Integrated Personal Data Protection Tool

by

Wangui Evalyn Wanjiru

069334

Master of Science in Information Systems Security

2023

Integrated Personal Data Protection Tool

by

Wangui Evalyn Wanjiru

069334

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems Security at Strathmore University

School of Computing and Engineering Sciences

Strathmore University

Nairobi, Kenya

July,2023

Declaration and Approval

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, this dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: Wangui Evalyn Wanjiru

Sign:_____ Date: _____

Approval

The dissertation of Wangui Evalyn Wanjiru was reviewed and approved by the following:

Dr. Humphrey Njogu,
Senior Lecturer,
School of Computing & Engineering Sciences,
Strathmore University

Dr.Julius Butime,
Dean,School of Computing & Engineering Sciences,
Strathmore University

Dr.Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

The privacy of personal data is an important focus area in today's information world, where personal data is easily collected, stored, processed, and shared. In recent years personal data protection has regularly featured as a topic of concern in the media and has become the target of legislation around the country. Organizations are collecting, re-using, and processing personal data on an unprecedented scale, without observing data limits, this has led to an increasing concern about the effectiveness of the existing data protection law and the need for a comprehensive framework for personal data protection. Based on the challenges, this study was aimed at designing, developing, and testing an integrated personal data protection tool. The tool employed an integrated approach that utilizes data encryption, one-time passwords and opt-in and opt-out mechanisms to ensure secure collection, secure storage, secure access, and secure sharing of personal data. Moreover, the tool was designed, developed, and tested using the agile software development methodology. Functional and performance tests were carried out to determine the performance and accuracy of the tool in protecting personal data. Validation testing results showed that the integrated tool could be used effectively in protecting personal data with 96 % accuracy.

Keywords .

Personal data; Personal data protection; Organizations; Data subject

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AES	Advanced Encryption Standard
API	Application Programming Interface
DES	Data Encryption Standard
DFD	Data Flow Diagram
DSS	Digital Signature Standard
ECC	Elliptic Curve Cryptography
ERD	Entity Relationship Diagram
GDPR	General Data Protection Rule
GPS	Global Positioning System
IEBC	Independent Electoral Boundaries Commission
ID	Identity Document
IDE	Integrated Development Environment
IPRS	Integrated Population Register Services
KRA	Kenya Revenue Authority
KYC	Know Your Customer
MAC	Media Access Control
MNO	Mobile Network Provider
PII	Personal Identifiable Information
PIN	Personal Identification Number
OTP	One Time Pin
RSA	Rivest Shamir Adleman
TIMS	Transport Integrated Management System
USB	Universal Serial Bus

Definition of Terms

Data Controllers	Legal person or any other body that processes data on behalf of the data controller (Ensight, 2019)
Data Privacy	The ability of a person to determine when, how and to what extent personal information about them is shared with or communicated to others (Mbanaso,2018)
Data Processors	Legal person or any other body that processes data on behalf of the data controller (Ensight, 2019)
Data Subject	Identified natural person who is the subject of the personal data (Ensight, 2019)
Education Data	Personal Identifiable Information that is maintained in education records (Education, 2022)
Financial Data	Any information that a consumer provides to obtain financial product or service Poremba (2021)
Health Data	Data collected by healthcare professionals to identify a patient and determine the appropriate health care (Target, 2021)
Insurance Data	Personal data collected by insurance companies for underwriting purposes and for more accurate pricing of risks (Githaiga, 2021)
Personal Data	Any information relating to an identified or identifiable natural person (Barrett, 2019)

Acknowledgements

I would like to thank the Almighty God for granting me this opportunity to pursue and complete my Master of Science in Information System Security at Strathmore University. I would also like to thank my supervisor Dr. Humphrey Njogu for motivating me to pursue this research and guiding me through the dissertation and the defense committee who generously provided feedback and expertise. I am also grateful to my family, friends, classmates, and colleagues. Their belief in me kept me motivated the entire period.

Chapter 1: Introduction

1.1 Background to the Study

The Data protection act (2019) describes personal data as any information relating to an identified or identifiable natural person. It is often referred to as the new oil of the internet and the new currency of the digital world (Berners, 2017). Mbanaso (2018) notes that every organization with digital presence collects personal data for one reason or another and that the web technology continues to influence the way personal data is collected, processed, stored, and shared.

International data protection laws typically distinguish categories of personal data depending on how strict the information must be protected. WFP (2018) explains that personal data refers to any data that is directly or indirectly related to an individual, which includes, but is not limited to the person's name, physical address, identity number, gender, age, and financial account numbers. Conversely, sensitive personal data refers to personal data that warrants extra security and confidentiality. This includes but not limited to person's racial or ethnic origin, physical or mental health status, political opinions or affiliations, religious beliefs, criminal record, biometric data such as fingerprints and refugee displacement status. Figure 1.1 indicates examples of personal data used to identify an individual.

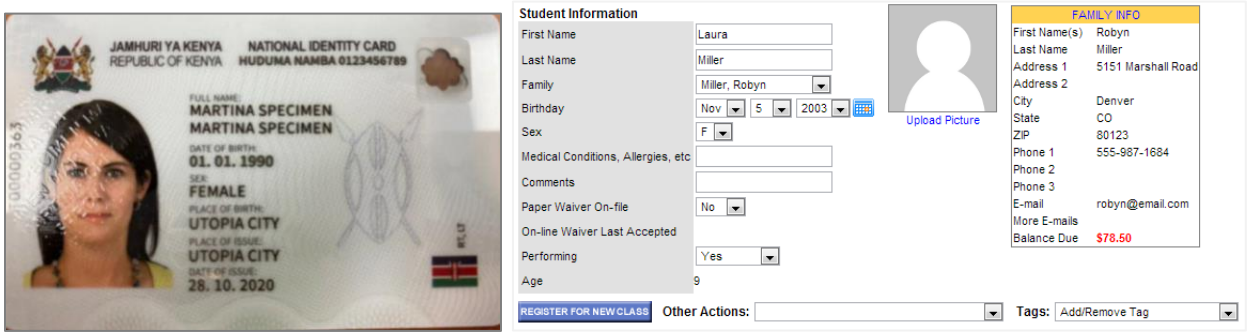


Figure 1.1:Examples of Personal Data

Source: Asamba (2020)

Initiative (2019) observes that in 2019 the Government of Kenya announced national wide biometric registration exercise to collect information for the National integrated identity management system (NIIMS). The government collected information on each person's nationality, place of birth, marital status, education background, employment status, disability status and biometrics including fingerprints and photographs. Upon registration, every registered citizen was expected to receive a unique number called 'Huduma number' which would facilitate individuals to access any government service in the future.

The roll out of this service was declared illegal because the government violated the law in processing of personal data and went against the data protection regulations (Muthoni, 2022).The main constitutional issues raised included; lack of public participation ,data privacy concerns, the right to information and that the system could further lead to discrimination of marginalized groups in Kenya (Initiative, 2019).

John (2019) notes that fintech organizations such as online lenders, microfinance providers ,mobile money service providers and traditional banks require their customers to register on the various platforms, and part of the know-your-customer (KYC) obligations, which financial entities are required to meet .Further, digital lenders share personal information of loan defaulters with third parties and other social media pages thus infringing the consumers right to privacy (Njana, 2021).

In the same context, Covid-19 pandemic created an urgent collection of personal data by the health ministry to help monitor, mitigate the social and economic effects of the pandemic and for contact tracing purposes. The data collection was conducted through phone interviews and surveys which covered; knowledge of Covid-19, mitigation measures, employment, income, food security educational activities and health services (Bank T. W., 2022). This exercise posed data privacy concerns as medical researchers might use health data of patients for secondary purposes, as well as analyze location data of patients to understand their movements and enforce social distancing rules that led to discrimination, stigmatization, unnecessary state surveillance and profiling (Malgieri, 2020).

Esteve (2017) outlines that the main privacy issues affecting the public and private sectors are lack of consent from the data subjects for the procedure of obtaining and sharing their personal data, insufficient access and control of their personal information and lack of encryption of sensitive data, while organizations are still using obsolete database systems in complex technological environments and risks of re-identifying anonymous personal data. According to Esteve (2017),privacy policies are hard to read and do not support rational decision making yet organizations impose them on their users. Their complexity makes them hard for the common user to read and understand. The lack of users consent also takes place when the organization's privacy practices change without obtaining new approval from the user.

In addition, most organizations fail to manage the risks associated with sharing personal data with third parties. Third parties and criminals who access personal data, frequently trade it on the dark web. More than twenty-two (22) billion new records of personal data are sold in the black market every year (Paro, 2021).A lawsuit was filed against a Kenyan mobile network operator for violating customers data privacy. The breach affected more than eleven (11) million customers

whose personal data was exposed including identity numbers, passport numbers, gender, age, and their betting history (Insights, 2021).Figure 1.2 demonstrates what customers think about companies using their personal data.

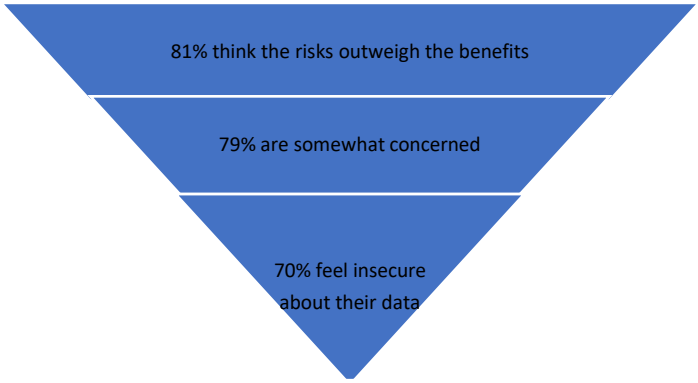


Figure 1.2: What customers think about how companies use their personal data

Source: Vizard (2019)

This study applied an integrated approach to ensure secure collection, secure storage, secure access, and secure sharing of personal data. One-time passwords were used to ensure secure access and sharing of data, and secure data encryption was employed to protect all the stored confidential data. Further, the tool enabled individuals to have more control of their data as opt-in and opt-out options were provided whenever they did not want their personal data shared for purposes it was not intended.

