This document is based on the Kenya Institute of Primate Research (KIPRE) 2023-2027 Strategic Plan document. Its aim is to contextualize the Data science and analytics section (DS&AS) as per the institute’s organogram. The section is under Research and Product development directorate.

The development of this document is cognizant of the draft data policy that outlined critical information on how data should be handled and the role of the ICT manager and team under the ICT division. The ICT division is under the corporate services directory.

While data policy draft is comprehensive and provides a comprehensive high-level framework, it has several notable weaknesses that call for strengthening. It is important to outline some of them here as an anchor to the main aim of this document.

The draft lacks:

1. A detailed implementation procedures and workflows for essential processes. This includes data sharing processes, breach response, and data quality assurance. Specific operational protocols for routine data management, quality control, and audit mechanisms are also insufficiently elaborated.
2. Emphasis on data science integration. Although the draft addresses data protection thoroughly, it offers little guidance on how data science and analytics teams can leverage data for advanced analytics, integration, and decision-making support. The policy could better define roles for analytic pipelines, data modelling, and results dissemination within the overall policy. The DS&AS team is completely not mentioned, and they are supposed to be great collaborators.
3. Accessibility and usability protocols. The draft strongly underscores data privacy and safety but lacks clear guidelines on balancing these priorities with legitimate research and collaboration needs. The policy should establish explicit rules and secure platforms for controlled data access and sharing would help reconcile security with usability.
4. A complete training and competency development guidelines. The draft mentions responsibilities for the Data Protection Officer (DPO) and staff training. However, it omits a detailed capacity-building strategy for the data policy and data science skills. Regular training programs and assessments for all staff involved in data handling are essential to sustain compliance and build institutional expertise.
5. Clearly defined enforcement and accountability mechanisms. Although disciplinary measures are mentioned, the draft provides limited detail on enforcement procedures, monitoring responsibilities beyond the DPO, and consequences of non-compliance. The policy needs a clearly assigned monitoring teams or committees that would enhance adherence.
6. Technological and infrastructural alignment. The policy offers little information on how current or planned ICT infrastructure and tools will practically support the data policy framework. The policy needs a better integration between policy requirements and ICT capabilities. This may include automated monitoring, audit trails, and secure data-sharing platforms. This are supposed to ensure effective implementation.
7. Stakeholder engagement and communication protocols. The draft focuses primarily on internal compliance. There is limited coverage on strategies for engaging external collaborators, research participants, and the public to ensure transparency in data use and protection. Mechanisms for feedback and dispute resolution with stakeholders should also be clearly articulated.
8. Lastly but not least, the policy lacks clarity on review frequency and adaptability. There is no clearly defined cycle for reviewing and updating the policy to keep pace with evolving governance standards, technological advancements, or regulatory updates. Adaptive policy is critical to maintain relevance and effectiveness over time.

These gaps will help the data policy to be more practical, comprehensive, and effective framework that supports both data protection and science-driven innovation.

The DS&AS section and the ICT division must work collaboratively to achieve the institution’s goals on data management while adhering to established guidelines and legal requirements.

The nature of these two important functions may lead to some overlapping or conflicting roles. It is therefore important to distinguish the roles of these two functions. We consult the two documents, KIPRE strategic plan as well as the data policy, to achieve the following distinctive roles for each.

* **ICT Division**

Focuses or should focus on:

* + IT infrastructure, communications technology, network security, databases, and systems that support data storage, security, and access.
  + Upgrading ICT infrastructure, deploying ERP systems, ensuring cyber security, and maintaining communication platforms.
  + Supporting digitization and automation of internal operation processes, financial systems, procurement, and document management.
  + Ensuring compliance with data privacy regulations and security protocols.
  + Providing platforms for virtual collaboration and enhances technology-enabled research operations.
* **Data science and analytics section**

Focuses or should focus on:

* + Developing data management, data governance, analytics, and data protection strategies.
  + Developing and deploying advanced data analytical tools, statistical models, machine learning, and visualization platforms to extract insights from research data.
  + Supporting epidemiological modeling, disease surveillance, mapping risky zones, and One Health data integration.
  + Building capacity for data-driven decision making, risk assessment, and research outcome monitoring.
  + Implementing data quality assurance, database management, meta-analysis, and knowledge dissemination via interactive reporting tools.
  + Developing performance tracking systems, facilitating data-driven resource allocation, and strengthens the analytic support for conservation, biodiversity, and climate response research.

**DS&AS should address the following Gaps.**

The gaps have been identified from the KIPRE legal notice, the Strategic Plan, opinions from key stakeholders and researcher as well as the global best data handling and processing practices as stipulated by the Royal Statistical Society (RSS) and the American Statistical Society (ASA).

1. **Data governance and management**
   * Lack of comprehensive data governance framework and policies.
   * Need to develop standardized data management procedures, metadata standards, and data sharing protocols to promote data quality and interoperability.
2. **Advanced Analytics Capacity**
   * Need for strengthening analytics skills to conduct complex epidemiological modeling, predictive analytics, and integration of heterogeneous data (molecular, ecological, clinical).
   * Enhanced capacity for environmental antimicrobial resistance monitoring and phage research analytics.
3. **Surveillance and Real-Time Data Integration**
   * Gaps in disease and ecosystem surveillance data analytics, including the use of spatial mapping and risk modeling.
   * Need for real-time data capture, integration, and visualization tools to support decision-making in One Health and pandemic preparedness.
4. **Research impact and monitoring**
   * Inadequate analytic systems for tracking research outputs, funding impact, and performance monitoring.
   * Tools for evaluating interventions, managing research portfolios, and reporting to stakeholders are lacking.
5. **Capacity building in data science**
   * Insufficient dedicated trained personnel and training programs in advanced data science methods.
   * Need for continuous professional development in bioinformatics, statistical computing, and AI for research data.
6. **Data-driven conservation and climate research**
   * Limited use of integrated analytics for biodiversity monitoring, ecosystem health assessment, and climate change response.
   * Development of digital platforms for data sharing, community-driven conservation data, and climate adaptation modeling needed.
7. **Ethics and data security in analytics**
   * Implement data protection strategies aligned with ethical standards for handling sensitive health and conservation data.
   * Ensure compliance with institutional policies and external regulations on data privacy and research ethics.

**Strategic priorities for data science and analytics**

* Develop a centralized data analytics and visualization platform.
* Establish robust data governance and protection frameworks.
* Enhance interdisciplinary analytics to connect public health, primate biology, and environment domains.
* Create analytic support for monitoring and evaluation to improve operational effectiveness.
* Invest in capacity building through training and mentorship in modern data science tools and techniques.
* Support research innovation by translating analytics into decision-support tools for conservation, health, and policy making.

By focusing on these gaps and strategic priorities, the Data Science and Analytics section can provide the analytical backbone that complements the ICT infrastructure, thereby advancing KIPRE’s research excellence and impact on public and ecosystem health.