Republic of Botswana

BUAN
BOTSWANA UNIVERSITY OF AGRICULTURE AND NATURAL RESOURCES
Inspiring Sustainable Growth

21st Annual General Meeting
01-05 December 2025
Botswana

RUFORUM
Capacity Building in Agriculture

RUFORUM 21ST ANNUAL GENERAL MEETING 2025

CONCEPT NOTE

Training Workshop on Data Management, Data Science and Bioinformatics (SADC Region)

THEME

Positioning Africa's Universities and the Higher Education Sector to effectively impact development processes on the continent

Hosts:
The Government of Botswana and
RUFORUM member Universities in Botswana led by Botswana University of Agriculture and Natural Resources

Venue:
BUAN BLOCK 306/1

Date: 24th -28th November 2025
Time: 08:30–17:00 SAT

Background

The Southern African Development Community (SADC) faces rising demand for professionals skilled in bioinformatics, data science, and data management to tackle regional challenges in agriculture, health, and the environment. Despite global advances in GenAI, big data, and cloud computing, many institutions lack the infrastructure, curricula, and expertise to harness these tools effectively. RUFORUM's recent virtual training, which reached over 1,000 participants across 40 African countries, demonstrated strong demand for practical, hands-on learning. Building on this momentum, the upcoming workshop at the 21st RUFORUM Annual General Meeting offers timely and strategic training to bridge technical gaps, foster collaboration, and strengthen Africa's data ecosystems for sustainable development.

Objectives

The workshop seeks to:

1. Build practical skills in Data Management, Data Science, and Bioinformatics through hands-on training tailored to agriculture, health, and environmental applications.
2. Introduce GenAI and advanced data-driven approaches to accelerate research, innovation, and evidence-based decision-making across African institutions.
3. Foster regional collaboration and knowledge exchange among SADC universities, research centers, and government agencies to strengthen data ecosystems and institutional capacity.

Expected outcomes

By the end of the workshop:

1. Enhanced technical capacity among participants in Data Management, Data Science, and Bioinformatics through hands-on, practical training.
2. Improved institutional readiness to apply data-driven approaches for research, innovation, and policy across the SADC region.
3. Strengthened regional network of professionals fostering collaboration, mentorship, and resource sharing among universities and government agencies.
4. Sustained learning and application through access to curated datasets and open-source tools like RStudio, QGIS, R Shiny, and GitHub.

Target participants

- University faculty, researchers, and postgraduate students in agriculture, animal science, life sciences, computer science, statistics, and public health.
- ICT/data officers in ministries, universities, and research institutions.
- Policymakers and technical staff working in agriculture, health, and environmental surveillance.

Workshop approach

The workshop will take Five full days (08:30–17:00), delivered by two trainers. Each day combines lectures, guided demonstrations, and intensive hands-on practice sessions.

Day 1: 25th November, 2025: Setup and Data Management Foundations
Venue: BUAN BLOCK 306/1
Registration Link: XXXX
Session Chair: XXXXXX
Rapporteurs: XXXXXXXXXXXXXXX

Time	Session	Content / Activities
08:30 – 09:00	Registration and Orientation	Welcome remarks, participant introductions, overview of objectives and tools to be used.
09:00 – 10:30	Software Installation and Setup	Installing and configuring RStudio, QGIS, R Shiny, GitHub, and sequence analysis tools.
10:30 – 10:45	Health Break	Hotel
10:45 – 12:30	Lecture: Data Management Principles	Overview of data collection, storage, metadata, governance, and data security frameworks.
12:30 – 13:30	Lunch Break	Hotel
13:30 – 15:15	Hands-on Session 1	Cleaning, structuring, and validating datasets in R; creating metadata templates.
15:15 – 15:30	Tea Break	Hotel
15:30 – 17:00	Group Exercise & Discussion	Metadata creation exercise; peer review and troubleshooting setup issues.

Day 2 – Data Management Applications

Time	Session	Content / Activities
08:30 – 09:00	Recap and Setup Review	Review of Day 1 and Q&A.
09:00 – 10:30	Lecture: Data Lifecycle and Integration	Data flow from collection to archiving; interoperability; ethical and policy considerations.
10:30 – 10:45	Health Break	Hotel
10:45 – 12:30	Lecture + Demo: Metadata Standards and Secure Sharing	FAIR principles, data citation, repositories, access control.
12:30 – 13:30	Lunch Break	Hotel
13:30 – 15:15	Hands-on Session 2	Applying metadata standards to sample datasets; collaborative sharing using GitHub.
15:15 – 15:30	Tea Break	Hotel
15:30 – 17:00	Practical Exercise	Spatial data management and visualization in QGIS.

Day 3 – Data Science (Methods)

Time	Session	Content / Activities
08:30 – 09:00	Recap and Setup Review	Key takeaways from Day 2.
09:00 – 10:30	Lecture: Data Analytics Framework	Descriptive and diagnostic analytics; understanding relationships and trends.
10:30 – 10:45	Health Break	Hotel
10:45 – 12:30	Lecture: Predictive and Prescriptive Analytics	Regression, classification, clustering, and optimization concepts.
12:30 – 13:30	Lunch Break	Hotel
13:30 – 15:15	Hands-on Session 3	Building predictive models in R (e.g., linear models, random forest); evaluating accuracy (RMSE, confusion matrix).
15:15 – 15:30	Tea Break	Hotel
15:30 – 17:00	Interactive Session	Creating interactive R Shiny dashboards for model visualization and decision support.

Day 4 – Data Science (Applications & GenAI Introduction)

Time	Session	Content / Activities
08:30 – 09:00	Recap and Setup Review	Review and discussion of model performance.
09:00 – 10:30	Lecture: Applied Data Science Case Studies	Examples from agriculture, health, and environmental monitoring.
10:30 – 10:45	Health Break	Hotel
10:45 – 12:30	Lecture: Introduction to Generative AI	Overview of GenAI tools (e.g., ChatGPT, etc) and their applications in research and analytics.
12:30 – 13:30	Lunch Break	Hotel
13:30 – 15:15	Hands-on Session 4	End-to-end analytics workflow: from data import cleaning -- modeling -- visualization in RStudio and QGIS.
15:15 – 15:30	Tea Break	Hotel
15:30 – 17:00	Practical Extension	Applying workflow concepts to domain-specific datasets.

Day 5 – Bioinformatics

Time	Session	Content / Activities
08:30 – 09:00	Recap and Preparation	Overview of the day and expected outcomes.
09:00 – 10:30	Lecture: Introduction to Bioinformatics	Overview of sequence analysis, genomics, metagenomics, and transcriptomics.
10:30 – 10:45	Health Break	Hotel
10:45 – 12:30	Demonstration: Sequence Analysis Workflows	Data formats (FASTA, FASTQ, GFF), quality checks), and alignment (BLAST).
12:30 – 13:30	Lunch Break	Hotel
13:30 – 15:15	Hands-on Session 5	Running sequence data analysis, interpreting outputs, and linking findings to research applications.
15:15 – 15:30	Tea Break	Hotel
15:30 – 17:00	Closing Session	Feedback, Q&A, wrap-up discussion