1.2 Problem Statement

Mostert (2020) note that in recent years, the data privacy landscape has changed because of rapid development in information technology. Organizations are collecting, re-using, and processing personal data on an unprecedented scale, without observing data limits, this has led to an increasing concern about the effectiveness of the existing data protection law and the need for a comprehensive framework for personal data protection (Absolute, 2019). Most of the customers do not understand the purpose for which the personal data is being collected or whether the data is being stored securely. For instance, mobile phone users often complain about receiving direct advertising in services they did not subscribe to and are not able to opt out from the service, (Phua, 2018). Further, enforcing instruments like data protection act of Kenya pose a challenge because organizations have not put in place mechanisms for compliance with the act. Based on the challenges, the study aimed to develop an integrated tool to guide organizations on how to process personal data more securely.

1.3 General Objective

The general objective of the study was to design and develop an integrated tool that ensures secure collection, storage, and processing of personal data based on data encryption and authorization techniques.

1.4 Specific Objectives

1. To identify categories of personal data processed by organizations.
2. To review techniques and approaches used by organizations to securely process personal data.
3. To design, develop and test an integrated tool that securely processes personal data.
4. To validate the effectiveness of the developed tool in securing the processing of personal data.

1.5 Research Questions

1. What categories of personal data are processed by organizations?
2. What techniques and approaches do organizations use to securely process personal data?
3. How will the tool be designed, developed, and tested to securely process personal data?
4. How effective is the developed tool in securing processing of personal data?

1.6 Justification

Tierney (2021) note that as organizations continue to collect massive volumes of personal data, individuals are not aware about the use of their data and do not have means to check whether their personal data is collected and processed in accordance with the data protection laws. This study applied an integrated approach to ensure secure collection, secure storage, secure access, and secure sharing of personal data. One-time access passwords were used to ensure secure access and sharing of data, and secure data encryption was employed to protect all the stored confidential data. Individuals gained more control of their data as they were provided with opt-in and opt-out options whenever they did not want their personal data shared for purposes that were not intended.

1.7 Scope and Limitation

The study focused on how organizations collect, store and process personal data, and the various approaches used to secure personal data and narrate the personal data protection principles and guidelines as stipulated by the Data Protection Act. Moreover, the study focused on ensuring

secure collection, secure storage, secure access, and secure sharing of personal data. The developed integrated tool employed RSA encryption technique to encrypt sensitive personal data, use of one-time passwords to authorize data processing as well as opt-in and opt-out mechanisms which enabled data subjects to have control of their personal data.

Chapter 2: Literature Review

2.1 Introduction

This chapter explores the literature and gives an overview of personal data and personal data protection. This is followed by detailed discussion on existing solutions and techniques that organizations use to protect their data subjects. This chapter also presents a detailed conceptual framework that illustrates the process that was undertaken in secure processing and storage of personal data.

2.2 Understanding Personal Data

Personal data refers to any information relating to an identified or identifiable natural person. An individual is identifiable if they can be distinguished from other individuals (ICO, 2019). Berners (2017) notes that a name is the most common means of identifying someone. However, a combination of identifiers may be used to identify an individual. Such identifiers include but not limited to:

a) Name: full name, maiden name, mother's maiden name or alias.

b) Personal identification numbers: social security numbers, passport numbers, driver's license number, taxpayers' identification number, financial account number, credit card numbers and patient identification numbers.

c) Personal address information: street address, email address and personal phone numbers.

d) Personal characteristics: photographic images and handwriting.

e) Biometric data: retina scans, voice signatures and facial geometry.

f) Property information such as title numbers.

g) Asset information: Internet protocols and media access control (MAC)addresses.

According to WFP (2018) personal data should be collected for specific and precise legal objectives. Furthermore, it should be used and processed in a way that is compliant with the aims for which the personal data have been collected, it should be accurate and updated, when need be, as well as be stored in a form that allows identification of the subject of the personal data and should not be stored longer than it is required.

2.3 Categories of Personal Data

Table 2.1 provides a brief description of the different categories of personal data collected by different organizations.

Table 2.1:Categories of Personal Data

Category	Description
Health Data	This is personal health data collected by healthcare professionals to identify a patient and determine the appropriate health care (Target, 2021)
Education Data	Education (2022) explains that education data is Personal Identifiable Information that is maintained in education records.
Financial Data	Poremba (2021) notes that financial data refers to any information that a consumer provides to obtain financial product or service.
Insurance Data	Insurance companies collect personal data for underwriting purposes and for more accurate pricing of risks (Githaiga, 2021).

Hofmann (2021) further notes that sensitive personal data is category of personal data that requires extra protection, it involves data that reveals a person's race, health status, ethnic social origin, conscience, belief, property details, marital status, mental health, political opinions or affiliations, religious beliefs, criminal record, biometric data such as fingerprints and refugee displacement status.

2.4 Personal Data Processing Life Cycle

Ng (2020) explains that personal data processing life cycle is a high-level process that describes how data flows within an organization. Figure 2.1 presents the phases of personal data processing lifecycle.

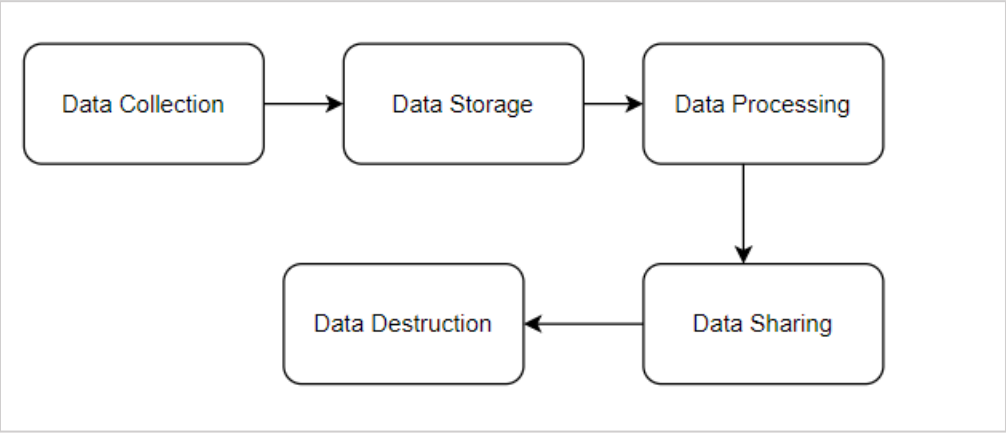


Figure 2.1: Personal Data Processing Lifecycle

2.4.1 Data Collection

In this phase, organizations must obtain consent prior to active and passive data collection. Ahmad (2017) note that active collection is when the data subject is aware of the data collection while

passive collection is when the collection of personal data is done through monitoring user activity or movements that could reveal a behavioral pattern.Beside consent,organizations are required to minimize the collection of data and collect only the personal data required to carry out the purpose for which it was intended (Ng, 2020).

2.4.2 Data Storage

According to Sheldon (2021) personal data must be stored in a secure environment to ensure its integrity,security and protection. Sheldon (2021) further notes that in this phase,data is encrypted,compressed and cleansed to ensure its availability,reliability and redundancy.

2.4.3 Data Processing

In this phase, Dataworks (2021) notes that personal data is accessed ,modified and used to support activities of the organizations such as collaboration,business intelligence and decision making.Further, an audit trail is maintained for all the critical data to ensure that all data modifications are traceable.

In addition, processing of personal data applies to all the ways in which different parties handle personal information (Newbanking, 2021). This includes collecting, recording, organizing, storing, editing, altering, searching, sharing, transmitting, and deleting information. Govuk, (2019)

2.4.4 Data Sharing

This phase consists of activities that prepare the previously accessed and retrieved personal data. Alshammari (2018) notes that this stage also involves disseminating the processed data to external actors such as third parties for historical ,statistical or scientific purposes.

2.4.5 Data Destruction

According to Assolini (2015), this stage involves a set of activities that include removing personal data from the storage in accordance with the Data Protection Act.These activities also include permanently deleting the personal data or destroying the digital storage media. Alshammari (2018) notes that the most important point in this stage is the permanent destruction,disposal and erasure of personal data.

2.5 Personal Data Privacy

In many jurisdictions, data privacy is considered a fundamental human right, and data protection laws exist to guard this right. Mbanaso (2018) explains that personal data privacy means the ability of a person to determine when, how and to what extent personal information about them is shared with or communicated to others. Data processors and data controllers often need to collect and

store personal data about their users to provide services. However, they may exceed users' expectations for data usage and collection. According to Ashford (2020) data processors may not place adequate safeguards around the data they collect, which can lead to compromise of data subjects' privacy and confidentiality. In the same context, it is acknowledged that personal data collected, processed, and stored and transferred without consent of the data subject. This violation of privacy has generated serious concerns across the globe as current privacy protections fail to promote data confidentiality.

Further, Mbanaso (2018) outline six axioms that strengthen the understanding of data privacy as a phenomenon. These axioms are explained as follows:

1) Who identifies the entities that collect personal data

2) What categorizes the personal data collected

3) Why connotes the purpose of collection or processing.

4) Where depicts the platform or the method used in the collection, sharing or processing.

5) When classifies the instance of collection or the situation under which the data is collected, processed, or shared.

6) How to identify the channel or means of collection, processing and sharing.

Table 2.2 depicts the personal data privacy map based on the six axioms.

