Data Science & Analytics (DS&AS) Policy Framework for KIPRE

Executive Summary

The Kenya Institute of Primate Research (KIPRE) is mandated to conduct biomedical pre-clinical research to improve human health. Its draft Data Protection and Sharing Policy outlines broad principles (e.g., data access rights in Art. 10 and breach notification in Art. 11.2) but lacks operational details. This DS&AS framework explicitly fills those gaps. It defines clear workflows, integrates DS&AS roles in governance, and recommends practical measures – e.g., data governance structures, training programs, audit procedures, interoperability standards, and secure access systems – to strengthen policy implementation. Each section below references the draft policy's articles to show how the DS&AS align with or correct the current deficiencies.

1. Implementation Workflows for Data Processes

The draft policy states general obligations for data sharing (Article 8) and breach notification (Article 12) but does not provide step-by-step procedures. The DS&AS will develop detailed workflows for key processes:

- Data Sharing Workflow: - A formal Data Access Request process will be introduced. When a researcher or external party requests data, DS&AS will coordinate a review that includes legal/ethical clearance, data classification (e.g., anonymization level), and approval by a Data Governance Committee (DGC). Secure transfer protocols (e.g., encrypted channels per Art. 8.2) will be mandated. Tracking logs and metadata will record each sharing event.
- Breach Response Workflow: - Although Article 12 requires breach reporting (e.g., "notify within 72 hours"), it lacks a sequence of actions. The DS&AS section will help define a Breaches Response Plan: automated alerts (e.g., intrusion detection), immediate containment steps, assignment of roles (ICT to secure systems, DS&AS to analyze incident data patterns, DPO to notify authorities), and post- incident review. This will align with Art. 12's requirements but add clarity on responsibilities and timelines.
- Data Quality Assurance Workflow: - The policy's vision (Art. 1.9.2) lists "Quality assurance standards," but no concrete process is given. DS&AS will institute routine quality checks: standardized data validation rules, periodic audits of datasets (completeness,

consistency), and error-correction feedback loops. Data scientists will use analytics to detect anomalies (e.g., outliers or duplicates) and coordinate with researchers to remediate issues. All datasets will include metadata and documentation (as intended in Art. 1.9.2) to support transparency and reuse.

Each workflow will be documented with flowcharts and checklists. For example:

- Data Sharing Workflow Steps: Request submitted (user portal) → 2. DS&AS triage (classify data sensitivity) → 3. DPO/Legal and ethics review → 4. DS&AS/ICT implement anonymization or encryption → 5. Secure data delivery (e.g., encrypted file share) → 6. Log completion and update data registry.
- Breach Response Steps: Detection (auto-monitoring or user report) → 2. Assessment (DS&AS and ICT determine scope) → 3. Containment (isolate affected systems) → 4. Notification (DPO alerts data subjects and regulators as per Art.12) → 5. Eradication & Recovery → 6. Post-incident Review (lessons learned, system hardening).

These specific procedures will ensure that KIPRE's responses are timely and coordinated – addressing the lack of clarity in Art. 12 and emphasizing data quality where the draft merely lists it as a principle.

2. Integrating DS&AS Roles and Advanced Analytics

The policy's administration section (Article 2) enumerates data-related roles (Board, DPO, ICT, etc.) but omits the DS&AS entirely. I propose amending Article 2 (e.g., new "2.9") to formally recognize DS&AS. The DS&AS, led by a Chief Data Scientist, would include data engineers, data analysts, and statisticians. Key duties would include:

- Advanced Analytics Guidance: Develop standards for applying machine learning and statistical methods to KIPRE's data. Ensure compliance with data privacy (e.g., privacy-preserving analytics) while enabling research insights.
- Data Stewardship: Oversee metadata management and data cataloging. Establish common data models (e.g., health research ontologies) for interoperability. Maintain the central research data repository.
- Policy Liaison: Work with the DPO, ICT, and researchers to implement data governance. Translate policy requirements into technical controls (e.g., encryption standards from Art. 8) and software features.

- Cross-Functional Support: Assist research teams with study design, data collection protocols, and statistical analysis plans. Ensure that data (especially personal or sensitive data) is handled per legal and ethical guidelines.

Defining these DS&AS roles fills a critical gap: currently, Article 2's ICT roles focus on technical infrastructure but do not address data analytics or strategy. By integrating DS&AS, KIPRE ensures that advanced analytics is governed alongside ICT (making DS&AS a "key complement" to ICT).

