Kenya Institute of Primate Research

Institutional Needs Assessment Report

Data Science & Analytics Section

Prepared by:

Patrick Mwaura
Data Science & Analytics Section

October 13, 2025

Executive Summary

Between 1–8 October 2025, the Data Science & Analytics Section (DS&AS) conducted a needs assessment across all KIPRE staff to identify gaps in data capacity, quality, security, and analytics, and explore collaboration opportunities. Out of 152 staff, 25 (16.4%) responded, predominantly from research divisions. Findings indicate:

- Data collection and analysis are mainly manual and research-focused, with inconsistent database use and limited advanced analytical tools.
- Data quality is moderate, but integrity, security, and governance are weak, with gaps in SOPs, access control, backups, and staff training.
- Analytical capacity is uneven, with high reliance on a few advanced staff and broad training needs in statistical modelling, data security, machine learning, and software (R, Python, SAS, SQL).

Staff view DS&AS as a strategic partner to strengthen data workflows. Key recommendations include implementing a structured analytical support framework, targeted capacity-building programs, formal SOPs, and improved infrastructure for secure data governance, starting with research divisions and gradually expanding to administrative units.

Contents

1 Introduction

Between 1^{st} and 8^{th} October 2025, the Kenya Institute of Primate Research (KIPRE) Data Science and Analytics Section (DS&AS), under the Research and Product Development Directorate (RPD), carried out a needs assessment on data science and analytics. The exercise targeted all 152 (as per the KIPRE SP 2023-2027) employees across directorates and units, aiming to identify gaps in data capacity, quality, security, and analytics, as well as explore opportunities for collaboration. A structured Google form was used, shared via email during the first four (4) days and through the institute Whats-app forum during the last two days, the other two being ignored since they fell on a weekend . The form took about five (5) minutes to complete, and employees were encouraged to respond honestly while maintaining anonymity. Data were analyzed using aggregation, with conclusions drawn from the proportion of responses in each category, applying critical thinking and objectivity. Generalizations were made cautiously, recognizing that participation was voluntary and random.

2 Who responded?

A total of 25 staff members responded to the assessment, representing a response rate of 16.4%. Most responses came from core research units (Table 1). The Research and Product Development (RPD) directorate contributed 20 responses (80%), followed by the Animal Science, Welfare and Ethics Directorate (ASWED) with 4 responses (16%), and Corporate Services Directorate (CSD) with 1 response (4%). No responses were received from the Capacity Building, Partnerships and Grant Management Directorate (CBPGMD), Internal Audit and Risk Assurance (IA&RA), or Corporate Services and Legal Services (CS&LS). This suggests low awareness or limited engagement with data initiatives in some non-research directorates.

The lack of responses from some directorates may be due to several reasons. Some of these directorates are still being operationalized as the institute transitions to its new structure. Others, such as CBPGMD, have very few staff who may also be overstretched. There may also have been communication gaps or limited interest in data-related initiatives among some administrative units.

Divisional trends (Table 1) show that participation was highest among teams potentially handling large scientific datasets. These include Reproductive Health & Biology 8 (32%), Infectious Diseases & One Health 7 (28%), and Veterinary Clinical & Diagnostic Services 4 (16%). Divisions such as Training, Supply Chain, HR, Finance, and Partnerships did not participate, pointing to possible gaps in institutional data culture beyond research programs.

Notably, the Kenya Snakebite Research and Intervention Program (KSRIC), though a key scientific program, did not participate. It is possible that staff in such specialized programs may not feel fully integrated into the wider KIPRE research community and may not yet see the value of central data initiatives. This observation is important and signals an opportunity. As DS&AS,

Table 1: Respondents by Directorate and Division

Directorate			**Division/Section**		
Name	Responses	%	Name	Responses	%
RPD	20	80	Reproductive Health & Biology	8	32
ASWED	4	16	Infectious Diseases & One Health	7	28
CSD	1	4	Vet Clinical & Diagnostic Services	4	16
			Evo. Ecology, Conserv. & Ecosystem Health	2	8
			Corporate Communication	1	4
			ICT	1	4
			Non-Communicable Diseases	1	4
			Planning & Strategy	1	4
Total	25	100	Total	25	100

our role is to build collaboration across all units—research and non-research—to strengthen data capacity institute-wide. This requires targeted engagement strategies moving forward.

