The Use of Technology to Enhance Compliance to Data Protection Regulations in Organizations

Esther Maina

United States International University-Africa, Kenya

Email: ewambuim@usiu.ac.ke

Paul Okanda

United States International University-Africa, Kenya

Email: pokanda@usiu.ac.ke

ABSTRACT

Data protection regulations implement legal obligations to ensure data security while processing personal data. The main aim of the regulations is to ensure accountability of those who collect information. Organizations are not only required to implement processes that are compliant to these regulations but also be able to document their compliance. This paper explores the impact of data privacy and security laws on organizations. This is motivated by the need to ensure data confidentiality as highlighted in the privacy laws. Specifically, the study is aimed at identifying the challenges Kenyan organizations face in achieving compliance to the Kenya Data Protection Act 2020 and how they can utilize emerging technologies such as data analytics to assist in compliance to the said regulation. The research findings discussed in this paper propose strategies that can help promote the adoption of data analytics as a strategy to monitor compliance of the regulation. The proposed strategies are evaluated using an analytics-based prototype that utilizes key performance indicators to monitor risk compliance. The results of the study highlight the need of integrating data analytics in the risk assessment of personal data maintained by an organization this enhancing the compliance of the Act in organizations.

The study investigates the impact of data privacy and security laws on organizations that collect or utilize personal data to maintain normal operations. This is motivated by the need to ensure data

confidentiality as highlighted in the privacy laws. The study also focused on validating and testing the solution to show how data analytics will be useful in helping organizations to be compliant with the Data Protection Act of Kenya.

Key Words: Data Analytics, Risk Assessment, Data Protection Act of Kenya, Key Performance Indicators, Personal Data, Compliance

1.0 INTRODUCTION

Personal data has gained significance in past few decades and is being used in different ways. Despite the legal restrictions on processing personal data, data subjects view the extensive collection of data with distrust. Legal battles between data subjects and processors have shown the need to redraft regulations to face today's privacy challenges hence leading to the introduction of data protection laws and regulations. According to a report by the Law Insider (Insider, 2020), data protection laws are legislation, regulation or code of practice which relates to the protection of individuals with regard to the processing of personal data. Data privacy and protection laws provide a legal framework to guide organizations on how to obtain, use and store private data. The data protection laws highlight the rights of an individual to control the use of their data and who has access to it. They also have the right to be informed on how their data is stored, the purpose for its use and they have the right to request its deletion.

Majority of countries worldwide have adopted comprehensive data protection regulations into their laws. The European Union introduced the General Data Protection Regulation (EU, 2018). The GDPR legal framework introduced a new era in global data protection and privacy by setting new regulations that apply to any organization that handles the data belonging to citizens of any European Union (EU) nation. In Kenya, the Data Protection Act of Kenya was passed into law in 2019 (No.181, 2019) and is designed to enhance the protection of personal data in Kenya. The Act places restrictions on the use of personal data, while strengthening the data subject's rights. Institutions that require the collection of personally identifiable data to be collected from participants are required to provide assurance through the research ethics process. They will need

to show that the processing of the data will adhere to the Act. An institution must therefore notify individuals how their data will be used. A Participant Information Sheet should be provided to participants and should contain the purpose of the data collection and processing and indicate who will have access to their information (Innovation, 2019). The institution must also maintain confidentiality of data and protect the identity of individuals and ensure that data is accurate and up to date. In case of collaboration with firms outside of Kenya, data may not be transferred to countries unless the firm has provided evidence of data protection regulations in place, the participants have provided explicit consent for the transfer data and the firm has a signed contract specifying data protection requirements that have been put in place.

This presents a need to equip Kenyan organizations, data controllers and producers with the necessary skills to adhere to the Act. Questions arise on how organizations can implement the Act given that the regulations were not imposed previously and there was no regulatory body to provide guidance. The Act also hinders implementation of fraud and cyber security management since the Act provides a leeway for hackers to mask their information such as Internet Protocol (IP) addresses. The regulation stipulates that personal data can't be transferred outside Kenya unless there is proof of adequate data protection safeguards or consent from the data subject (Section28(b), 2019). The Act however does not provide clear guidelines for organizations and individuals that require transfer of data outside of Kenya and how the different jurisdictions can cooperate in case action is needed after a data breach.