3. Balancing Accessibility and Protection

The policy's Article 10 (Data Access Rights) and Article 8 (Data Sharing) emphasize transparency and security. Data subjects "have the right to obtain information" (Art.10), yet the policy does not detail how to manage these rights in practice. Similarly, Art. 8 allows sharing with safeguards but lacks operational guidance on de-identification or tiered access. DS&AS will implement concrete protocols to balance openness with protection:

- Data Classification Framework: All datasets will be labeled (e.g., Open, Restricted, Confidential) based on sensitivity. This governs who may request access and what anonymization is required. For example, Restricted health datasets might only be available via secure analysis platforms.
- Tiered Access Controls: Implement role-based access and data enclaves. Sensitive data can be accessed only on designated secure servers with logging; de-identified data may be downloaded under agreement. This operationalizes Art.8.2 encryption/pseudonymization requirement and Art.11.3's "need-to-access" principle.
- Automated Governance Tools: Use data catalogs and APIs to track who accesses which data. DS&AS will maintain a "Data Governance Dashboard" that monitors access requests, compliance with consent, and detects unusual patterns. This ensures real-time enforcement of Art. 10 rights without manual bottlenecks.
- Consent & Transparency: Enhance mechanisms so data subjects can easily view and manage their consents. For example, a web portal may allow research participants to see what data of theirs is held, in line with Art. 10.1, and to update preferences.

By formalizing these measures, DS&AS addresses the policy's current vagueness on protecting data while enabling research use. Advanced techniques (e.g., differential privacy, federated learning) will also be explored for analyses on sensitive data, further bridging the access-protection trade-off.

4. Training and capacity building

The draft policy tasks the DPO with "facilitating capacity building of staff"
but provides no details on scope or frequency. We recommend a comprehensive DS&AS-led training program:

- Data Governance Training: Regular workshops for all staff on KIPRE's data policies, legal requirements (e.g., Kenya's DPA 2019), and ethical data use. Training will cover subjects from data privacy basics to advanced topics like machine learning ethics.
- Technical Skills Development: Offer courses in statistical programming, database management, and cybersecurity best practices. DS&AS analysts can mentor researchers on data analysis tools (R, Python, SQL) and reproducible research workflows.
- Onboarding & Refresher Courses: New hires will receive mandatory data policy and tools training. Existing staff will attend annual refreshers and compliance quizzes to reinforce understanding.
- Certifications and Incentives: Encourage relevant certifications (e.g., Certified Data Privacy Professional) and create a recognition program for "Data Champions" who exemplify best practices.
- Partnerships and E-Learning: Collaborate with academic partners (e.g., local universities) for joint training and provide access to online modules (e.g., Coursera or WHO data courses) for continuous learning.

These programs will be tracked by DS&AS to measure effectiveness (e.g., pre-, and post-training assessments). Involving DS&AS in training ensures a data-driven approach to capacity building, fulfilling Art.2.3's intent with more structure, and addressing the absence of explicit training mandates in Article 15.

5. Enforcement and Monitoring Structures

Currently the policy designates only the Internal Audit Unit and the DPO for oversight. To strengthen enforcement, DS&AS will collaborate in the following:

- Data Governance Committee (DGC): Establish a standing committee (including the DPO, ICT lead, DS&AS head, and external

advisor) to oversee policy compliance. The DGC will review audit findings, approve data sharing requests above a sensitivity threshold, and handle escalations.

- Enhanced Audits: In addition to biannual IT audits by Internal Audit, conduct specialized Data Audits led by DS&AS. These will examine data pipelines and analytics processes for compliance with standards (e.g., encryption at rest, access logs).
- KPIs and Dashboards: Define key performance indicators (KPIs) for data governance (e.g., breach incident rates, data request turnaround time, data quality scores). DS&AS will maintain dashboards (internal metrics portals) to track these KPIs in real time, alerting leadership to anomalies.
- Reporting Lines and Consequences: Clarify that serious data breaches or misuse led to formal review by the DGC, with sanctions per Art.5.2 and possible legal action. Responsibilities and consequences will be documented so that "knowingly circumventing safeguards" carries accountability.
- External Audits and Accreditation: Periodically invite external experts (e.g., Ministry of ICT assessors) to review KIPRE's data management. Seek compliance certifications where possible (e.g., ISO 27001), reinforcing trust beyond the DPO's annual review

.

By creating these structures, DS&AS expands monitoring beyond the DPO-centric model. The DGC ensures collective oversight, and metrics-driven monitoring addresses Art.15's call for review with concrete data.

6. Aligning ICT Infrastructure with Policy

The policy's ICT provisions (Art.2.4–2.5) task ICT with "secure data from loss" and developing databases, and Article 11 cites the ICT Security Policy's role. However, the link between KIPRE's infrastructure and policy enforcement is weak. DS&AS will work closely with ICT to ensure the technical environment supports data governance:

- Data Architecture & Platforms: Implement a centralized data warehouse/repository that enforces access controls and audit logging at the system level. This could be a secure cloud or on-premises data platform that integrates with analytics tools. ICT will provision the infrastructure (servers, networks) while DS&AS defines requirements (e.g., encryption, redundancy) aligned with Art.11's secure handling.