Role-wise distribution (Table 2) indicates strong participation from technical/research roles, with limited representation from admin/governance roles.

Table 2: Respondents by Role

Role	n	%
Research Scientist	10	40
Asst Research Scientist	6	24
Director	2	8
Admin Officer	1	4
Asst Director/HOD	1	4
Deputy Director/HOD	1	4
Director General	1	4
ICT Officer	1	4
Senior Technician	1	4
Vet Doctor	1	4
Total	25	100

Taken together, the response rates show stronger engagement from research and technical units compared to administrative and governance units. The low responses from certain units, particularly the ICT division, raise important concerns. ICT plays a central role in data infrastructure, digital systems, and information security, so their limited participation may suggest either low awareness of the exercise, competing priorities, or a perception that DS&AS activities are separate from their mandate. It may also reflect gaps in internal collaboration where ICT functions more as a support unit rather than an active partner in data-driven development within the institute.

Other possible reasons for low participation across some corporate units include ongoing structural transitions, limited communication reach, or uncertainty about the relevance of data initiatives to their daily responsibilities. These findings show that while data practices and analytical activity are strongest within research divisions, institutional data development will require broader engagement beyond research.

Moving forward, DS&AS will build on the strong interest already shown by researchers while giving special attention to underrepresented units like ICT. Strengthening collaboration with ICT will be a priority, especially in areas such as data governance, secure storage, system integration, and deployment of data tools. This will help create a more coordinated and inclusive data environment across the institute.

3 Current data capacity

The survey shows that KIPRE's data practices are mainly research-driven. Data collection is still dominated by basic tools, with Excel (43%) and paper-based forms (30%) being the most commonly used. Modern digital tools such as REDCap, ODK/KOBO Collect, and Google Forms are used but remain underutilized. Almost half of the respondents work with biological data (48%), while 36% indicated that they handle all types of data, reflecting a strong research orientation in the institute.

Database use is limited and inconsistent. MS Access (21%) and text-based storage formats like CSV/TXT (21%) are used at similar levels, but a significant number of respondents (21%) reported that they do not use any database system or were not sure. This shows a lack of standardized data management frameworks across programs and divisions.

For data analysis, most staff rely on familiar tools. Excel remains the most commonly used analytical tool (36%), followed by R (22%) and GraphPad Prism (16%). Use of programming-based tools such as Python, SAS, and development environments like VS Code is minimal, indicating a gap in advanced analytics and reproducible research practices. Figure 1 summarises these trends.

Taken together, these findings highlight that while research units are actively engaged in data work, KIPRE still has gaps in modern data capacity. Data collection remains largely manual, database use is inconsistent, and analytical work depends heavily on basic tools. This limits efficiency, data quality, and scalability, especially as research output grows. There is a clear need to standardize data practices, introduce structured data management systems, and promote wider use of digital tools and programming skills. Strengthening these areas will not only improve research quality but also build a stronger institutional data foundation that supports evidence-based decision-making across the Institute.