This presents a need for an examination of data collection methods and processing practices within an organization. The major reason for this is the blur that exist between what is personally identified information (PII) and what is not. A study by Paul Ohm (Ohm, 2013) shows one can link non-PII to individuals. The study revealed that one was able to identify eighty percent of 500,000 Netflix subscribers who were classified as anonymous. Their identity could be linked to their ratings in the Netflix database for at least three films. The study was able to establish a link between the data sample and movie ratings made in the Internet Movie Database (IMDb) by the same individuals. The study has drawn on this research and other studies to argue, "No matter

what the data administrator does to anonymize the data, an adversary with the right outside information can use the data's residual utility to reveal other information".

Organizations that collect personal data are required to comply with a wide range of data privacy regulations due to the cross-border nature of personal data transactions, and extra-territorial nature of various data privacy regimes in order to maintain the confidentiality, integrity and availability of data. This means that the costs that companies would need to expend to comply with such standards could potentially be high. Collecting personal data from many individuals also means that the organizations will need to be tasked with ensuring data subjects' requests on how their personal data is used. Organizations that suffer a data breach and are found to have inadequate security measures are susceptible to penalties from regulators. The negative publicity from data breaches could have a huge impact on the company's reputation that could result in individuals refraining from providing their personal data to the company therefore negatively impacting the organization's operations and revenue streams. This paper focused on the risks that personal data creates to privacy and the kinds of cautionary processes that should be taken to comply with data protection laws.

2.0 OBJECTIVES AND IMPORTANCE OF THE RESEARCH

The main objective of this research was to identify the challenges research institutions face in achieving compliance to the Kenya Data Protection Act 2020 and to develop a conceptual prototype that utilizes data analytics to guide institutions on how to integrate the act in their data protection strategy.

This paper aimed to achieve the following objectives:

- I. Review of the challenges faced by organizations in adopting compliance to the Data Privacy Act of Kenya.
- II. To demonstrate how data analytics can be used to monitor risk compliance using key performance indicators to help organizations adhere to the Data Privacy Act.
- III. To analyze the effectiveness of integrating data analytics to monitor compliance metrics through selected verification and validation processes.

The paper highlights previous literature on the research topic, legislation in various jurisdictions such as the GDPR and the Data Protection Act of Kenya, commentaries on privacy and data protection from industry players, working papers of international bodies on privacy and data protection, just to name a few. In doing this the study aimed to meet the research objectives. The study highlighted the principles of data processing stated in the GDPR and how they are implemented. Finally, the study focused on how the elements of data analytics can help an organization to become compatible with the principles of data processing under the GDPR. This study will contribute to research focusing on the measures organizations must put in place to comply with data protection regulations specifically the Data Protection Act of Kenya. The topic has not been comprehensively undertaken, despite the law being introduced in November 2019.

3.0 LITERATURE REVIEW

Various studies have been conducted to investigate the relationship between data analytics and compliance to data protection regulations. Some studies focus on the benefits and risks of embedding analytics in adopting data protection strategies. For instance, (Viktor Mayer-Schönberger, 2016) studies the restrictions GDPR place on processing of personal data and how data analytics practices such as data minimization can help in masking personally identifiable information (PII) to allow processing of personal information while maintaining security and privacy of the data subject. Norjihan Abdul Ghani et al. (al, 2016) focus on how data analytics through pseudonymization and anonymization can help repurpose data which encourages data reuse while in the limits of the restrictions imposed by the GDPR. They pointed out that data analytics creates new data that does not require consent from the data subjects as PII has already been removed and if done correctly, might not be linked back to the original data.

A study by Zarsky (Tal, 2014) argues that utilizing big data strategies can be useful in enforcing the regulations within an organization and prevent fines imposed due to data security and privacy challenges. Some literatures suggest that there is need of a balance of the use of data analytics with data protection laws. For instance, Tene and (Polonetsky, 2013) argue that a data analytics approach should be able to accommodate the benefits of big data for businesses and the data

subjects' right to privacy. They argue that such a strategy will help determine the justification of processing data based on legitimate business interest through provision of data subjects' consent to process the data.

Risk management is a critical tool for complying with data protection laws and ensuring that data is processed appropriately, and the fundamental rights and interests of individuals are protected effectively. Risk management does not alter rights or obligations. Rather, it is a valuable tool for calibrating the implementation of and compliance with privacy requirements (Leadership, 2014). Risk management involves three key elements: identifying and assessing negative impacts; avoiding or mitigating those that cannot be justified by the benefits of positive impacts; and finally accepting and managing the remaining risks. This paper addresses the role of risk management in data protection and highlights the growing consensus around risk management as an essential tool for effective data protection and addresses key considerations that affect the role of risk in data protection law and practice.

