresses the creation, management, storage and publication of data and the production of descriptive metadata; ensure that Research Data are deposited and preserved in appropriate repository, unless specified otherwise in the DMP. Project leaders shall inform students and scientists of the Centre's Policy and mentor them regarding good practice in data management.

4.8.8 Project Team in collaboration with the DMMG unit, are responsible for data quality, develop data templates, DMP, maintain and update all metadata information relating to the Research Data.

4.8.9 Capacity Building and Institutional Development Unit in collaboration with the DMMG unit will develop staff skills on data management competencies.

4.8.10 Human Resources Unit will ensure that competencies, skills and attributes related to data management are included in relevant staff job description and performance assessment; ensure that custodial responsibilities for data are included in the staff clearance handover notes to ensure continuity, smooth transition and operations.

5. APPENDIX 1: FAIR PRINCIPLES

(Excerpt from: https://www.force11.org/group/fairgroup/fairprinciples)

5.1   Preamble

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Here, we describe FAIR - a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. The term FAIR was launched at a Lorentz workshop in 2014, the resulting FAIR principles were published in 2016.

Based on these 15 principles, a set of 14 metrics have been defined to quantify levels of FAIRness. The latest developments on FAIR are available at GO-FAIR.

5.2   To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

5.3   To be Accessible:

A1   (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2   metadata are accessible, even when the data are no longer available.

5.4.   To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

4.5   To be Re-usable:

R1.   meta(data) have a plurality of accurate and relevant attributes.
R1.1 (meta)data are released with a clear and accessible data usage license.
R1.2 (meta)data are associated with their provenance.
R1.3 (meta)data meet domain-relevant community standards.