

Research Data Management WorkFlow

Henri Tonnang

The **PASET** Regional Scholarship
and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



1

Outline

- Introduction (Is Data the new oil?)
- Data knowledge areas
- Data Management Plan (DMP)
- Data Acquisition / Cleaning/ Conversion
- Data Analyses
- Publishing/ Storage / Reuse – FIAR Principle
- Data publication and citation support

The **PASET** Regional Scholarship
and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



2

Introduction

- In the 1980 most profitable firms were oil companies
- Currently the most most profitable firms are data oriented

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



3

Data knowledge areas

- ✓ Data engineering
- ✓ Data analytics
- ✓ Data Analysis
- ✓ Artificial Intelligence (AI)
- ✓ Machine learning
- ✓ Deep learning

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



4

Data knowledge areas

- ✓ Data engineering
- ✓ Data analytics
- ✓ Data Analysis
- ✓ Machine learning
- ✓ Deep learning

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



5

Data Management Plan (DMP)

When DMP is needed?

- A DMP is a formal document you develop at the start of your research project which outlines all aspects of your data (i.e. what you will do with your data during and after your research project)
- DMP is not a static document, but needs adjustment at regular intervals

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



6

Why manage data?

- Data is the new oil
- Make your research easier
- Stop yourself drawing in irrelevant stuff
- Save data for later
- Ovoid accusation of fraud or bad science
- Write a data paper
- Share your data for re-use
- Get credit for the data

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



7

Benefit of DMP for the institution

- Opportunity to engage with researchers and improve RDMA
- Raise awareness of support available
- Collate information to inform service delivery
- Ensure the institution is not exposed to risk
- Provide gratification and openness
- Ability to recovery cost via grants

The **PASET** Regional Scholarship and Innovation Fund

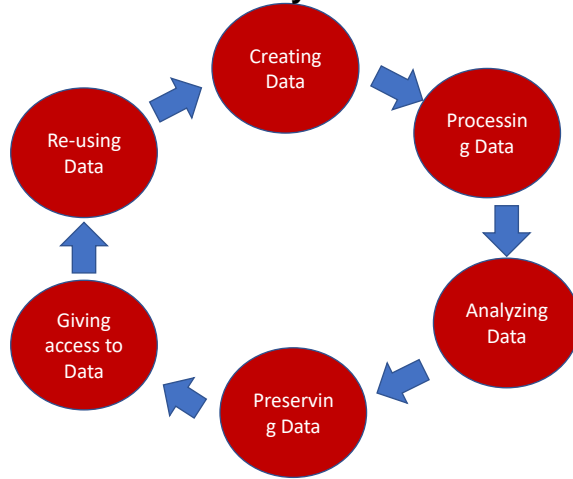


An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



8

Research data life cycle



The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



9

Tips on writing DMP

- Keep it simple, short and specific
- Seek advice –consult and collaborate
- Base plans on available skills and support
- Ensure the institution is not exposed to risk
- Make sure implementation is feasible
- Justify any resource or restrictions needed

The **PASET** Regional Scholarship and Innovation Fund



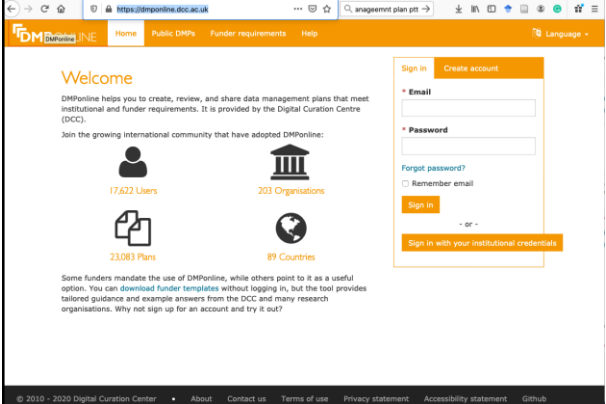
An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



10

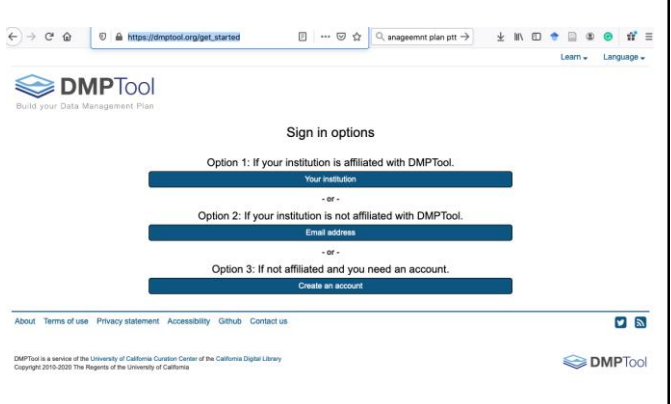
Tools for DMP

<https://dmponline.dcc.ac.uk/>




The screenshot shows the DMPonline homepage with a navigation bar, a 'Welcome' section, and a 'Sign in / Create account' form. The form includes fields for email and password, and options for 'Remember email' and 'Sign in with your institutional credentials'.