A study by Hamed et al. evaluated how users' activities are tracked when they are online and the threat this poses to their privacy (Hamed, 2013). Their findings showed the various occurrences of different racking components such as cookies, local shared objects (LSOs), JavaScript and iframes on the users' visit to certain websites. Websites that are not compliant with the privacy protection laws do not inform or seek the users' consent before setting cookies. Compliant websites that seeks the user's consent will confirm to the user that cookies can be used for different purposes ranging from tracking to personalization. By setting an unauthorized cookie set on a user's computer, it may be impossible for such a user to know, resulting in a possible exposure to privacy concerns arising from the use of cookie for tracking such individual's online activities.

This study aimed to examine the arguments included in previous literature to fulfil the research objectives. However, these studies are not specific to the Data Protection Act of Kenya and how it can be adopted in Kenyan institutions. The findings, therefore, do not automatically apply to the Kenyan context. The provisions on the Act will be used to examine how data analytics can be useful in helping organizations to use personal data without compromising privacy. The study also looked at analytics and how it can be used in purpose limitation, data retention and minimization

and risk assessment to create a vantage point for organizations looking to comply with the data protection regulations.

4.0 METHODOLOGY AND IMPLEMENTATION

The study was supported by a conceptual prototype that utilizes key performance indicators to monitor risk compliance to help organizations adhere to the Act. The study and the proposed solution contribute to research on compliance to the Data Protection Act of Kenya. The research used the design science methodology as its research design. Research design aims to ensure that the evidence obtained helps us to answer the initial questions unambiguously (Broadhurst, 2012). Obtaining relevant evidence involves specifying the type of evidence needed to answer the research question, testing a theory, and evaluating a program to accurately describe some phenomenon. The third objective of this study was to analyze the effectiveness of integrating data analytics to monitor compliance metrics through selected verification and validation processes. Therefore, a prototype was developed in line with the said objective. Research science aims to provide instructions in the design of the prototype for actions that are applicable to the study. To demonstrate the usefulness in using data analytics in risk assessment and key performance indicators to monitor compliance of organizations, the study revolves around the compliance of research institutions to the Data Protection Act of Kenya, which include the Public and Private Universities. The study focused on USIU-Africa students especially in graduate and post graduate programs. Most students at this level are required to carry out research or experiments for their thesis. Therefore, these students will have to regulate their data collection methods to be compliant with the data protection act of Kenya.

The research used a holistic approach for the evaluation of the artifact using evaluative techniques that use logic and simulations. The methodology was evaluated using a simulation of a research institution. The study evaluated the effectiveness of data analytics methods in helping a researcher comply with the data protection regulations while conducting research by assessing processes from data collection to archival. On data collection we assessed how a researcher might use data analytics methods such as pseudonymization or anonymization to mask personal data as per the regulation. While processing the data, we review how data analytics methods and

what KPI metrics can be used to manage data subject rights. For data analysis the study focused on data analytics methods to monitor limits on data processing and consumer profiling such as data minimization. And finally, on data archival, the study focused on analytical methods to monitor the data lifecycle and flow of data.

4.1 SYSTEM REQUIREMENTS

From the problem analysis, the tool's requirements were guided by the challenges faced by researchers in being compliant with data protection regulations while carrying out their research. The prototype's objectives are used to guide the system requirements for the prototype. Using the principles highlighted by the Act, the study analyzes the key procedures to ensure data processing compatibility with the principles of the Act.

The study focused on the analysis of the following principles of the Data Protection Act: purpose limitation, data minimization, accuracy, integrity, and confidentiality. The principles guided the requirements of the prototype and their implementation within the scope of research. The proposed prototype aims to implement data analytics components such as key performance indicators and risk assessment to comply with data protection regulations. The proposed prototype aims to help research institutions navigate past the challenges posed by restrictions introduced by the Act. Therefore, the main stakeholders of the prototype are researchers in any research institution. This study mainly focused on USIU-Africa as a research institution and the post graduate students as the main stakeholders. The prototype supported the study by showing how data analytics can help them to be compliant with the Act while carrying out research for their theses. The researchers will use the prototype to analyze data privacy risk, secure and maintain data collected during research. The system is therefore constrained to be used in research institutions for research activities by researchers and subjects of their study.

