Institute of Primate Research

STANDARD OPERATING PROCEDURE (SOP) DOCUMENT

Database and workflow management

SOP No.	Issue Number	Issue Date	Revision Status	Revision Date
SOP/KIPRE/RPD/DSAS/3.1.76	Version 01	October 2025	-	-

Approvals

	Name	**Signature**	**Date**
Developed by:	_Patrick Waweru Mwaura_	_____	**_6th October; 2025_**
	_____	_____	_____
	_____	_____	_____
Reviewed by:	_____	_____	_____
Approved by:	_____	_____	_____

Table of Contents

1. PURPOSE

To establish **standardized procedures for designing, managing, and maintaining research databases and associated workflows** within DS&AS, ensuring **data interoperability, integrity, quality, reproducibility, and compliance**.

This SOP ensures that:

- Data pipelines and workflows adhere to **institutional policies and governance frameworks (SOP 1)**.
- Database operations comply with **ethical and legal requirements (SOP 2, DPA 2019)**.
- Access, versioning, and storage practices are aligned with **SOP 6 (Data Access) and SOP 7 (Storage and Backup)**.
- All workflow processes support **reproducible research and efficient analytic pipelines (linked to SOP 3–5)**.

2. SCOPE

Applies to all **DS&AS-managed databases, data processing workflows, and metadata systems**, including:

- Relational databases (**PostgreSQL, MySQL**), graph databases (**Neo4j**), and document-based/text repositories (**CSV, JSON, XML**).
- Workflow pipelines for **data cleaning, transformation, analysis, and reporting**.
- Integration, versioning, and documentation of all analytic and data-processing workflows.

3. PERSONS RESPONSIBLE:

- **Data Engineer / Database Administrator:** Designs, implements, and maintains databases and workflow pipelines, ensuring integrity and efficiency.
- **Biostatistician / Data Scientist:** Validates data pipelines, verifies reproducibility, and ensures statistical and analytic accuracy.
- **Head of DS&AS:** Monitors adherence to standards, approves workflow changes, and ensures compliance with institutional and regulatory requirements.

- **Data Protection Officer (DPO):** Ensures all database and workflow operations comply with the Kenya Data Protection Act (2019) and institutional governance policies.

4. FREQUENCY

- **Database and workflow audits:** Conducted **semi-annually** to verify integrity, performance, and compliance.
- **Updates and versioning:** Implemented with every major research project, schema change, or workflow modification.
- **Routine monitoring:** Workflow logs and data pipelines reviewed **weekly** to detect errors, bottlenecks, or unauthorized changes.
- **Ad-hoc reviews:** Triggered by software upgrades, regulatory updates, or identified risks.

5. MATERIALS

- **Database Platforms:** PostgreSQL, MySQL, Neo4j (graph), and document-based repositories (CSV, JSON, XML).
- **Workflow Management Tools:** Apache Airflow, Nextflow, Snakemake, R/Python scripts, or other approved pipeline systems.
- **Metadata and Schema Standards:** FAIR principles, HL7 FHIR, CDISC, and institutional metadata guidelines.
- **Version Control Systems:** Git, GitHub, GitLab, or institutional repositories for workflow and database scripts.
- **Documentation Tools:** Data dictionaries, workflow diagrams, standard operating procedures, and analytic logs.
- **Institutional Policies:** Data governance, access, and security frameworks applicable to DS&AS-managed systems.

PROCEDURE

- **Design:**

  • DS&AS defines database schemas, metadata standards, and workflow structures in alignment with FAIR principles and institutional guidelines.

  • Project-specific data requirements, formats, and interoperability needs are documented.

- **Implementation:**

  • Data Engineer sets up the database and workflow pipelines with **role-based access controls**.

  • Integration with analytic and reporting systems is configured, ensuring secure data flow.

- **Validation:**

  • Biostatistician/Data Scientist tests workflows and pipelines for **accuracy, reproducibility, and statistical validity**.

  • Errors, inconsistencies, or missing data handling are addressed before production deployment.

- **Audit and Monitoring:**

  • Semi-annual audits assess **data quality, workflow integrity, and compliance** with regulatory and institutional standards.

  • Routine monitoring logs and error reports are reviewed **weekly**.

- **Documentation and Version Control:**

  • All database configurations, workflow scripts, and changes are documented in the DS&AS repository.

  • Version control is maintained using **Git** or equivalent systems to track updates, ensure reproducibility, and enable rollback if needed.

- **Updates and Maintenance:**

  • Updates are applied with each new project, schema modification, or regulatory requirement.

• All changes are reviewed and approved by the Head of DS&AS before implementation.

6. REFERENCES

1. Kenya Data Protection Act (2019) and Regulations.
2. KIPRE Institutional Data Protection and Sharing Policy (2024).
3. DS&AS SOP 1 – Policies and Strategies.
4. DS&AS SOP 2 – Alignment with Institutional and National Regulations.
5. DS&AS SOP 6 – Data Access and Authentication Procedures.
6. DS&AS SOP 7 – Data Storage, Backup, Encryption, and Disaster Recovery.
7. FAIR Data Principles (Wilkinson et al., 2016).
8. HL7 FHIR Standard for Health Data Interoperability.
9. CDISC Standards for Clinical Data Management.
10. ISO/IEC 27001:2022 – Information Security Management Systems.

7. APPENDICES

Appendix 8.1 – Forms and Templates

1. Database Configuration Checklist
2. Workflow Validation Log Template
3. Data Pipeline Versioning Record
4. Data Dictionary Template
5. Workflow Audit Report Template
6. Change Request and Approval Form

Appendix 8.2 – Reference Systems and Tools

- **Databases:** PostgreSQL, MySQL, Neo4j, CSV/JSON/XML repositories.
- **Workflow Management Tools:** Apache Airflow, Nextflow, Snakemake, R/Python scripts.
- **Version Control:** Git, GitHub, GitLab, or institutional equivalents.
- **Documentation Tools:** Workflow diagrams, metadata templates, data dictionaries.