Table 2.2: Privacy Data Map

Personal Data Privacy Map					
Who	**What**	**Why**	**Where**	**When**	**How**
Corporate organizations Government Learning Institutions Hospitals Others	Biodata Historical data Location data Career info Political views Health data	Marketing/Advertisement Political Recruitment Fraud Access Control National Security User Experience	Government offices Access Control Social Network Sites Personal Devices Apps	Account creations Online application forms Social Interactions Banking Online shopping	Filling forms Enrolments Registrations Tracking Opt-in /opt-out

2.6 Legal Framework

International (2018) observe that the right to privacy is a fundamental right that is enshrined in many constitutions around the world. Protecting privacy in the current digital space is important for effective democratic governance and income generation.

Initiative (2019) notes that there are few controls in place and there is little transparency in the decision-making processes regarding personal data. The rule of law and the responsibilities of the private sector are unregulated and there is often little public engagement and oversight of how personal data is processed and utilized, thus the need for personal data protection.

Ensight (2019) notes that the Data Protection Act is law that is designed to protect personal data in the modern society. It further controls and regulates how personal data is shared and processed by government and organizations. The Data Protection Act 2019 was enacted and became effective in November 2019.The objective of this act is to regulate the processing of the personal data; it ensures that the processing of personal data complies with the constitution and provides data subjects with the rights and remedies to protect their personal data from any form of processing that is not in accordance with the Constitution.

According to Ensight (2019) the act describes data controllers as any natural, legal person, public authority or any other body that has the power to determine the purpose and means of processing data. It further describes data processors as natural, legal person or any other body that processes data on behalf of the data controller while data subjects are described as identified natural person who is the subject of the personal data.

Baguma (2018) observe that to enforce protection of personal data, data controller or data processor must identify any foreseeable internal and external risks to personal data and establish appropriate measures against the identified risks as well as ensure that the safeguards are continually updated to address new risks.

The Data protection act of Kenya outlines the rights of data subjects as follows:

a) Right to be informed of the use of their personal data.
b) Right to access their personal data which is in the custody of data processor.
c) Right to object to the processing of all or part of their personal data.
d) Right to correction of any false or misleading personal data about them.
e) Right to deletion of false or misleading data about them.

2.7 Grounds and Principles of Data Protection

Ensight (2019) notes that consent is the primary lawful ground of processing personal data. Consent is any freely given, specific, informed, and unambiguous indication of the data subjects wishes. It can be a statement or clear affirmative action which signifies agreement to the processing of personal data relating to him or her (Ensight, 2019). In this case, data subjects are required to tick unchecked boxes to the consent and sometimes they are provided with written statements to append their signatures to show explicit consent.

Belal (2017) note that the Data Protection Act 2019 outlines the guiding principles for processing personal data in accordance with the constitution of Kenya. To comply with this policy, information must be collected and processed fairly, stored securely and not disclosed to any unauthorized users. The principles applied in the Act are based on best practices in data protection globally. Table 2.3 describes the principles of data protection.

Table 2.3:Principles of Data Protection

Principle	**Description**
Fairness, Lawfulness and Transparency	Personal data must be processed lawfully, fairly and in transparent manner in relation to the data subject (Ensight, 2019).To enable fair processing in respect to the data subject, the processing must have a legal or legitimate basis, data subjects must be informed the purpose for processing the data as well as the name of the data processor.
Purpose Limitation	Ensight (2019) note that personal data must be collected and processed for specific and explicit purposes and should not be processed in a manner that is incompatible with those purposes.
Data Minimization	Assolini (2015) explain that before processing personal data, data controllers are required to determine the extent of processing of personal data to establish the purpose for which the data was required and purpose for data collection.
Storage Limitation	Insights (2021) recommend that personal data should not be kept for longer than is necessary and for the purpose which it was collected. Data stored for historical purposes should have adequate controls against access use by unauthorized persons.

Accuracy	According to Assolini (2015) data must be correct, complete, and up to date. Consequently, data processors are expected to take appropriate measures in ensuring that inaccurate data is discarded, corrected, or updated.
Confidentiality	Assolini (2015) observe that to retain confidentiality and integrity, personal data should be preserved by establishing organizational and technical measures to prevent unauthorized access, processing, and distribution of data as well as its modification.
Accountability	Data processors are responsible for protecting personal data and are expected to comply with the principles of data protection (Ensight, 2019).

2.8 Personal Data Protection Techniques

2.8.1 Data Masking

Rabi (2022) explains that data masking is a high-level security feature used to mask plain text data. The process of data masking is designed to de-identify any sensitive data such that the data is not in human readable format. Rabi (2022) further notes that in the data masking process, sensitive data is obfuscated and replaced with special hash and asterisk characters, to preserve data integrity during storage.

2.8.2 Data Encryption

Olufohunsi (2019) describes data encryption as the process of converting plaintext information into unreadable ciphertext using various encryption algorithms. This process is accomplished by altering the string of characters contained in the information, to produce a new string of coded information. Olufohunsi (2019) notes that data encryption is necessary to secure privacy and to ensure data is only accessible by authorized users .Encryption further allows the receivers to verify that the message sent has not been altered during transmission and that the message is not from unauthorized sender (Simmons, 2018).The components of data encryption are as follows;

a) Plaintext. Olufohunsi (2019) refers to it as the original data that the sender wants to send to a specific recipient. This data is normally input to the encryption algorithm.
b) Encryption technique. Udaidullah (2016) notes that this is a set of processes that transform plaintext into ciphertext using a secret key.

c) Secret key is the value that is combined with the plaintext to transform it to cipher text, this value is independent of plaintext (Udaidullah, 2016).
d) Ciphertext is the output of the original plaintext which was put to the encryption algorithm (Olufohunsi, 2019).
e) Decryption algorithm is a set of processes that are executed to ciphertext to produce original plaintext using a secret key (Olufohunsi, 2019).

Figure 2.2 categorizes data encryption into Symmetric or Assymetric and presents encryption technique under each category.

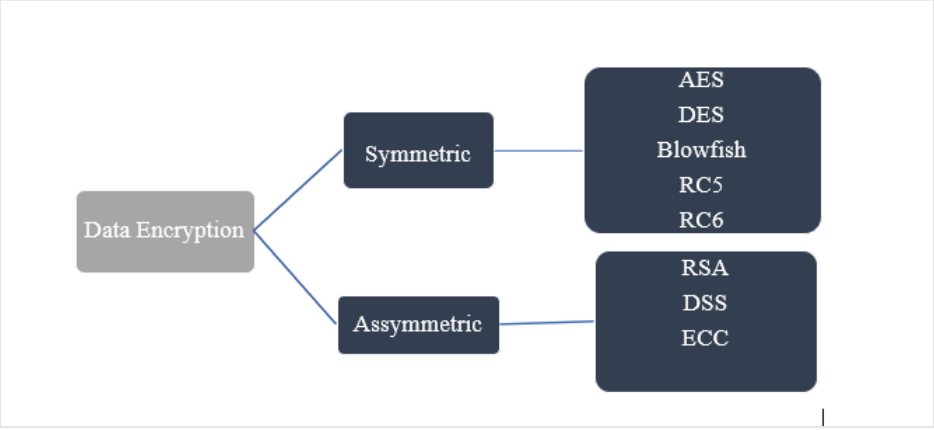


Figure 2.2:Data Encryption Techniques

Table 2.4 presents different encryption methods and what each method entails.

Table 2.4:Encryption Methods

Encryption Method	Description
Data Encryption Standard	According to NIST (2019) DES is a symmetric key algorithm that uses the same key for encrypting and decrypting data. Baguma (2018) notes that during encryption,DES converts 64-bit plaintext to a 64-bit ciphertext,while during encryption it converts 64-bit ciphertext into 64-bit block of plaintext and a 56-bit cipher key is used for both encryption and decryption.
Advanced Encryption Standard	Yehya (2020) explains that AES is an iterative symmetric-key block cipher that uses 128,192 and 256-bit keys, which is the main reason why it is deemed to be exponentially stronger and secure than DES. In addition, AES uses a substitution permutation with several rounds to generate ciphertext, and the key length used determines the number of rounds for every permutation.

RSA Encryption	Lal (2019) notes that RSA algorithm is a block cipher in which the plaintext and the ciphertext are integers between 0 and n-1. Encryption and decryption are guided by the formula (C = Pe mod n). Both the sender and the receiver must know the value of n and the sender is expected to know the value of e. This technique is computationally intensive which makes the encryption process slower and expensive. However, it is a secure algorithm because the longer the length of the key pairs that more time attackers take to crack the message.
Elliptic Curve Cryptography	Feldshuh (2020) notes that ECC is an asymmetric encryption algorithm that is based on elliptic curves and has both private key and public key. The public key is used to verify any process that is signed with private key. Further, ECC provides the best security against current cracking methods and has great performance, however many applications do not support ECC encryption technique.

2.8.3 Data Anonymization

According to EFD (2020) data anonymization refers to protection of personal or sensitive data by encrypting identifiers that identify an individual to the stored data.In most cases, data anonymization is irreversible and the recipient of the anonymized data cannot recreate the original data. Berdardino(2020) observes that the principles of data protection do not apply to anonymous information ,mainly because the information does not relate to an identifiable natural person as the data subject is no longer identifiable.

Table 2.5 describes some of the data anonymization techiques used depending on the nature of data and the specific purpose for the anonymization.

Table 2.5:Data Anonymization Techniques

Technique	Description
Attribute Suppresion	In this case, irrelevant and unnecessary attributes are removed from the data set and are permanently deleted making it impossible to retrieve the original data (Keerthiga, 2019).

Character Replacement	This technique involves replacing values and attributes of data with a predefined symbol such as the asterisk (*). Character replacement partially hides an attribute to sufficiently anonymize data (Berdardino, 2020)
Shuffling	Shuffling is normally used when analyzing one attribute that is not necessarily related to other attributes (Berdardino, 2020).
Noise Addition	This technique involves slight modification of data attributes to make them less accurate (Berdardino, 2020).
Generalization	This technique generalizes data attributes to change the respective scale of data. For instance, replacing the date attribute with the year attribute (Asamba, 2020).

