

**Institute of Primate Research**

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

**STANDARD OPERATING PROCEDURE (SOP) DOCUMENT**

**Database and workflow management**

|  |  |  |  |
| --- | --- | --- | --- |
| **Approvals** |  |  |  |
|  | **Name** | **Signature** | **Date** |
| **Developed by:** | \_Patrick Waweru Mwaura\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_6th October; 2025\_** |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Reviewed by:** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Approved by:** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

**Table of Contents**

[1. PURPOSE 4](#_Toc144316958)

[2. SCOPE 4](#_Toc144316959)

[3. PERSONS RESPONSIBLE: 4](#_Toc144316960)

[4. FREQUENCY 4](#_Toc144316961)

[5. MATERIALS 4](#_Toc144316962)

[6. PROCEDURE 4](#_Toc144316963)

[7. REFERENCES 5](#_Toc144316964)

# 

# PURPOSE

To establish procedures for managing research databases and workflows in DS&AS, ensuring interoperability, quality, and compliance.

# SCOPE

Covers all DS&AS-managed research databases, workflows, and metadata systems.

# PERSONS RESPONSIBLE:

* **Data Engineer:** Designs and maintains databases/workflows.
* **Biostatistician/Data Scientist:** Validates data pipelines.
* **Head of DS&AS:** Ensures standardization and compliance.

# FREQUENCY

* Database/workflow audits conducted **semi-annually**.
* Updates made with each major research project.

# MATERIALS

* Database platforms (PostgreSQL, MySQL, Neo4j).
* Workflow management tools (Airflow, Next flow, R scripts).
* Metadata and schema standards (FAIR, HL7 FHIR, CDISC).

# PROCEDURE

1. **Design:** DS&AS defines schema and metadata standards.
2. **Implementation:** Data Engineer configures database with access controls.
3. **Validation:** Biostatistician tests pipelines for accuracy and reproducibility.
4. **Audit:** Semi-annual checks for data quality, interoperability, and compliance.
5. **Documentation:** All workflows documented and stored in DS&AS repository.

# REFERENCES