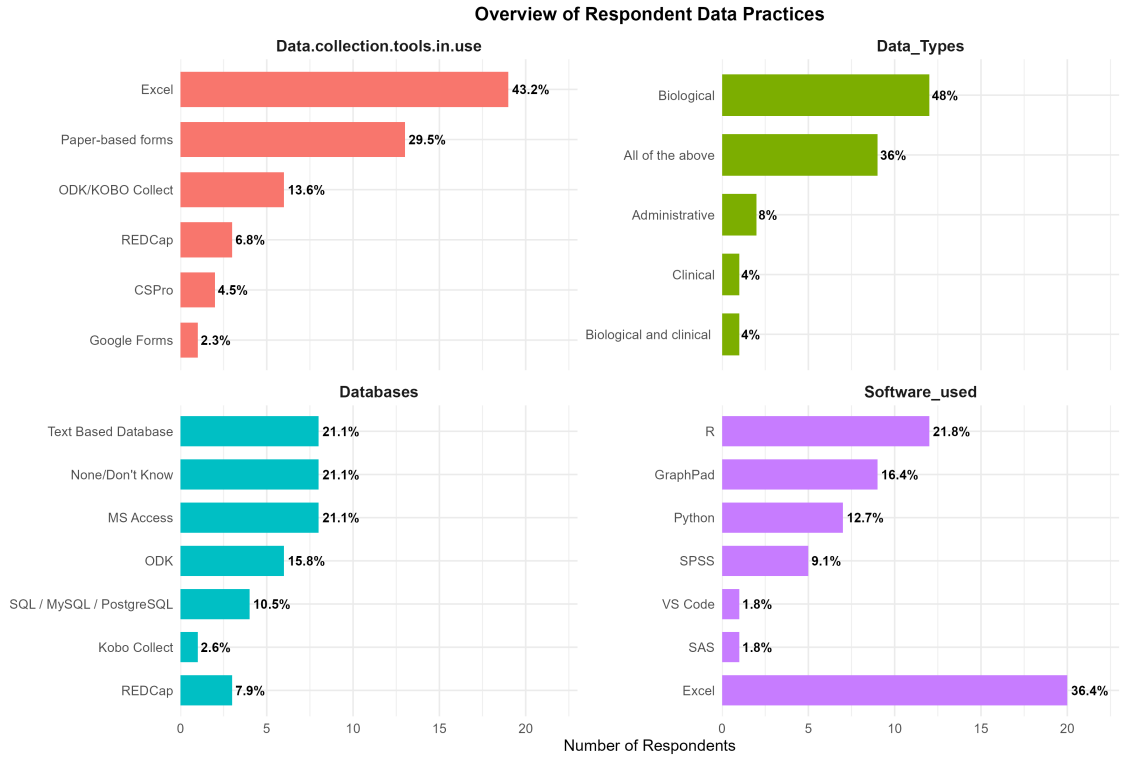

Figure 1: Overview of respondent data practices across categories: data types, software, data collection tools, and databases. Percentages indicate relative usage among respondents.

4 State of data quality & security

The survey reveals a mixed picture of the Institute's data environment. As shown in Figure 3, most respondents rated data quality as Good or Fair across the four dimensions assessed: accuracy, completeness, consistency, and reliability. Only a small proportion rated their data as Excellent or Poor. This suggests that research and technical units maintain a reasonable standard of data quality. However, the variation across categories points to occasional gaps in documentation, standardization, and completeness, which may affect analysis and decision-making.

In contrast, perceptions of data integrity, protection and security are more concerning. A significant share of respondents rated these areas as Not Protected or Low. For example, 44% of staff rated data integrity as Not Protected or Low, while 48% rated data protection at the same level. Data security followed a similar pattern, with very few respondents reporting High or Very High levels of protection. These findings suggest that many staff feel exposed to risks such as data loss, unauthorized access, or manipulation of datasets. The low ratings also indicate gaps in institutional safeguards, including absence of Standard Operating Procedures (SOPs), weak access controls, and limited guidance on secure data handling.