2.8.4 Data Pseudonymization

Commission (2019) desribes data pseudonymization as a process of replacing any identifying characteristics of personal data with a value which does not allow the data subject to be directly identified.Unlike anonymization,pseudonymisation only provides limited protection for the identity of data subjects as it is possible to identify the data subject by analysing related data. Table 2.6 presents the basic pseudonymisation techniques .

Table 2.6:Psedonymization Techniques

Technique	Description
Counter	In this case,identifiers are substituted by numbers which provide for pseudonyms without directly relating to the initial identifiers (Enisa, 2021).
Random Number Generator	This technique provides strong data protection because it is impossible to extract information regarding the initial identifiers (Berdardino, 2020).

2.8.5 Data Minimization

According to Biega (2021), the General Data protection rule requires that the personal data be adequate, relevant, and limited to what is necessary in relation to the purposes for which it was collected, and that personal data should only be processed for specified and explicit reasons and not further be processed in a manner that is incompatible with these purposes. Belal (2017) explains that data minimization technique ensures that the collected personal data is not retained or further used for purposes that were not clearly stated in advance.

Biega (2021) divides data minimization into three distinct requirements.First, data must be adequate,this means that data must be well established without omission of certain data,to achieve

accurate results.Secondly,data ought to be relevant ,meaning that organizations should process only data that is directly related to the purpose for collection,to safeguard it against accumulating data for the sake (Biega, 2021).Lastly,the data must be limited ,meaning that data controllers should identify the minimum amount of data required to fulfil the stated purpose.

2.9 Existing Tools for Personal Data Protection

2.9.1 Dataflow Map Tool

According to Governance (2022) Dataflow Map tool is a cloud based tool that enables organizations and individuals to gain full visibility over the flow of their personal data .The tool helps in streamlining the data protection process and mitigating the risks of unauthorized data access.Further,the tool allows organizations to create consistent visual representations of the flow of data.

However,Bearer (2022) notes that mapping and arranging personal data collected from users is time consuming and error prone and that creating an accurate data flow map proves to be the hardest part of fulfilling data protection.

2.9.2 Incydr

Dearmer (2022) describes Incydr tool as a data risk and detection platform that focuses specifically on internal risks.The tool evaluates how organization's data is used,and identifies any anomaly behaviours and flags them to prevent any data security threats.The activities that Incydr tracks include:file creation,file deletion,file modification and file movement.Berdardino (2020) acknowledges that Incdyr tool is great at recognizing already existing patterns but requires constant improvements and updates ,for it to recognize the emerging data threats.

2.9.3 PADRES

PADRES is a front end web application that is built for data privacy,data regulation and data security (Crocker, 2021).It is developed in Angular framework and the backend connects to a database that is prior filled with the GDPR questions that need to be addressed. Crocker (2021) further notes that the major functionalities of this tool are the integration of security assessment tools and the creation of a final report with comments and recommendations.The tool is rigid in nature because it requires manual filling of the GDPR questions.Figure2.3 presents the architecture of the PADRES tool.

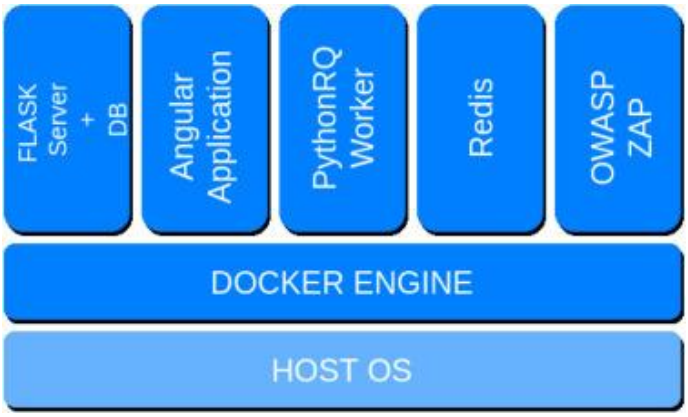


Figure 2.3:PADRES Architecture

Source: (Coker, 2020)

2.9.4 Proxymate

Faith (2018) notes that proxymate is an anonymizing tool that allow automatic generation of pseudonyms with corresponding passwords and emails that users send to websites. The tool uses a cryptographic function to create the pseudonyms and takes a user's username, password and the website domain as input and produces a pseudonymized username ,email address and password as output.Further, Faith (2018) notes that the main limitation of Proxymate tool is that different username and passwords are provided to every website that a user visits which is a tedious process.

2.9.5 Redstor Protector

Becta (2019) explains that Redstor Protector is a tool that allow users to secure data that is stored across the personal data storage and computing devices,by combining features such as backup,encryption,synchronization and data destruction.Becta (2019) notes that the tool has embedded data shredding and destruction feature that can be remotely exploited by users.

2.9.6 Crypt Sync

Ahmad (2017) observe that Crypt Sync is a cloud storage tool that uses encryption to secure personal data.Crypt Sync synchronizes two folders while encrypting the data stored in one of the folders.This basically means that one of the two folder contains all the unencrypted data while the other folder stores the encrypted data. Ahmad (2017) further notes that the synchronization process works in both ways ,such that a modification in one folder gets synchronized to the other folder.Furthermore,when data is added or modified in the unencrypted folder,it gets encrypted,however,if data is added or modified in the encrypted folder,it gets decrypted to the other folder.

Table 2.7 presents a summary of the existing personal data protection tool with their limitations.

Table 2.7:Summary of Gaps of Existing Tools

Tool	Limitation
Dataflow Map Tool	It is error prone and that creating an accurate data flow map proves to be the hardest part of fulfilling data protection
Incydr	Requires constant improvements and updates ,for it to recognize the emerging data threats.
Proxymate	The tool requires username and passwords are provided to every time a user visits a website which is a tedious process
PADRES	The tool is rigid in nature because it requires manual filling of the GDPR questions.
Redstor Protector	The tool has embedded data shredding and destruction feature that can be remotely exploited by users
Crypt Sync	When data is added or modified in the encrypted folder,it gets decrypted to the other folder.

2.10 Conceptual Framework

Figure 2.4 illustrates the conceptual framework for the proposed tool that aims to help organizations protect personal data by ensuring secure storage and secure processing of the data. The first step of the conceptual framework is collection of personal data by data processors and data controllers. Data processors collect different categories of personal data. The collected data is stored in a centralized data storage. In data storage, personal data is classified into sensitive and non-sensitive data.

Further, data subjects are required to authenticate into the solution using username and password. The framework incorporated access control mechanisms to ensure only legitimate users have access to the system and personal data. opt in and opt-out mechanisms are used to record consent of the data subjects for collection and processing of their personal data. The activities of the framework aim to achieve the main objective of the study which was to develop an integrated tool that ensures secure collection, storage, and processing of personal data based on data encryption and authorization techniques.

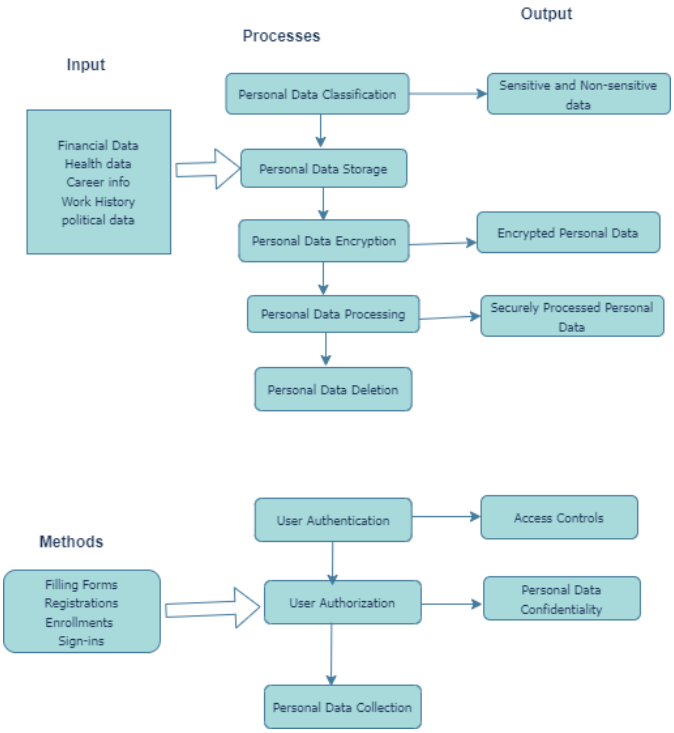


Figure 2.4:Conceptual Framework

2.11 Conclusion

This chapter presented an overview of the related literature to this study by exploring different categories of personal data processed by organizations. This section also outlined the laws and regulations that control personal data processing, the different techniques, and tools that organizations use to protect personal data. Additionally, existing data protection tools were reviewed and examined to point out their strengths and weaknesses in protecting personal data. Lastly, a conceptual framework was presented to illustrate how the developed solution handled inputs to produce the expected results.

Chapter 3: Research Methodology

3.1 Introduction

This chapter describes the methodology used in the study to answer research questions and meet the research objectives. Systematic research review was used to address research questions 1 and 2 while the Agile methodology was used to answer research questions 3 and 4.

3.2 Methodology for Objectives One and Two

The first objective of the study aimed at analyzing different categories of data, while the second objective was to analyze the techniques, approaches and tools used to securely process personal data. These two objectives were addressed through the systematic literature review methodology. Sturt (2022) explain that systematic literature review methodology is the process for identifying and critically appraising relevant research as well as for collecting and analysing data from the said research with the aim to identify all empirical evidence. Systematic literature review uses explicit methods to identify the basis of the study in a systematic and transparent way. Griffith (2022) notes that the key steps of systematic literature review are.

a) Formulating the research question.
b) Developing and validating review protocol.
c) Searching the literature.
d) Screening for inclusion.
e) Assessing quality.
f) Extracting data.
g) Analyzing and synthesizing data.
h) Interpreting the findings.