https://dmptool.org/get_started




The screenshot shows the 'Sign in options' page for DMPTool. It offers three options: 1) signing in with an institutional affiliation, 2) signing in with an email address if not affiliated, and 3) creating a new account if not affiliated and needing one.




The PASET Regional Scholarship and Innovation Fund

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



PASET
Partnership for Skills in Applied Sciences, Engineering & Technology



icipe

11

Data Acquisition / Cleaning/ Conversion

Project Folders are a way to organize the projects on your My Projects page by putting them into groups. You may create new folders below and then assign your projects to them on the right. A project can be assigned to multiple folders at the same time. To reorder your folders, you can drag and drop them in the left-hand table. Note: Only you can see your folders; they are for your own personal organization.

Illustration with RedCap

Digital Data Collection Tool

STEP 1: Create Folders


My Folders	New Folder	Add
Human Health		✎ ✖
Plant Health		✎ ✖
Environmental Theme		✎ ✖
Animal Health		✎ ✖
Molecular Biology, Bioinformatics and Biostatistics		✎ ✖
Social science		✎ ✖
Behavioural and chemical ecology		✎ ✖
Communication, Capacity Building, Delivery and Impact Assessment		✎ ✖

STEP 2: Assign Projects To Folders

--- Select a folder ---


Hide projects already assigned

Hide archived projects




The PASET Regional Scholarship and Innovation Fund

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



PASET
Partnership for Skills in Applied Sciences, Engineering & Technology



icipe

12

Data Acquisition / Cleaning/ Conversion

Illustration with RedCap Digital Data Collection Tool

Role name <small>(click role name to edit role)</small>	Username or users assigned to a role <small>(click username to edit or assign to role)</small>	Expiration <small>(click expiration to edit)</small>	Data Access Group <small>(click DAG to assign user)</small>	Project Design and Setup	User Rights	Data Access Groups	Data Export Tool	Reports & Report Builder	Graphical Data View & Stats	Calendar
—	aespira (Andrew Espira)	never	—	✓	✓	✓	Full Data Set	✓	✓	✓
Data manager	abneel (Abneel matharu)	never	Tunga_2020							
	amina (Amina Abubakar)	never	Tunga_2020							
	francis (Francis Mutebi)	never	Tunga_2020	✗	✗	✗	Full Data Set	✓	✓	✓
	htonnang (Henri Tonang)	never	—							
	ibrahim (Ibrahim kiche)	never	Tunga_2020							
PI	lynye (lynye elson)	never	Tunga_2020	✓	✗	✓	Full Data Set	✓	✓	✓
	ufilingler (Fillingler Ulrike)	never	Tunga_2020							

	No Access	Read Only	View & Edit	Edit survey responses
Demographics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Baseline Data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pain Management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Survey - PRE (survey)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Intervention Data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pain Management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Survey - POST (survey)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Study Completion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



13

Data Acquisition / Cleaning/ Conversion

SOLANUM SPECIES ON TETRANYCHUS EVANSI

Experiment site = icipe Duduville Campus

DATA COLLECTOR NAME/INITIALS: L.K. Murungi

Experimental year= 2009-2011

Treatments = Host plant species

Pest	Host Plant Species	Source
<i>Tetranychus evansi</i>	<i>S. sarrochoides</i> (GBK 028726)	Gene bank
	<i>Solanum villosum</i> Miller (MW 13)	AVRDC
	<i>Solanum scabrum</i> SS 52	AVRDC
	<i>Solanum americanum</i> Miller (SA)	AVRDC
	<i>Solanum tarderematum</i> Bitter (MW 03)	AVRDC

LEGEND(For each measurement made; use this convention)

spp = species
no_eggs = number of eggs
Gla_tricho = glandular trichomes
gla_tricho1 = glandular trichomes1
Non-gla_tri = Non-glandular trichomes
non-gla_tri1 = Non-glandular trichomes1
Rep = replicate