Taken together, the responses highlight that while KIPRE maintains moderate levels of data quality, there are clear weaknesses in data integrity and security. This has implications not only for research credibility but also for compliance with ethical and institutional standards. Moving

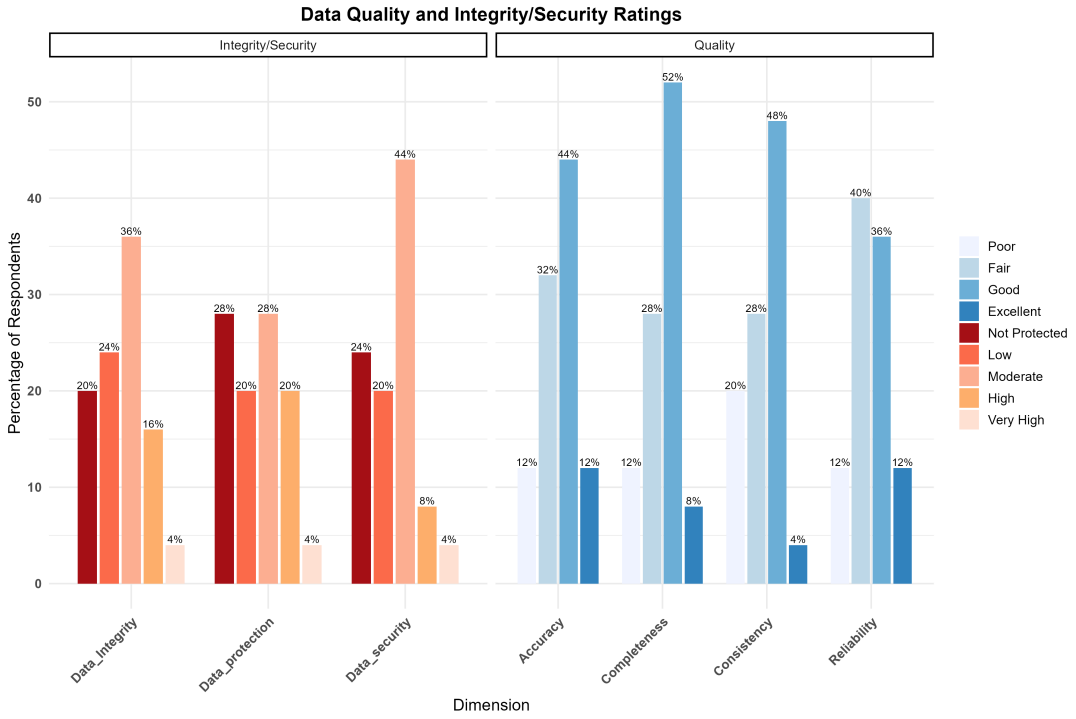

Figure 2: Survey ratings of data quality and integrity/security across KIPRE units. Ratings range from Poor/Not Protected to Excellent/Very High.

forward, DS&AS should prioritize development of data governance frameworks, strengthen SOPs for data handling, establish clear access control policies, and provide practical training on secure data management. Addressing these gaps will build confidence among staff and ensure that data at KIPRE is both high-quality and well protected.

5 Data analysis & reporting

The results (see Table 3) reveal important insights into both demand for analytical support and perceptions of data utilization in institutional decision-making. The majority of respondents reported needing analytical support *Ad-hoc or As Needed* (45.3%), followed by those requiring support on a *Monthly* basis (28.3%), *Quarterly* (15.1%), and a smaller group seeking *Weekly* support (11.3%). This pattern suggests that most staff interact with data reactively, often in response to donor reporting deadlines, management requests, or project-specific outputs, rather than through continuous and structured analytical processes. The limited proportion of staff requiring weekly support further indicates that routine, data-driven operational management is not yet widely institutionalized.

When this is compared with perceptions of data use in decision-making, the results highlight a disconnect between data generation and its translation into institutional action. While 73.1% of respondents believe that data is used in decision-making, a significant 25% remain uncertain, suggesting limited visibility on how data flows into decision processes. More critically, even among those who frequently request analytical support, a notable proportion are unsure whether

their analytical outputs translate into actionable decisions. This indicates that for many users, current analytics are report-oriented—produced to satisfy information requests—rather than decision-oriented, driving program learning or strategic planning.