From our main objective, we identified that the major challenges faced by organizations while complying with the act were the restrictions on processing personal data. The major reason for this is the blur that exist between what is personally identified information (PII) and what is not. Recent studies show one can link non-PII to individuals. The second challenge is the management of data subject rights. The study identified four areas of data protection regulation

compliance where rule-based technologies such as data analytics may be relevant in mitigating the challenges. The areas were implemented as functionalities in the prototype. These are:

- Data Subject Rights Management: Use of KPIs and dashboards to identify, monitor and track responses to Data Subject requests to respond within the timelines provided by the regulations
- 2. Consent management where dashboards to monitor and respond to the data subject's request and consents preferences.
- 3. Risk Assessment: One can use KPIs to evaluate risk and balance the cost of keeping and securing data and responding to individual rights requests against your processing needs.
- 4. Data minimization and length retention: analytics can be used to manage data and ensure data governance procedures are adhered to.

4.2 SYSTEM IMPLEMENTATION

The developed prototype utilizes analytic methods of data masking to minimize data collected during research. The prototype allows a researcher to analyze the data collected during research to identify personally identifiable information that can be a hindrance to the data subject's privacy. The figure 1 next page shows data generated by a data generator that contains information that classifies as PII in a dataset meant for research.

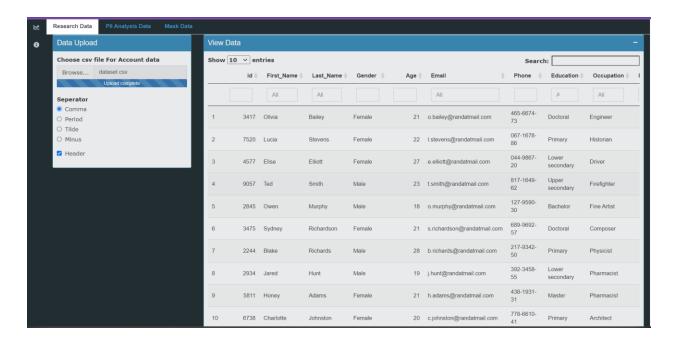


Figure 1: Sample data uploaded into prototype

The data would be subject to minimization to mask the data that can be linked to an individual in this case, the first name, last name, email, and phone number. The tool incorporated security measures to de-identify and delete personally identifiable information. Using the tool, a researcher can generate a synthetic dataset based on the original data collected and continue performing analysis on the synthetic dataset, retaining the data intelligence without the personal data.

The solution analyses the columns in the dataset to classify the data as PII or non PII. Figure 2 below shows the implementation of PII classification of data using the solution. This feature in the prototype will allow researchers to classify their data columns to know which columns require data anonymization.

The solution utilizes a masking algorithm to anonymize the data uploaded to the tool. As shown in figure 2 below, the tool masks the PII columns. The data can therefore no be linked to a particular individual and the researcher can therefore carry on with data processing within the confines of the data protection regulations.

							∳ Anonyi	mize Data					
Anonymization Results -													
Show 10 v entries											Search:		
	id ϕ	First_Name	Last_Name	Gender	Age 🔅	Email	Phone	Education	Occupation	ExperienceYears.	Salary	Marital.Status	Number.of.Chile
		All	All					A	All	All		All	All
1	3417	12213208	950e82c6	Female	21	3cd22cad	dfda4778	Doctoral	Engineer	7	1096	Married	
2	7520	2b63b22f	4d7dac61	Female	22	eef0fdb7	91ea615d	Primary	Historian	0	1259	Married	
3	4577	395bee64	b18ce503	Female	27	63f9f285	0a6d8b1d	Lower secondary	Driver	11	4980	Married	
4	9057	803275fb	e5872733	Male	23	b78153e5	3a71bcbf	Upper secondary	Firefighter	0	9317	Single	
5	2845	541f2030	dc0a4d4c	Male	18	5f3e992e	4d7665b2	Bachelor	Fine Artist	5	6452	Married	
6	3475	e5136edf	807af5d3	Female	21	a27ae9f9	157ad74a	Doctoral	Composer	8	1802	Married	
7	2244	3d6904dc	28807e81	Male	28	4691c09f	d1e8c874	Primary	Physicist	3	9251	Single	
8	2934	65387567	d988cf07	Male	19	8b46384e	03464397	Lower secondary	Pharmacist	2	5353	Single	
9	5811	e9445117	d1f3b5a8	Female	21	d237961c	fe1d420d	Master	Pharmacist	6	1946	Single	
10	6738	fee04a7d	9d75f869	Female	20	03387f3b	f444c8c8	Primary	Architect	5	6488	Married	
4													