3.2.1 Formulating the Research Problem

Formulation of research question involved identifying the area of interest which was clearly stated as the main objective in Chapter 1. Further, the research problem clearly stated the issues that this study addresses and research objectives as well as research questions formulated to conduct meaningful research. The main research problem formulated in the study was to identify the categories of personal data collected by organizations and the various approaches used by organizations to protect personal data. Further, the research design selected to conduct this study involved qualitative research and quantitative research. Qualitative research was used to determine the relationship between the personal data collected.

3.2.2 Developing and Validating Review Protocol

Griffith (2022) explains that review protocol sets out the methods to be used in the literature review.This study involved describing the study selection,research questions ,data collection,data synthesis and the inclusion criteria for relevant literature.

In this stage,the researcher came up with the seach strategy ,developed research questions and screened different literature sources to find out how organizaions can securely collect personal data,store personal data and process personal data.

3.2.3 Searching the Literature

According to Griffith (2022) searching the literature involved creating search strategies for systematic reviews. Literature search was based on journals, reports, and electronic databases. Further, keyword search was conducted on Google and Google scholar to explore how organizations collect and process personal data, how personal data encryption is implemented as well as different approaches and tools used by organizations in securing personal data.

3.2.4 Screening for Inclusion

In this stage, a list of references was used to identify only the relevant research that met the inclusion criteria. Different articles and publications were screened independently to filter only the relevant information related to this study. The Data protection Act 2019 was used to guide on the data protection guidelines and principles of data protection described in the study.

In addition, the study recruited a sample from the population of interest to conduct a survey on personal data protection. The population consisted of individuals who have previously shared their personal data with different parties. It also included representatives of organizations who collect personal data for marketing and other processing activities. Basically, the population of interest included individuals who work in industries such as fintech, health, government, Insurance, education, and consultancy. The rationale for choosing the inclusion and exclusion criteria also depended on whether individuals have shown consent to participate in the survey by accepting to the terms and conditions. Sample of the population consisted of twenty individuals whose personal data was collected to conduct the research. Lastly, the individuals tested the performance and accuracy of the solution.

3.2.5 Assessing Quality

The study included a description of the area of interest as described in chapter one, further, the process used to select and gather data was clearly stated, as well as key concepts and deliverables.

This step also involved analyzing research data and selecting appropriate tools to analyze the findings and obtain the necessary information for the study. Chapter 4 describes the various tools applied to assess the quality of data and the application. The quality of data was maintained by collecting accurate and minimal data of the data subjects used for this study.

3.2.6 Extracting Data

This stage involved obtaining the necessary information about study characteristics and the findings from the study. Data extraction and study quality were undertaken simultaneously. Electronic form was used to combine data extraction and data entry to reduce bias and improve data quality and reliability. Adequate data was collected to avoid omission of certain data,to achieve accurate results in line with purpose for collection.Appendix A indicates the electronic form used to collect personal data.

3.2.7 Analyzing and Synthesizing data

In this stage personal data was organized according to the review methodology. To illustrate how the tool protects personal data, test cases were created in table format, clearly indicating the performance and reliability of the tool. Further the relationships between different modules were presented in table format. Personal data was stored in structured format into tables and fields. Individual entries within the database were recorded in columns and rows to illustrate the interaction of data, users, and the database itself.

In addition, descriptive statistics described the relationship between the population of interests and the personal data collected. The statistical methods used involved planning, designing, collecting data, analyzing, drawing meaningful interpretation, and reporting of the research findings.

3.2.8 Interpreting the Findings

This step outlines all the important findings and how the research questions were addressed. The section also compared findings of previous studies and outline the relevance of the findings to the field of study in general. Based on the research problem identified in this study, the activities to promote research utilization included providing technical assistance to data processors and data subjects and holding one-on-one meetings with data subjects to facilitate in personal data collection and testing of the developed solution. The research findings were utilized instrumentally and conceptually at different stages of secure data collection, secure data storage and secure personal data processing.

3.2.9 Dissemination of Results

This stage outlines the process of selecting the target audience and the setting in which the research findings are to be communicated in a way that will facilitate decision making processes. This study identified an audience of interest that included individuals who work in industries such as fintech, health, government, Insurance, education, and consultancy. The audience was engaged through one-on-one meeting and zoom meetings to achieve high quality research.

3.2.10 Utilization of Results

The study implemented activities to promote research utilization. These activities included conducting dissemination meetings, providing technical assistance to the audience, holding one-on-one meetings, and joining data protection social networks. Case studies were used to indicate how the research findings were utilized instrumentally at different stages of the research.

3.3 Agile Methodology for Software Development

Objective three of the study aimed at designing, developing, and testing the integrated tool that securely processes personal data, while the fourth objective was to validate the effectiveness of the developed tool in ensuring secure collection and processing of personal data. Agile software development methodology was used to address the two objectives.

Guru (2018) explains that Agile software development methodology is a set of practices that promote continuous iteration of development and testing throughout the software development lifecycle. This methodology was applied in the study because it is iterative with short cycles and allowed modularity on the development process. Agile methodology is composed of six steps as illustrated in Figure 3.1

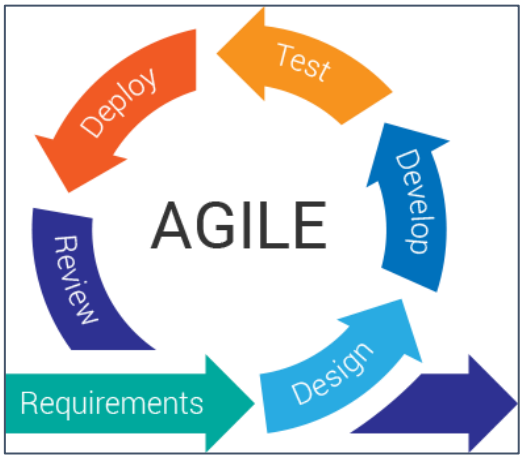


Figure 3.1:Agile Methodology

Source: (Technology, 2016)

3.3.1 Requirements

The purpose of this phase was to document all the functionalities and scope of the system. This involved generating the functional, system and technical requirements to achieve the research objectives. This was achieved through online forms, that were used to gather information on how organizations process personal data as well getting the data subjects opinions and feedback on personal data collection and processing. See Appendix A for data collection and Appendix B for feedback.

3.3.2 Design

In this phase the system requirements were translated into a conceptual design. The system architecture, modules, interfaces, and data were defined in this stage. In this research a flowchart diagram was used to graphically represent the flow of data. The conceptual design was modelled using sequence diagram and entity relationship diagrams. Further, the use case diagram was used to illustrate the relationship between the data subjects and data processors who are the major actors of the system. The following tools were used in this phase.

a) Postman-This is an API platform that was used to design and build the Application Programming Interfaces.
b) SmartArt-This is a Microsoft product that was used to draw conceptual frameworks and the data flow diagrams.
c) Draw.io-This is proprietary software for making diagrams and charts. This tool was used to draw sequence diagram, use case diagram, context diagram, DFD and ERD.

The overall design techniques used in this study include.

a) Flowchart was used to illustrate how the data flow across the tool and the processes and data interact with each other.
b) A use case was used to describe the actors within the application. The main actors for the developed personal data protection tool are the data processors/controllers and data subjects whose personal data is collected for processing activities.
c) Sequence Diagram was used to illustrate the interaction between the objects depicted in a single use case,and illustrated how different parts of the developed tool interact with each other to achieve the objective.
d) Entity Relationship Diagram was used to represent the database design structure and indicated the relationship between the personal data stored in the database.

e) Context Diagram was used to show the interactions between the developed action and other actors that interface the system.
f) Data flow diagram visually represents the flow of data within a system.

3.3.3 Development

This phase involved converting the conceptual design into a working tool, based on the system requirements. This stage entailed writing the source code using the Java programming language which provided an object-oriented approach. The following tools were used in this phase.

a) Postman-This API platform was used to build the Application Programming Interfaces.
b) IntelliJ IDEA-This an integrated development environment (IDE) that was used for coding the developed tool. IntelliJ IDEA was a suitable development environment because it offers a wide variety of Java frameworks and instant code analysis.

3.3.4 Testing

Testing was carried out to ensure that the application works accordingly based on the requirements, to measure the tools effectiveness in addressing the research questions. Functional testing, manual testing and automated testing were carried out in this phase to ascertain the functionality and reliability of the developed tool. Test cases were developed to form part of the manual testing while Postman API testing tool was used to carry out automated testing. These testing methods ensured that the tool requirements were met.

3.3.5 Deployment

This step involved packaging and updating the application for use. It entailed delivering the product to prospective users for testing and feedback. JBoss webserver, which is an open-source platform developed by Red Hat, was used to run, deploy, and manage the developed personal data protection tool.

3.3.6 Review

This phase incorporated activities that track and validate the application's performance. Review was done to validate the model performance. To review the tool's requirement, online google forms were used to collect feedback from potential users, see Appendix B

3.4 Research Quality and Ethics

The study conducted the research in accordance with the university guidelines and standards. The problem statement and research objectives were evaluated to ensure that the output meets the set requirements. Further the data protection tool was aligned with the principles of data protection

and the guidelines provided by the Data Protection Act of Kenya. This generally included the values of preventing harm, ensuring privacy, maintaining confidentiality, and ensuring that the benefits of data collection and processing outweigh the risks.

