

**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**

## **Development and validation of computational tools**

|  |  |  |  |
| --- | --- | --- | --- |
| **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 provide a standardized framework for the development, validation, and maintenance of computational tools—including algorithms, software, and machine learning models—used in DS&AS research, ensuring reproducibility, accuracy, usability, and compliance with institutional and regulatory standards (building on SOPs 4, 7, 8, 12, and 13).

# SCOPE

Covers the design, development, testing, validation, deployment, and maintenance of all in-house computational tools and customized open-source software applied in DS&AS research, including genomics, proteomics, epidemiology, and predictive modelling workflows.

# PERSONS RESPONSIBLE:

* **Computational Biologist / Data Scientist:** Leads design, development, and validation of computational tools.
* **Software Engineer / Data Engineer:** Supports coding, optimization, workflow integration, and deployment.
* **Head of DS&AS:** Reviews and approves final tool release, ensures compliance with institutional and regulatory standards, and oversees maintenance.

# FREQUENCY

* **Initial Validation:** All computational tools must be validated before deployment for use in any research project.
* **Re-Validation:** Required whenever major updates, methodological changes, or modifications to data inputs occur.
* **Periodic Review:** Tools should undergo scheduled reviews (e.g., annually) to ensure continued accuracy, reproducibility, and compliance.

# MATERIALS

* + **Coding Platforms:** Python, R, C++, Java, or other relevant programming environments.
  + **Version Control:** Git, GitHub, GitLab, or equivalent for source code management and change tracking.
  + **Test Datasets:** Benchmarking datasets, simulated or real, for tool validation and performance assessment.
  + **Continuous Integration / Deployment (CI/CD) Tools:** For automated testing, build verification, and workflow deployment.
  + **Documentation Templates:** Standard templates for recording tool specifications, validation results, and usage instructions.

# PROCEDURE

1. **Development:** Design and implement prototype tools following reproducible coding standards and modular practices.
2. **Testing:** Evaluate tool functionality and performance using benchmark datasets; assess accuracy, efficiency, and reproducibility.
3. **Validation:** Conduct internal peer-review within DS&AS; document performance metrics, limitations, and compliance with institutional guidelines.
4. **Deployment:** Release validated tools for internal use or as open-source (if applicable), accompanied by comprehensive user manuals and documentation.
5. **Maintenance:** Continuously monitor tool usage, collect bug reports, implement updates, and track changes through version control.
6. **Archiving:** Maintain comprehensive records of all versions, validation reports, test datasets, and change logs in the DS&AS repository.

# REFERENCES

1. ISO/IEC 25010:2011 – Systems and software engineering: Software product quality requirements and evaluation (SQuaRE).
2. FAIR Principles – Wilkinson et al., 2016 (linked to SOP 1: Policies & Strategies; SOP 12: Genome and Proteome Data Management).
3. Kenya Data Protection Act (2019) (linked to SOP 2: Alignment with Institutional and National Regulations; SOP 6: Data Access and Authentication; SOP 9: Data Sharing and Anonymisation).
4. KIPRE Institutional Data Governance and Software Development Guidelines (linked to SOP 1: Policies & Strategies; SOP 7: Data Storage, Backup, Encryption, and Disaster Recovery; SOP 8: Database and Workflow Management).
5. Best practices for reproducible computational research: Sandve et al., 2013, PLoS Comput Biol (linked to SOP 3: Study Design and Statistical Consultation; SOP 4: Statistical Analysis Plans; SOP 5: Reporting Research Results; SOP 13: Bioinformatics Pipelines; SOP 14: Development and Validation of Computational Tools).
6. Git/GitHub/GitLab documentation and version control guidelines (linked to SOP 4: Statistical Analysis Plans; SOP 13: Bioinformatics Pipelines; SOP 14: Development and Validation of Computational Tools).

# ****APPENDIX****

**Appendix A: Computational Tool Development & Validation Forms**

1. Tool Development Request Form – Submitted by DS&AS team or PI.
2. Tool Validation Checklist – Includes test datasets, expected outcomes, reproducibility metrics, and performance evaluation.
3. Peer-Review Record – Documenting internal review comments, approvals, and recommendations.
4. Version Control Log – Tracks all code versions, updates, and major changes.
5. Deployment & User Documentation Template – Provides standardized user instructions and usage guidelines.