Navigation: Metadata | Protocol | Rawdata

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



14

Data Acquisition / Cleaning/ Conversion

Stage of Trial	Protocol	Notes
Pre-planting of host plant	Establish list of required host plant species and their source. The planting site (greenhouse) temperature should be monitored using a HOBO Pro Series Temp, relative humidity(RH) data logger. Average daily temperature should be 23C±1°C	
Seedling Transplant	Seedlings are transplanted into pots (29cm in diameter) filled with a mixture of soil, compost and sand (3:2:1, v/v) 28 days after sowing. Plants are watered daily and each pot is fertilized with 3g calcium ammonium nitrate (26% N) 2 weeks after transplanting.	
Mites Sourcing	Mites are sourced from a colony maintained on potted tomato plants (variety Moneymaker obtained from the East African Seed Company, Nairobi, Kenya), in a rearing room at a temperature of 25±1°C, 50-70% RH and 12h photoperiod.	
Trichome Identification and quantification	Twelve fully expanded young leaves are collected at random from 4-week-old plants of each respective nightshade species. Identification and classification of the trichomes are made based on the presence or absence of glands. Trichome counts are made under a 32x dissecting microscope fitted with a square grid to assist in counting. Ten squares (each 0.11mm ²) are selected at random on the abaxial surface of each leaf. Three replicates each with 36 leaves are carried out for respective plant species.	*Densities are expressed as the number of trichomes/mm ² .
Mite placement for fecundity	Four leaf disks (25mm in diameter) of the respective species are maintained individually in one Petri dish (86mm in diameter) stacked with cotton wool moistened with tap water and placed into plastic trays (36x 23x 2.3cm). A single female deutonymph and two males are carefully picked from the colony and transferred to the respective leaf disks for oviposition. These rearing units are placed in an incubator maintained at 25 ± 1°C, 70-80% RH and a 12h photoperiod. Males are removed 48h later, after the female has emerged. The number of eggs laid per female is monitored daily during the first 10 days of the oviposition period. The leaf disks are changed every 4 day	*Fecundity tests are carried out on leaf disks of the same age as leaves on which the number of trichomes is counted and representing each investigated species. *Five replicates, each with 20 deutonymphs, are evaluated for each plant species.
Evaluation of the effect of trichomes on mite movement	Evaluation should be quantified with a no-choice thumbtack bioassay. One leaf of each species is attached to a board of Styrofoam through a metallic thumbtack (9mm in diameter) placed at the centre of its abaxial surface. One replicate should consist of four leaves of each species individually placed on the Styrofoam boards. Ten female spider are transferred with a fine camel-hair brush to the head of each thumbtack. The trial is carried out on a laboratory bench at 23±1°C. Distances travelled by each mite onto the leaf surface are measured as the shortest distance (in cm) between the mite and the thumbtack edge, and are recorded after 15, 30, 45 and 60min. Three replicates, each with 40 spider mites, should be carried out for each plant species.	*Mites that stay on the thumbtack are considered to have travelled a distance equal to zero.

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



15

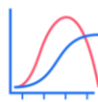
Data Analyses



Bar chart



Stacked bar chart



Line graph



Gantt chart



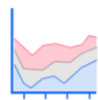
Polar area diagram



Scatter plot



Calendar heatmap



Stacked area chart



Sparkline



Column sparkline



The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



16

Publishing/ Storage / Reuse – FIAR Principle

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.

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.

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.

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

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



17

Data publication and citation support

The screenshot shows the Dataverse Project website at <https://dataverse.org/publications>. The page features a navigation menu with 'About', 'Community', 'Best Practices', 'Software', and 'Contact'. The main content area is titled 'Publications' and includes a 'Download Citations' link. The page lists several publications from 2020 and 2019, such as 'Qualitative data sharing and synthesis for sustainability science' and 'Advancing Computational Reproducibility in the Dataverse Data Repository Platform'. Social media icons for Facebook, Twitter, and LinkedIn are also visible.

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



18

Direct Financial support to icipe from:-



The PASET Regional Scholarship and Innovation Fund
 An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



19

The PASET Regional Scholarship and Innovation Fund

PASET
Partnership for Skills in Applied Sciences, Engineering & Technology

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

For more information contact
Regional Coordination Unit
 Regional Scholarship and Innovation Fund
 International Centre of Physiology and Ecology (*icipe*)
 P.O. Box 30772-00100, Nairobi, Kenya
 Tel +254 (20) 8632000

✉ dg@icipe.org, icipe@icipe.org, rsif@icipe.org
 📘 facebook.com/TheRSIF
 🐦 @pasetrsif
 🔗 linkedin.com/in/PASET-RSIF/

20