3.5 Conclusion

This chapter summarized the research methodologies used to meet the research objectives. Systematic literature review was used to address research objectives one and two stated in Chapter one while Agile software development methodology was used to design and develop the data protection tool. Further, the chapter gives a detailed description of the specific steps that were applied in the software development life cycle. To close the chapter, the ethical considerations were presented.

Chapter 4: System Design and Architecture

4.1 Overview

This chapter describes the system analysis and architectural design of the data protection tool. A detailed requirement analysis was conducted in this chapter to define the functional and nonfunctional requirements of the tool. To close the chapter, the design tools used to design the developed tool were well presented.

4.2 Requirement Analysis

4.2.1 Functional Requirements

Canedo (2020) defines functional requirements as the functionality of a system or system components. Canedo (2020) further notes that functional requirements describe the behavior of function of the system. In this study, functional requirements were defined by presenting a high-level description of the second research objective which aimed at reviewing the existing data protection tools and analyzing their gaps. The requirements were defined as follows:

a. The application should allow secure collection of personal data.
b. The application should enable secure storage of personal data.
c. The application should enable secure processing of data.
d. The application should enable opt-in and opt-out mechanisms.
e. The application should enable personal data deletion.

4.2.2 Non-functional Requirements

Eriksson (2018) notes that non-functional requirements describe how a system should behave and the aspects that limit the system's functionality. These requirements cover all the system requirements that are not covered by the functional requirements. The requirements were defined as follows:

a. Security-The application provided access to authorized and legitimate users only.
b. Availability-The application operated with zero or minimal downtimes to ensure users can use it any time they need it.
c. Performance-The tool performed data access and enabled secure data processing without interruptions.
d. Scalability-The application was designed to accommodate increased data volumes and users.

e. Usability-The application followed the Data Protection Act guidelines on personal data protection.

4.3 System Architecture

Faisandier (2020) explain that system architecture defines a comprehensive solution based on the principles of the system requirements, and has features, properties and characteristics that help to achieve the system requirements Figure 4.1 represents logical architecture of the developed tool for personal data protection. The architecture consists of six major components that include personal data collection, data storage, opt-in and opt-out, data access and personal data processing activities.

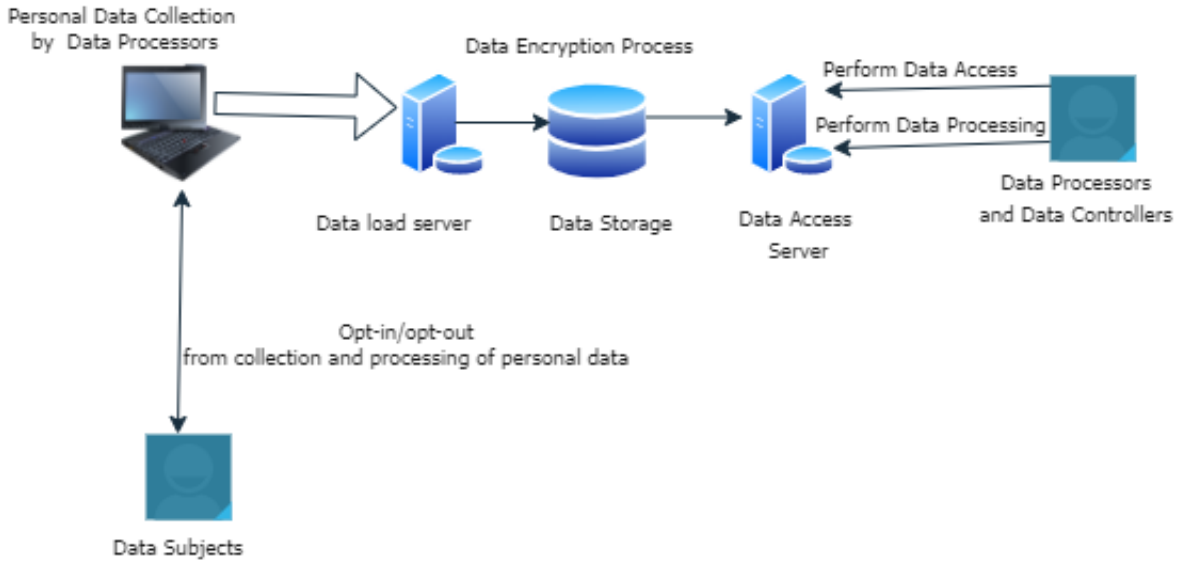


Figure 4.1: System Architecture

a) Personal Data Collection

In this stage, data processors directly collect personal data from the data subjects. Google forms and online survey were used to collect personal data from the data subjects and get feedback refer to appendix A. The collected personal data was then input to the API. Only the necessary personal data was collected for the study. After collection, the data was organized and stored for further processing.

b) Personal Data Storage

This component illustrated the database that was used to store the personal data, and as a repository for retrieving personal data. In the database, personal data was arranged in tables and columns to logically represent it.

Further, data was classified into sensitive and non-sensitive personal data. To achieve security, all the sensitive data was encrypted for confidentiality and integrity of the data.

c) Personal Data Access

In this stage data processors were given granted access to view the personal data stored in the database,to modify and alter it as well as delete the data. In the same context,data subjects were allowed to view the personal data they had submitted and check any change or modification on their personal data. Further, data processors and data objects were granted access to record organize, edit, alter, search, and delete data.

d) Opt-in/Opt-out

The opt-in component illustrated explicit consent from the data subjects to authorize data processors and controllers to collect and process their personal data. This indicated an affirmative action taken by the data subject indicating their consent to allow collection and processing of their personal data.

In the same context, the purpose of the opt-out component was to allow data subjects to withdraw their consent to the collection, storage, and processing of their personal data.

e) Personal Data Deletion

This process involved a set of activities that included removing personal data from the storage in accordance with the Data Protection Act.These activities also included permanently deleting the personal data.This stage was approached in two dimensions.First,the data subjects would request the data processor to permanently erase their data.Second,the data processors would delete the data once the purpose for which data was collected was achieved.

4.4 System Design Tools

Edition (2021) explains that system design is the process of defining elements of a system such as models, architecture, data, components, and their interfaces based on the specified requirements.

4.4.1 Flow Chart

Lucidchart(2021) explains that a flow chart depicts a process, system, or computer algorithm. It is commonly used to clearly communicate complex processes. Figure 4.2 illustrates how the data flows across the tool.First personal data collected was input in the application ,once in the data storage,the data processor verifies whether data subject has opted in for data processing,if this is the case the data processor is granted access to process the data.

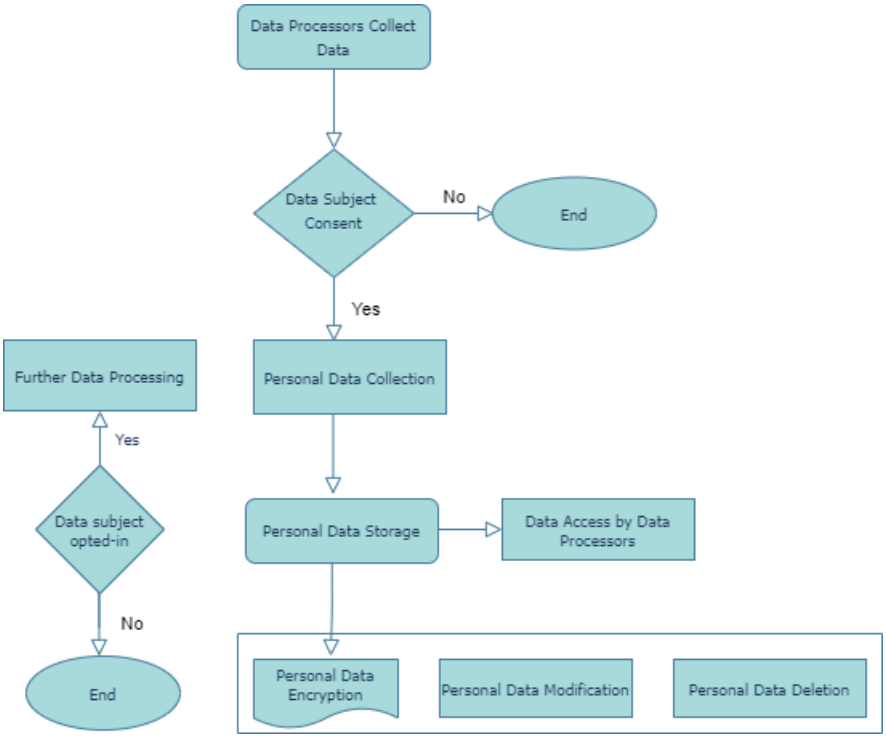


Figure 4.2:Flow Chart Diagram

4.4.2 Use Case

Contributor (2021) states that usecase diagram summarizes the the interactions of the actors within the system,and specifies the specific events in the system.Actors within the usecase diagram represent users of the system.Figure 4.3 represent the use case diagram for the developed intergrated tool for personal data protection.The data processors and data subjects were the main actors of the developmed application.Data processors represented the organizations that collect personal data for processing activities.In the same context, data subjects represented the data owners.

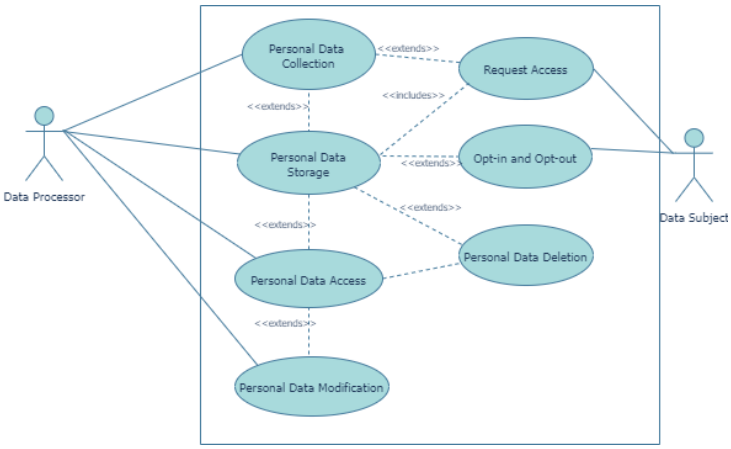


Figure 4.3:Use Case Diagram

i. Get Personal Data Use Case Description

Table 4.1 illustrates Get Personal Data Use Case

Table 4.1:Personal Data Use Case

Use Case Title	Get Personal Data
Description	This use case describes the process followed to acquires data subject's personal data such as name, ID number, Addresses among others.
Actors	Data Processor
Basic Flow	a) Get data subject's personal data. b) Create profiles for data subjects.
Pre-condition	None
Post condition	The data should be automatically stored in data store.

ii. Personal Data Storage Use Case Description

Table 4.2 illustrates Personal Data Storage Use Case.