Figure 2: Anonymization results of dataset using the prototype

The tool also utilizes data analytics to manage consent. Using Key Performance indicators, the tool utilizes a dashboard to help the user get a summary of the amount of data records being used in a processing activity. This should help the researcher get a high-level summary of the data in possession during their research and know how many respondents provided their data. The dashboard also monitors consent and when it was first granted to keep track of the ninety-day reauthentication period required by the Act to refresh access to the data. By implementing a dashboard, a user can estimate the period for which consent is valid for cases when data is for a one-off use. Figure 3 below shows how consent management has been integrated in the tool to help researchers keep track of their data subjects consent management and requests.

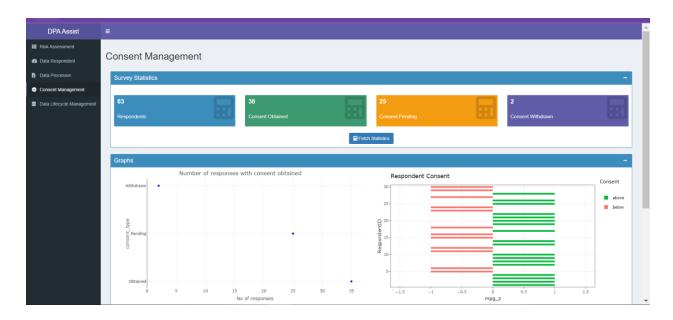


Figure 3: KPI monitoring dashboard implemented in the prototype

5.0 PRELIMINARY EVALUATION AND TESTING

The architecture evaluation approach used in this study is the Simulation-based evaluation. This method relies on a high-level implementation of some of the components in the software architecture for the purposes of evaluating quality attributes such as performance and correctness of the architecture. To evaluate the prototype, sample datasets from internet databases such as Kaggle were used to measure the efficiency of the features implemented in the tool. To evaluate the system's ability to analyze and classify PII, we simulated records for dummy individuals and observed the masked data generated by the prototype. The results of the algorithm are displayed in Figure 2. To model a dashboard showing key performance indicators, the prototype utilized data uploaded to provide metrics based on the system features that were modelled on the data protection principles. The prototype was able to compute metrics that can be used to monitor consent agreements provided and the data subject requests that were pending.

6.0 FINDINGS

The first research objective aims to identify the challenges faced by organizations in adopting compliance of the Data Privacy Act of Kenya. Privacy laws are paid more attention to due to the increased use of platforms like the internet. This introduces a need to make sure that personal data

is protected. While these protection laws pose an advantage to those transacting online, it is a disadvantage to those who are using this data and have to adhere to the strict regulations (Acharya, 2019). The Data Protection Act fails to clearly define some recurrent terms under the law such as sensitive personal data. This impacts data processing as the thresholds under which organizations can consider sensitive personal data are not clearly defined.

According to Section 18(a) of the Act, data controllers and processors are required to be registered with the Commissioner. The role of the Commissioner is to set thresholds for mandatory registration according to the nature of industry, the quantity of data being processed and considerations on whether sensitive personal data is being processed. These thresholds are not yet put in place hence one might argue that registration is not mandatory. This will hinder data processing in organizations as they would not know the thresholds in which their data falls under and which regulations they will need to adhere to.

Data that is considered sensitive is data that reveals health status, race, ethnicity, belief, genetic data, biometric information, property details, marital status, family details and sexual orientation of an individual (Section(2), 2019). Specific rules apply to the use of such data. Health data relating to an individual may only be processed by a healthcare practitioner. There is no clarity provided by the regulation on the length of time the personal data can be stored. Researchers are therefore required to apply judgment in assessing retention durations since there is no prescribed length of time stipulated in storing research data. The Act imposes strict restrictions on data being transferred outside Kenya (Section(48), 2019). Organizations might find it hard to share data outside Kenya due to this restriction. Researchers are permitted to transfer personal data to another country after providing proof to the Commissioner that the other party has put in place appropriate to protect the personal data being transferred.

The Act has not defined the processes to be followed. This might impede cross country research as there are no stipulated guidelines on the proof required. However, the act does offer an exemption where publication of data is in the public interest which are all under specific circumstances. Nonetheless, organizations should consider, as part of their processing activities,

demonstrating accountability for the protection of the data they collect and use. The institutions can appoint a Data Protection Officer to help in drafting data protection policies and ensure they are properly implemented. It would also be prudent to comply with the Act by conducting risk assessments and implementing procedures and practices as set up by the Act.