Table 3: Analytical Support by Data Utilization (Row Percentages)

Analytical Support	Maybe (%)	No (%)	Yes (%)	Total (%)
Ad-hoc / As needed	20.8	4.2	75.0	100.0
Monthly	11.1	0.0	88.9	100.0
Quarterly	50.0	0.0	50.0	100.0
Weekly	100.0	0.0	0.0	100.0
Total	25.0	1.9	73.1	100.0

Further analysis of demand by type of analytical output (Table 4) reveals varied levels of analytical maturity across the institution. Requests for standard reports (69.2%) and simple ad-hoc analyses (54.5%) dominate among those who engage with data only when needed. This underscores a reactive analytics environment in most departments, with analysis performed mainly to meet accountability requirements rather than to drive insights or improve performance.

Table 4: Analytical outputs by frequency of analytical support needed

Analytical Output	Ad-hoc (%)	Monthly (%)	Quarterly (%)	Weekly (%)
Ad-hoc Analyses / As needed	54.5	36.4	9.1	0.0
Dashboards / Visualizations	50.0	30.0	20.0	0.0
Exploratory models	27.3	45.5	18.2	9.1
Predictive models / forecasts	14.3	42.9	28.6	14.3
Standard reports Docs/PDFs	69.2	23.1	7.7	0.0
Total	43.4	35.8	17.0	3.8

Departments that engage monthly with the analytics team demonstrate more advanced analytical behaviors, with higher demand for exploratory modelling (45.5%) and predictive analytics (42.9%). This group is transitioning from descriptive reporting towards diagnostic and forecasting insights, signaling growing analytical capability. On the other extreme, while a small subset (3.8%) of respondents reported weekly analytical engagement, their demand is focused on advanced analytics, indicating pockets of innovation where data is actively leveraged for operational responsiveness.

Taken together, the findings show that analytical demand across the institution is increasing but remains largely unstructured and reactive, with limited integration of analytics into decision systems. While there is emerging analytical capacity in some programs, significant institutional gaps remain in data governance, analytical planning, automation, and feedback mechanisms. A structured framework is needed to transition the organization from fragmented reporting practices

to a sustainable evidence-driven culture. This will require standardizing analytical processes, strengthening human capacity, institutionalizing dashboarding and automation to reduce ad-hoc workloads, and linking analytics explicitly to program review and strategic decision forums.

6 Staff capacity, analytical outputs, and training needs

An assessment of staff engaged in data activities across divisions reveals an uneven distribution of analytical capacity. Table 5 shows that divisions such as Veterinary Clinical & Diagnostic Services have substantial intermediate-level capacity (16 staff) alongside beginner-level personnel (15), whereas divisions like Evolutionary Ecology Conservation & Ecosystem Health and Information Communication Technology report minimal or no staff actively working with data. Infectious Diseases & One Health, although having only advanced-level staff, lacks intermediate and beginner-level personnel, highlighting gaps in layered analytical expertise across the institution. These disparities imply that operational decision-making may be heavily reliant on select individuals, risking bottlenecks and inconsistent data use.

Table 5: Staff Distribution by Division and Level of Expertise

Division / Section	Beginner	Intermediate	Advanced
Corporate Communication	4	0	0
Evolutionary Ecology Conservation & Ecosystem Health	0	0	0
Infectious Diseases & One Health	0	0	3
Information Communication Technology	0	0	0
Non-Communicable Diseases	4	0	0
Planning & Strategy	2	0	0
Reproductive Health & Biology	0	0	5
Veterinary Clinical & Diagnostic Services	15	16	0

The analysis of analytical outputs indicates that most divisions rely heavily on standard reporting and ad-hoc analyses, with advanced outputs such as exploratory and predictive models concentrated in a few divisions with consistent analytical engagement. This pattern suggests that while some units are gradually moving toward structured, insight-driven workflows, the majority still operate in a reactive, report-oriented mode. Only a small proportion of divisions demonstrate high-frequency engagement with predictive analytics, indicating that real-time decision-making using complex data is limited institution-wide.