Table 4.2:Personal Data Storage Use Case

Use Case Title	Store Personal Data
Description	This use case describes how the collected personal data is stored in the data store.
Actors	Data Processor
Basic Flow	a) Get data subject's personal data. b) Initiate Data storage
Pre-condition	Personal data has been collected.
Post condition	Data access request

iii. Request Access

Table 4.3 illustrates request access use case.

Table4.3: Request Access Use Case

Use Case Title	Request Access
Description	This use case describes the steps taken to request for data access.
Actors	Data Processor
Basic Flow	a) Authenticate into the system.

	b) Request data access
Pre-condition	System Authentication
Post condition	Data processing which involves data modification, data processing, data sharing and data deletion.

iv. OPT-IN/OPT-OUT Use Case

Table 4.4 illustrates OPT-IN/OPT-OUT Use Case.

Table 4.4:Opt-in and Opt-out Use Case

Use Case Title	OPT-IN/OPT-OUT
Description	This use case describes the opt-in/opt-out mechanism for data subjects.
Actors	Data Subject
Basic Flow	a) Personal data is already collected. b) OPT-IN /OPT-OUT from data processing.
Pre-condition	Personal data is collected.
Post condition	None

v. Grant Access Use Case

Table 4.5 illustrates Data Access Use Case.

Table 4.5:Data Access Use Case

Use Case Title	Data Access
Description	This use case describes the process of granting access for data processing and sharing.
Actors	Data Subject
Basic Flow	a) Personal data is already collected. b) Personal Data is stored in the data stores c) Request Data Access d) Grant Data Access
Pre-condition	Data Access Request
Post condition	Data Modification Data Deletion

vi. Data Modification

Table 4.6 illustrates the Data Modification Use Case.

Table 4.6:Data Modification Use Case

Use Case Title	Data Modification
Description	This use case describes the modifications activities on the personal data collected. This includes editing the personal data.
Actors	Data Processor
Basic Flow	a) Get data subject's personal data. b) Store Personal Data c) Create profiles for data subjects. d) Request Data Access e) Personal Data Modification
Pre-condition	Data Access has been granted
Post condition	None

vii. Data Deletion Use Case

Table 4.7 illustrates the Data Deletion Use case.

Table 4.7:Data Deletion Use Case

Use Case Title	Data Deletion
Description	This use case describes the process of deleting personal data after it has fulfilled its intended purpose.
Actors	Data Processor
Basic Flow	a) Get data subject's personal data. b) Store Personal Data c) Create profiles for data subjects. d) Request Data Access e) Access Granted f) Delete Personal Data
Pre-conditions	Get data subject's personal data. Store Personal Data Create profiles for data subjects. Request Data Access Access Granted
Post condition	None

4.4.3 Sequence Diagram

Creately (2020) explains that sequence diagram illustrates the interaction between the objects depicted in a single use case. A sequence diagram also illustrates how various parts of the system interact with each other to execute a task. Figure 4.4 illustrates how different modules interacted in the developed tool.

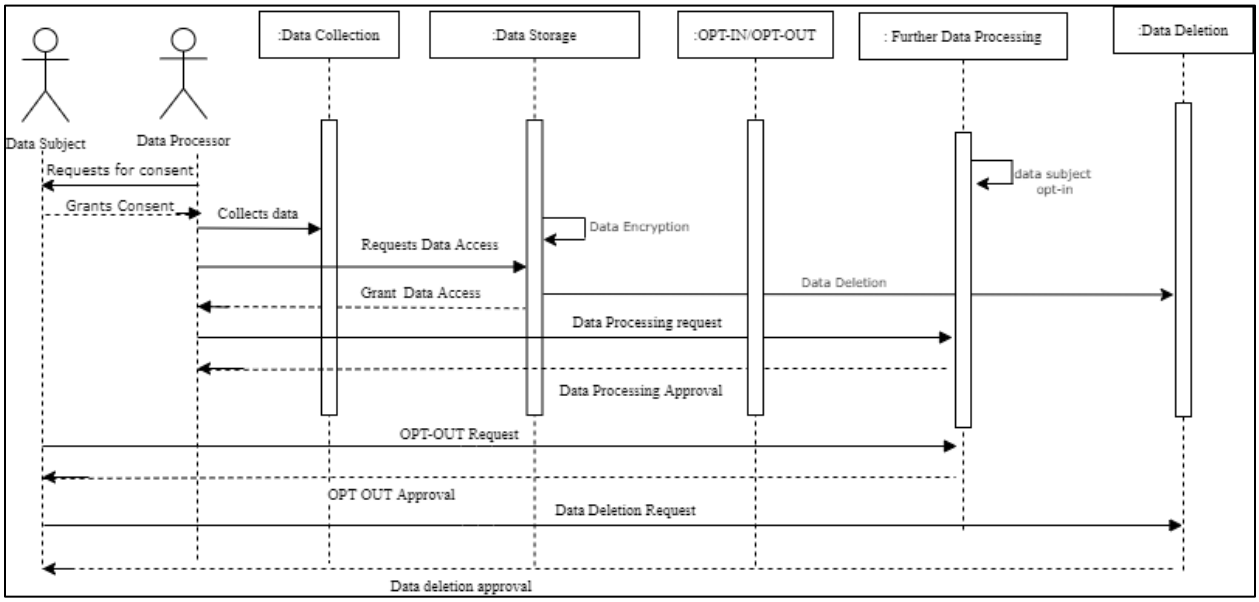


Figure 4.4: Sequence Diagram

4.4.4 Entity Relationship Diagram

An entity relationship diagram (ERD) represents the database design structure and indicates the relationship between entities stored in the database. Figure 4.5 represents the model of the integrated personal data protection tool. It showed all the visual instruments of the database tables and the relations between personal data and data subjects. Further, the database used structure data to define the relationship between the data groups and the tools functionality. The main entities of the personal data protection tool were personal data, data subjects and data processors.

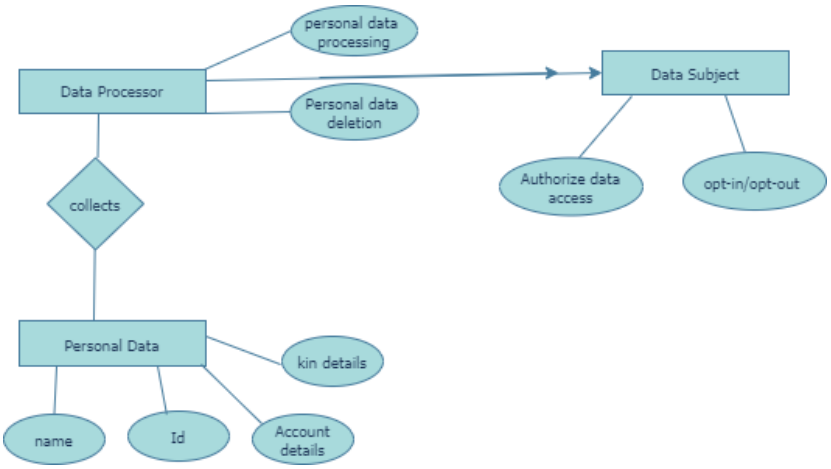


Figure 4.5:Entity Relationship Diagram

4.4.5 Data Flow Diagrams

4.4.5.1 Context Diagram

Context diagram shows the interactions between a system and other system actors that interface the system. Figure 4.6 highlights the major components of the developed application for protecting personal data. The data processors and data subjects were the main actors of the developed tool.Data processors get consent from data subjects before collecting and processing data.Once the data subjects consent to the collection of their data,the data is stored in and modified to fulfil the purposed for which the data was collected.Data subjects are also provided with the option to opt out from the application when they do not want their personal data processed.

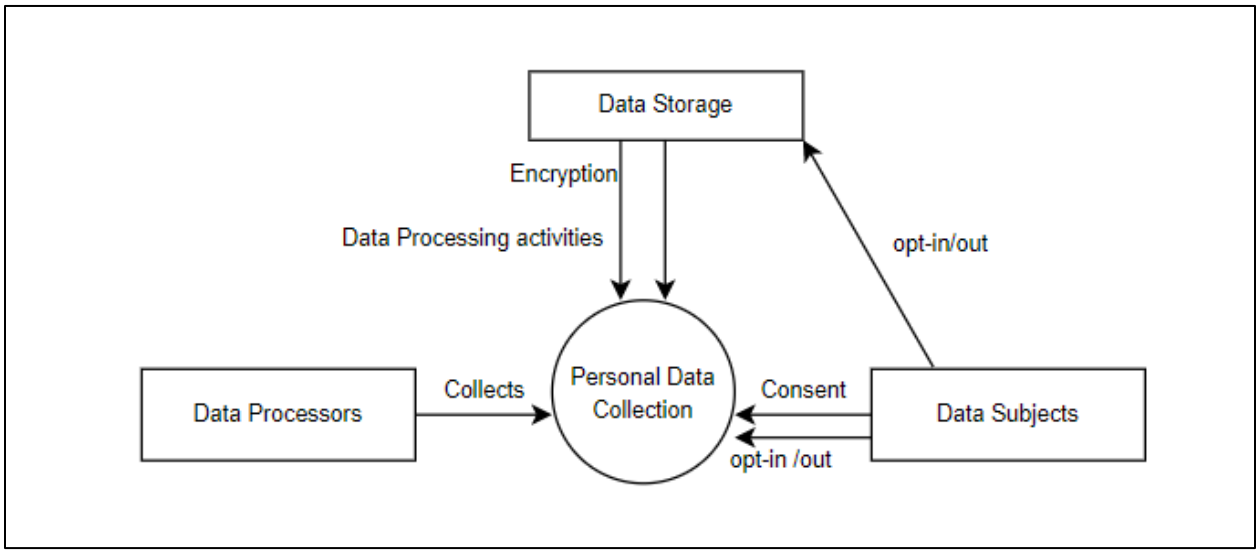


Figure 4.6:Context Diagram

4.4.5.2 Level One Dataflow Diagram

According to Chi (2021) Data flow diagram is a visual representation of how the information flows within a system. It gives better understanding of the system operation to discover potential problems, improve efficiency and develop better processes. Figure 4.7 illustrates the data flow diagram for the developed personal data protection tool.