- Interoperability Standards: Adopt open standards (e.g., HL7 FHIR for health data, CDISC for clinical studies) so that data collected by different departments can interoperate. DS&AS will maintain metadata schemas and ensure all new systems comply with these standards, reflecting the policy's aim for "interoperability".
- Secure Access Systems: Introduce multi-factor authentication and VPN access for remote data work, beyond the base "need-to-access" rule. Develop a secure data portal (HTTPS/SSL, per the illustration) for internal and external queries – enabling Art.10's electronic copy provision in a protected manner.
- Monitoring and Logging: ICT systems will generate detailed logs of data access, which DS&AS will analyze for irregularities. This aligns with Art.11.1's emphasis on "secure handling" by adding visibility into data usage.
- Scalability and Updates: Design infrastructure with scalability (to support growth in genomic or imaging data) and automatic security updates. Regular penetration testing and patch management will be scheduled to enforce the ICT security policies referenced in Article 11.

This alignment ensures that technical measures (networks, databases, security protocols) are not just ICT deliverables, but also tools for enforcing the data policy. DS&AS will act as the bridge, translating policy requirements into infrastructure specifications and ensuring new ICT initiatives (e.g., a data lake project) incorporate policy controls from the design phase.

7. External Engagement and Dispute Resolution

Article 14 governs dispute resolution but is internally focused, assigning complaints to the DPO and Director General. To enhance stakeholder trust and transparency, DS&AS will expand engagement with external parties:

- Liaison for Data Collaborations: Appoint DS&AS staff as official contacts for external researchers or public health agencies requesting data. They will vet requests against policy criteria (ensuring "adequate safeguards") and negotiate Data Use Agreements.
- Public Data Sharing: Develop procedures for publishing de-identified datasets in open repositories (in line with the commitment to Open Access). This proactive sharing enhances transparency and scientific impact.
- Advisory Panels: Include external experts (e.g., ethicists,

community representatives) in the Data Governance Committee or ad hoc review panels. This extends Art.14's resolution mechanism beyond internal actors and provides neutral arbitration for disputes.

- Conflict Resolution Channels: Establish clear escalation paths for disagreements over data (e.g., a dispute review board separate from the DPO). Document procedures for arbitration that reference both KIPRE policy and national ICT dispute frameworks.
- Stakeholder Feedback: Regularly survey internal and external stakeholders (researchers, data subjects) about the policy's effectiveness. Use feedback to adjust practices between formal 3-year reviews (addressing Article 17's rigidity).

By codifying these engagement steps, DS&AS ensures that KIPRE's data policy operates transparently with its partners and the public. This fills the gap in Art.14, which currently only specifies internal reporting and DPO-led inquiries and reinforces KIPRE's mandate to disseminate knowledge safely.

8. Adaptive Policy Review Mechanisms

Article 17 mandates a three-year review cycle (with annual data updates)
but does not specify interim adjustments. DS&AS will institute a more dynamic review process:

- Continuous Monitoring: Using the enforcement dashboards (Sec.5), DS&AS will track policy-relevant metrics (e.g., number of breach incidents, data request backlogs, training completion rates). Significant deviations (such as new threat patterns) will trigger immediate policy evaluations.
- Annual Policy Impact Report: Each year, DS&AS and the DPO will produce a report on policy effectiveness. This report will recommend any needed amendments in procedures or guidance (e.g., new tech standards), rather than waiting three years.
- Benchmarking and Best Practices: DS&AS will scan emerging data governance best practices (e.g., WHO guidelines, new DPA regulations) and propose updates. If, for example, new analytic methods raise privacy concerns, the policy can be updated rapidly rather than on a strict timetable.
- Stakeholder Review Sessions: Host annual meetings where staff and external partners review how well the policy is working. Gather suggestions for improvement to feed into the next formal review.

This adaptive cycle ensures the policy evolves with KIPRE's data needs and technological advances. It supplements Article 17's fixed window with ongoing

oversight.

Conclusion

By explicitly addressing the draft policy's gaps, these DS&AS-driven enhancements ensure robust data governance at KIPRE. Clear implementation workflows and defined DS&AS responsibilities make data sharing, security, and quality processes transparent and actionable. Expanded training, monitoring, and stakeholder engagement fill the policy's silences in Articles 2.3, 2.6, 8–10, 14–15. Together with ICT, DS&AS will implement modern infrastructure and analytics (e.g., secure data platforms, interoperable standards) that align practice with policy. This comprehensive approach leverages data science as a key complement to ICT, ensuring KIPRE fulfills its mandate for data-driven research and public health innovation.

Written by Patrick Mwaura,

For DG KIPRE.