The use of both structured and unstructured data, large storage of data, collection of data without a pre-established purpose all relates to the large volume of data collected by organizations. The benefits and risks of big data become more obvious. To demonstrate how data analytics can be used to monitor risk compliance using key performance indicators to help organizations adhere to the Act, the study relied on the principles of data as included in section 25. These principles were considered as the basis of the scope of the study. It was discovered that it is upon these principles that the GDPR was built and they relay the rules of which any organization that wants to process personal data should adhere to.

7.0 CONCLUSION

The main findings of this study showed that data analytics can be integrated into an organization approach to comply with the data protection regulations within their jurisdictions. The contribution that was made through the study was that it highlights how data analytics methods can be used to monitor key performance indicators that are guided by the principles of data protection through the implementation of the prototype solution. This helps build a case for the introduction of data analytics processes in compliance strategies in organizations. For further studies, it is recommended that the study incorporates how other data analytics methods such as machine learning can be useful in helping organizations comply with data protection regulations. Using machine learning can broaden the scope to include data threat detection and further classification of data used by an organization. These additional enhancements can be included in the prototype to enhance accuracy and user acceptance.

During the study, few challenges were encountered which provided a basis for future improvements. A notable challenge is the different provisions implemented by the data protection regulations on the use of data collected and reason for processing. Although the Act creates restrictions for organizations that process personal data, there are exemptions provided for special

cases such as research. The Act exempts research from the restrictions imposed on processing personal data to allow researchers to process personal data. The Act also allows researchers to process and transfer personal data to foreign countries without providing proof of protection in specific cases. To benefit from these exemptions, researchers must put in place appropriate safeguards and maintain ethical standards which will lower the risks of data privacy breaches or act against the rights of data subjects. As a result of these challenges, additional configurations will be required to attain acceptable levels of accuracy of the prototype. For example, the metrics for the risk posed by research organizations are different from the risks posed to other organizations such as businesses or financial institutions. Future work calls for evaluating the prototype on a different organization, improving algorithms to enhance the solution, and investigating trade-offs of various features based on the scope of the data processors and controllers.

REFERENCES

- Acharya, M. (2019, November 22). *Mondaq*. Retrieved from ENSafrica: https://www.mondaq.com/data-protection/867010/data-protection-in-kenya-what-youneed-to-know
- al, N. A. (2016). Big Data and Data Protection: Issues with Purpose Limitation Principle.

 International Journal of Advanced Soft Computing and Data Mining, 116-121.
- Broadhurst, H. &. (2012). What is Research design? Explanatory/descriptive research. *Qualitative Social Work*.
- Data Protection Laws. (n.d.). Retrieved from Law Insider: https://www.lawinsider.com/dictionary/data-protection-laws
- Data, S. (2021). *Data Protection Principles*. Retrieved from SafeDataGov: https://www.safedatagov.com/data-protection-principles/
- EU. (2018). General Data Protection Regulation. European Union.

- Hamed, A. (2013). Evaluation of Third Party Tracking on the Web. *IEEE Conference Publications*, 471-477.
- Innovation, U. I. (2019). GDPR Regulations for Researchers. Retrieved from https://www.ukri.org/news/general-data-protection-regulation-guidance-for-researchers/
- Leadership, C. f. (2014). THE ROLE OF RISK MANAGEMENT IN DATA PROTECTION.

 Privacy Risk Framework and Risk-based Approach to Privacy.
- No.181, K. G. (2019). Data Protection Act, 2019. Kenya: Government of Kenya.
- Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. (UCLA LAW REVIEW 1701).
- Polonetsky, T. O. (2013). Privacy in the age of Big Data: A Time for big decisions. Northwestern Journal of Technology and Intellectual Property, 239-273.
- Section(2). (2019). Restriction of Processing. *Kenya Gazette Supplement No. 181 (Acts No. 24)*.

 Retrieved from http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24 of2019.pdf
- Section(48). (2019). Conditions for transfer out of Kenya. Data Protection Act of Kenya.
- Section 28(b). (2019). Section 28 (b). Data Protection Act of Kenya.
- Tal, Z. (2014). Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society'. (13).
- Viktor Mayer-Schönberger, Y. P. (2016). Regime change: enabling big data through Europe's new data protection regulation. *The Columbia Science & Technology Law Review*.