The distribution of training needs, summarized in Table 6 and illustrated in Figure 3, reinforces these observations. Technical skills such as data analysis and statistical modelling (10%), data security and compliance (9.2%), and machine learning/AI methods (8.8%) dominate the training demand. Related software competencies—including Python, R, SAS, and SQL—collectively account for over 30% of identified needs, highlighting the breadth of technical up-skilling required. Operational skills such as data management, project management, research design,

and communication constitute another significant portion, underscoring that analytical capacity is not solely technical but also managerial and interpretive.

Table 6: Percentage of Training Needs Across Staff

Training Needed	Percent (%)
Data analysis and statistical modelling	10.0
Data security and compliance	9.2
Machine learning / AI methods	8.8
Python	7.6
R	7.6
SAS	7.6
SQL	7.6
Other softwares	7.6
Data management and governance	7.2
Project management and coordination	7.0
Research design and methodology	6.7
Communication and knowledge translation	6.3
Data visualization and reporting	5.7
None	0.8
Bioinformatics	0.2
All of the above	0.2

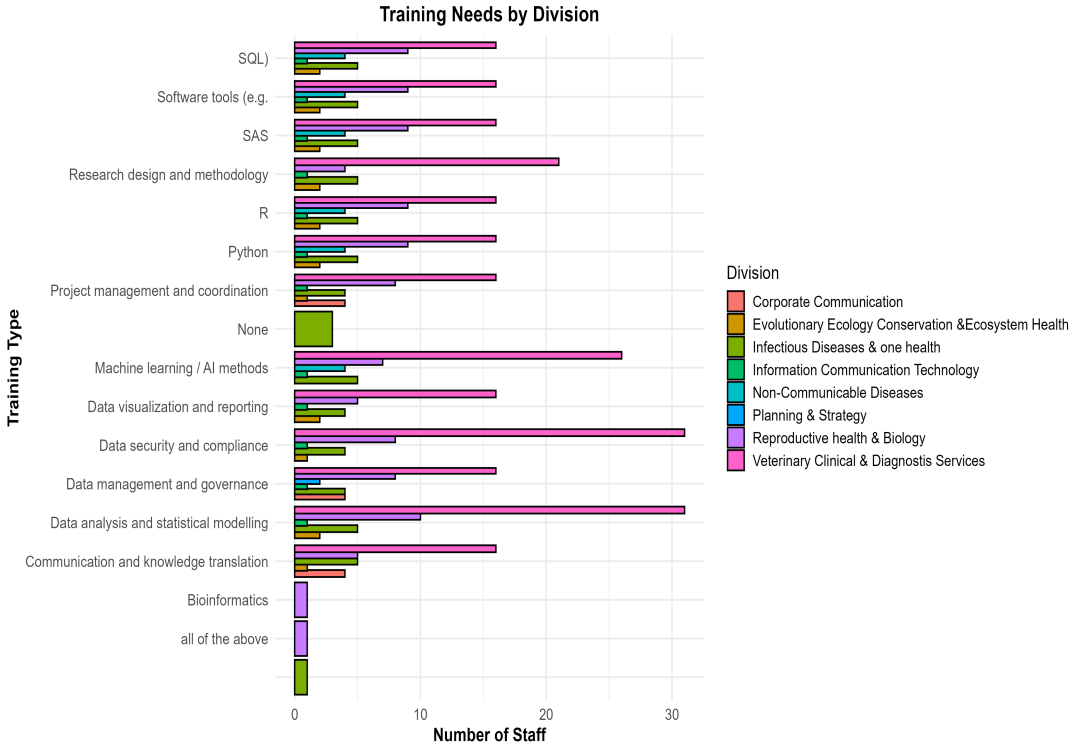

Figure 3: Training needs by division or section

The integrated view of staff capacity, analytical outputs, and training needs reveals three critical

