DS&AS

Overview

Our **Data Science & Analytics Section (DS&AS)** has a much broader role than just data cleaning and analysis; it is a **strategic research enabler.**

Here are the main roles of the section:

- **Policy & strategy support** (analytics-related guidelines, standards, review).
- **Applied statistical and data science support** (design, analysis, reporting).
- **Data governance** (safeguarding, workflows, database management, compliance).
- **Monitoring & evaluation** (of projects, disease programs, interventions).
- **Specialized research** (epidemiology, bioinformatics, genomics, computational tools, machine learning, predictive modelling).
- **Capacity building** (training, mentorship, dissemination).

Given these roles, the section SOPs need to ensure that:

1. **Every role has an SOP backbone** to clear any ambiguity in how the section delivers.
2. **Every project has an SAP** (Statistical Analysis Plan) as part of its lifecycle, since analysis is central to most of section functions.
3. **Compliance and reproducibility** are embedded (so data handling, modelling, genomic analysis, and reporting are auditable).

Core Domains for the SOPs

1. **Policy & Strategy Support SOPs**
    - SOP on developing, reviewing, and updating data analytics policies/strategies.
    - SOP on aligning DS&AS processes with institutional/national regulations (Data Protection Act, ethical approvals, etc.).
2. **Research Study Support SOPs**
    - SOP on study design and statistical consultation.
    - SOP on **Statistical Analysis Plans (SAPs)** (pre-specified, version-controlled, and approved).
    - SOP on reporting research results (Word, Latex/Overleaf tables/PDF, figures, dashboards).
3. **Data Governance & Security SOPs**
    - Data access and authentication procedures.
    - Data storage, backup, encryption, and disaster recovery.
    - Database and workflow management.
    - SOP on data sharing, anonymisation, and compliance.
4. **Monitoring & Evaluation SOPs**
    - SOP on project performance monitoring (M&E framework, KPIs).

- SOP on evaluating disease control programs (epidemiological and cost-effectiveness frameworks).

5. Bioinformatics & Computational Biology SOPs
    - SOP on genome and proteome data management.
    - SOP on bioinformatics pipelines (from raw sequence data to analysis).
    - SOP on development and validation of computational tools.
6. Statistical Modelling & Machine Learning SOPs
    - SOP on predictive modelling and ensemble modelling.
    - SOP on handling large datasets and trend detection.
    - SOP on reproducible coding practices (Git, R Markdown, Jupyter, etc.).
7. Capacity Building & Dissemination SOPs
    - SOP on training/mentorship delivery.
    - SOP on scientific dissemination (conferences, workshops, publications).
8. Data Collection & Capture
    - Covers survey instruments, field/lab measurements, electronic data capture, cameras, biometric tools, REDCap, ODK, CSPro, Excel, and other relevant tools.
9. Data Cleaning & Preprocessing
    - Standard procedures for cleaning, deduplicating, validating, and preparing datasets for analysis.
10. Data Integration & Harmonization
    - Guidelines for linking multiple datasets, reconciling formats, and ensuring consistent variable definitions.
11. Ethics & Regulatory Compliance
    - Covers ethical approvals, informed consent tracking, and regulatory reporting for all DS&AS research projects.
12. Data Retention, Archival & Secure Disposal
    - Procedures for long-term storage, archival policies, and secure deletion of datasets in line with institutional and regulatory standards.
13. Automated Reporting & Dashboards
    - SOP for generating real-time or periodic reports, dashboards, and visualizations for projects or institutional monitoring.