insights. First, analytical capacity is unevenly distributed, with some divisions heavily reliant on a few advanced staff, while others have minimal or no engagement. Second, skill gaps are substantial and span technical, methodological, and operational domains, reflecting a need for structured, comprehensive training. Third, the existing analytical workflow is predominantly reactive, focusing on standard reports and ad-hoc requests rather than systematic, evidence-driven decision-making. Addressing these gaps requires coordinated interventions: targeted training aligned with division-specific needs, development of structured analytical support frameworks, investment in workflow automation, and mechanisms to link data analysis to operational and strategic decision-making. Such measures are essential to move the institution toward a mature, sustainable data culture that supports consistent and actionable insights across all divisions.

7 Staff capacity, training needs, and data governance

The assessment of data security, governance, and policy awareness among staff reveals several critical gaps and highlights priority areas for capacity building (see Table 7 below) . Key security challenges reported include inadequate staff training on data protection (18%), limited infrastructure or secure storage systems (18.9%), risks of data breaches or unauthorized access (15.3%), and lack of clear institutional policies and guidelines (15.3%). These findings indicate that staff capacity to maintain secure and compliant data practices is inconsistent, exposing the Institute to operational and regulatory risks.

Table 7: Data Security, Governance, and Policy Awareness Among Staff (%)

Key Security Challenges	**SOP Documentation Status**	**Awareness of Data Sharing Policy**
Challenges in managing consent and confidentiality: 9.9	Yes: 24	Yes: 32
Difficulty complying with regulations: 8.1	No: 48	No: 52
Inadequate staff training: 18	Maybe: 28	Maybe: 16
Lack of clear policies: 15.3		
Limited infrastructure: 18.9		
Limited resources: 13.5		
Risks of data breaches: 15.3		
Lack of understanding of org structure: 0.9		
Awareness of DPAct	**Perception of SOP Documentation**	**Gaps in Governance / Security**
Yes: 56	Yes: 4	Backup systems: 36.5
No: 24	No: 60	Access control / permissions: 30.2
Maybe: 20	Maybe: 36	Staff training and sensitization: 33.3

Regarding formal governance structures, only 24% of staff confirmed the existence of documented policies or standard operating procedures (SOPs), while 48% reported their absence and 28% were unsure. Similarly, awareness of data-sharing policies and the Data Protection Act (DPAct) is moderate, with 32% and 56% of staff, respectively, reporting familiarity, while others remain unaware or uncertain. This demonstrates a clear need for targeted training and sensitization to ensure all staff understand the legal and ethical frameworks guiding data handling.

Perceptions of SOP documentation remain low, with only 4% confirming active documentation, suggesting that formal guidance is either insufficiently developed or poorly communicated. Identified gaps in governance and security are concentrated around backup systems (36.5%), staff training and sensitization (33.3%), and access control mechanisms (30.2%), highlighting concrete areas where interventions could strengthen data integrity.

Overall, these findings underscore the importance of comprehensive capacity building, development and dissemination of SOPs, investment in secure infrastructure, and ongoing awareness campaigns. Addressing these gaps is essential for ensuring that staff can manage sensitive information securely and responsibly, thereby enhancing the Institute's overall data governance framework.

8 Support from DS&AS and opportunities for collaboration

The survey responses reveal a clear consensus among staff on the crucial role of the DS&AS in strengthening data management, analysis, and decision-making across the Institute. Respondents consistently highlighted the need for structured engagement with staff to identify gaps, streamline processes, and provide technical guidance. Many emphasized the importance of collaborative initiatives, particularly in sections handling complex scientific data, including bioinformatics, virology, and ecological research, where DS&AS expertise can directly enhance research quality and efficiency.

Staff expressed a strong desire for training and capacity building, ranging from basic data literacy and software usage to advanced analytics, including statistical modeling, data visualization, and machine learning. Several respondents noted the value of mentorship programs and regular data review sessions to foster sustainable skills development. Centralized access to data tools, clear SOPs, and assurance of data confidentiality were also recurrent themes, underscoring the need for organized support structures that enable all staff to utilize data resources effectively.

A notable concern was the handling of sensitive and personal data, including biometric information such as personnel profiles, facial recognition identifiers, and other unique personal data. Respondents highlighted that robust data governance frameworks, encryption protocols, and strict access control mechanisms are essential. Clear guidelines on data ownership, consent, sharing policies, and continuous staff training were suggested to ensure ethical and compliant data use.

Furthermore, respondents indicated that DS&AS could provide practical, project-focused support, such as designing digital data collection tools, automating data cleaning and integration pipelines, and developing interactive dashboards for real-time decision-making. Some highlighted smaller projects that currently lack dedicated support, emphasizing the importance of inclusive strategies that benefit both high-profile and routine research efforts.

Overall, staff see DS&AS not merely as a service provider but as a strategic partner, capable of integrating data science expertise into research, governance, and operational workflows. By addressing gaps in training, governance, and data infrastructure, the section can ensure that both research and administrative units benefit from a more data-driven, secure, and collaborative institutional environment.

9 Conclusion

Taken together, this needs assessment provides a focused and actionable picture of KIPRE's current data landscape. Responses show a clear concentration of analytical capacity within core research units, a predominantly reactive demand for analytics (heavy reliance on ad-hoc requests and standard reports), and substantive gaps in data governance and security — most notably in staff training (18%), secure infrastructure (18.9%), access control and backup arrangements. Training priorities cluster around statistical modelling, data security and compliance, and software proficiency (R, Python, SAS, SQL), while operational skills in data management and communication remain important support needs. On this evidence, DS&AS should prioritize (i) a structured analytical support framework that moves teams from one-off requests to scheduled, programmatic engagement; (ii) a focused capacity-building program that pairs technical training with practical, project-based mentorship; and (iii) clear, disseminated SOPs together with targeted investments in access controls and backup systems so that improved analytical outputs are matched by secure, auditable data practices. The findings should be treated as indicative rather than definitive because the response rate was modest (25 responses, 16.4%); nevertheless the patterns are consistent and sufficient to inform near-term priorities and pilot interventions.

The assessment itself has several strengths that increase confidence in its conclusions and utility for planning. First, the instrument captured both quantitative proportions and open text, enabling triangulation between numeric summaries and qualitative insights; this produced nuanced evidence about where needs are practical (tool training, SOPs) versus structural (infrastructure, governance). Second, participation was concentrated in the divisions most actively handling scientific data, so the results reflect the lived experience of the primary data users and therefore point to high-impact interventions. Third, the analysis was transparent and reproducible: responses were cleaned, multi-response fields parsed and summarized, and results presented in clear tables and figures that directly map to recommendations. Finally, the report's emphasis on operationally specific actions — SOPs, scheduled support, pilot automation, and division-level training — makes the findings readily implementable by DS&AS and partners. Together, these

strengths mean the assessment can be used immediately to design a staged program of work: pilot capacity building in research divisions, co-develop SOPs with ICT and program leads, and roll out monitoring to measure uptake and impact before broader scale-up.

Acknowledgments

We thank all KIPRE staff who generously took time to complete the needs assessment. Your candid responses and practical suggestions made this analysis possible and provide a solid basis for targeted improvements in data practice, governance and capacity building. We are especially grateful to the Director General for his support and encouragement of this exercise, and for taking time from his busy schedule to complete the questionnaire. His leadership in championing stronger data systems across the Institute is much appreciated. DS&AS will use these findings to prioritize pilot interventions, co-develop SOPs with program and ICT leads, and deliver practical training closely aligned to day-to-day work.

Patrick Mwaura
Data Science & Analytics Section
Kenya Institute of Primate Research
October 13, 2025