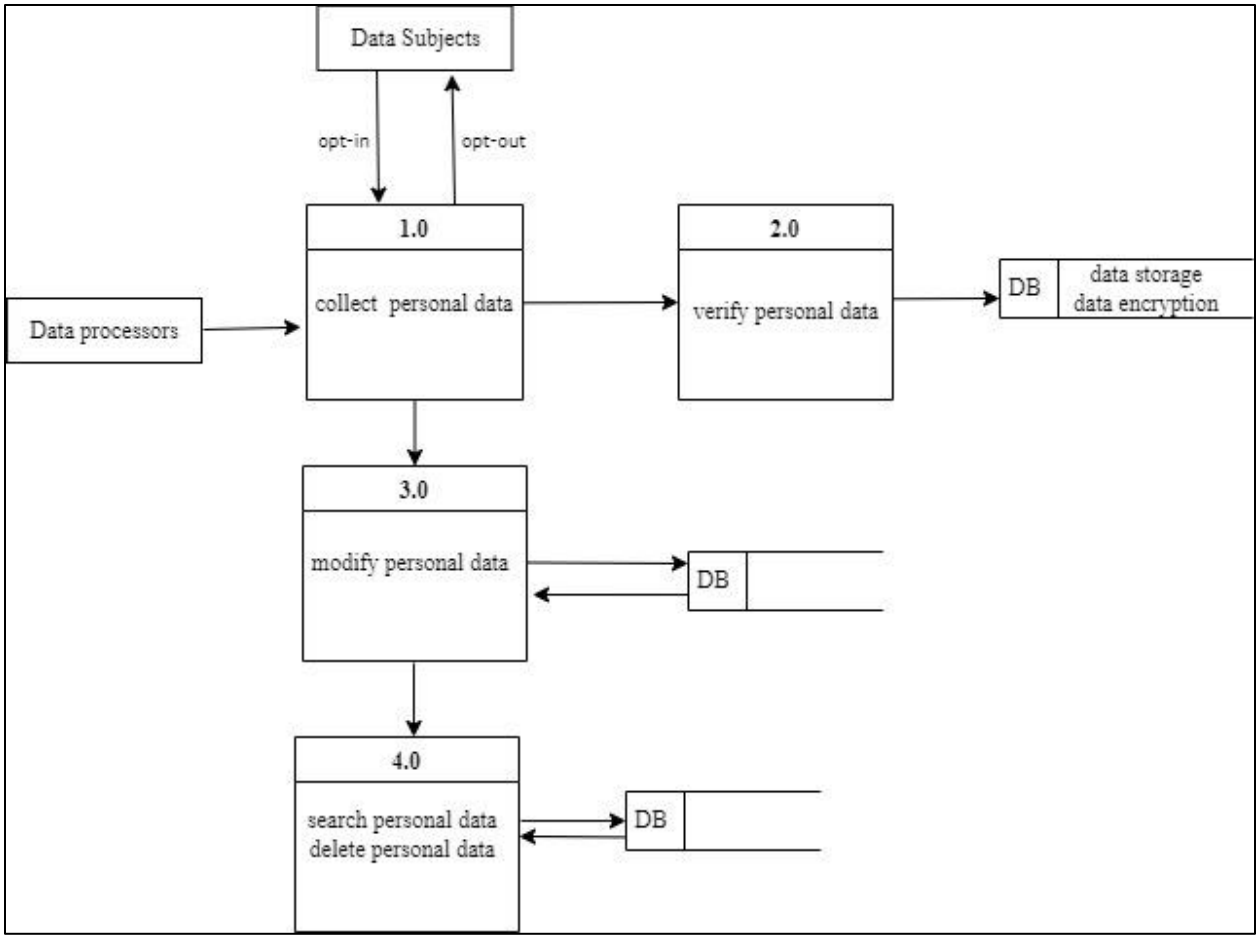


Figure 4.7:Level One Data Flow Diagram

4.5 Security Design

Wigmore (2020) notes that security design is an application development approach that seeks to protect applications against exploitation and resilient to attacks through continuous testing, authentication safeguards and adherence to best application development practices. The developed tool for personal data protection focused on five key security controls as depicted by figure 4.8 The security design adopted during the development of the tool provided consent to personal data collection, authentication, authorization, data validation and encryption controls.

Encryption was the main method used to secure stored personal data. Data encryption reduced the risk of personal data misuse as access was limited only to authorized people with the right key. In addition, Privacy by design approach was applied to safeguard the privacy of data subjects,' by collecting adequate and minimal personal data in relation to the purposes of this study. During data collection, data subjects were immediately informed about the purpose of data collection, the processing purposes and information about the duration of storage of their personal data. The

developed tool also provided the data subjects with the option to withdraw consent from processing activities.

The developed application makes use of SSL certificates to securely transmit data to the backend server through the APIs. The developed tool transmits data through HTTPS for secure communication.

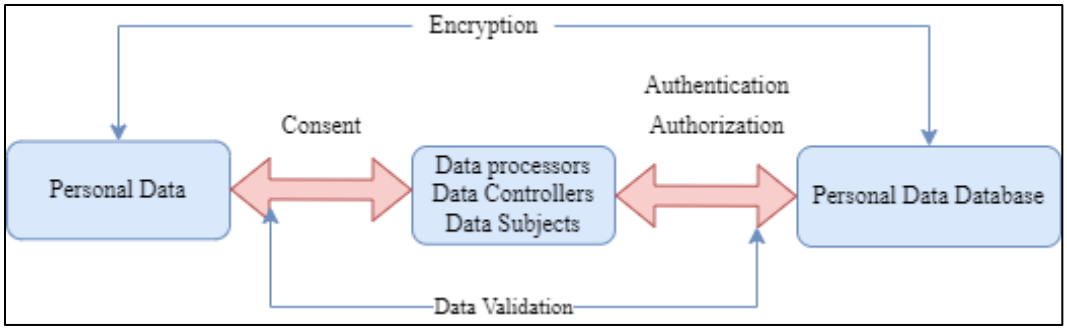


Figure 4.8:Security Design

4.6 Network Design

Technopedia (2021) explains that network design is planning the implementation of the network infrastructure. It involves evaluating, understanding, and scoping the network implemented. This study used the client-server architecture as part of the network design. The architecture consisted of computers connected to a central server over internet connection.

Figure 4.9 illustrates the network diagram that served as a blueprint for the network physically. Typically, the network design included the following.

a) The endpoint URL
b) Backend Server
c) PostgreSQL database
d) IP Addresses

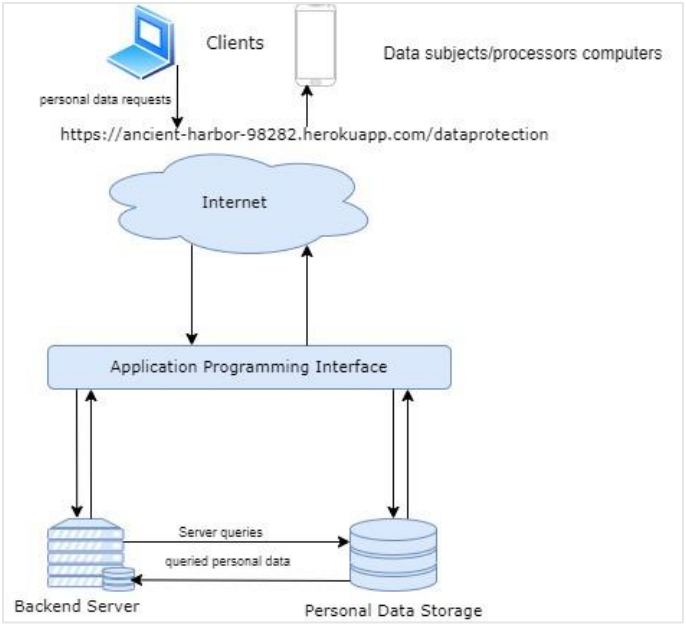


Figure 4.9:Network Design

The flow of data within the network is unidirectional and forms a cycle. As illustrated in figure 4.9, the client (data processors using their computers) initiated requests to the server. The server processes the requests and sends back data to the client. The typical flow of data within the network was as follows:

1. Client requests data from server.
2. Server processes the client's requests.
3. Server queries appropriate database for data.
4. Database returns the queried data back to the server.
5. The server processes the data and sends it back to the client.
6. This process repeats.

4.7 Wireframes

As illustrated in figure 4.10 the developed application data subjects and data processors were required to select their profiles and input their username and passwords for secure authentication. Once the users confirm their identity, they are allowed to have access to the application.

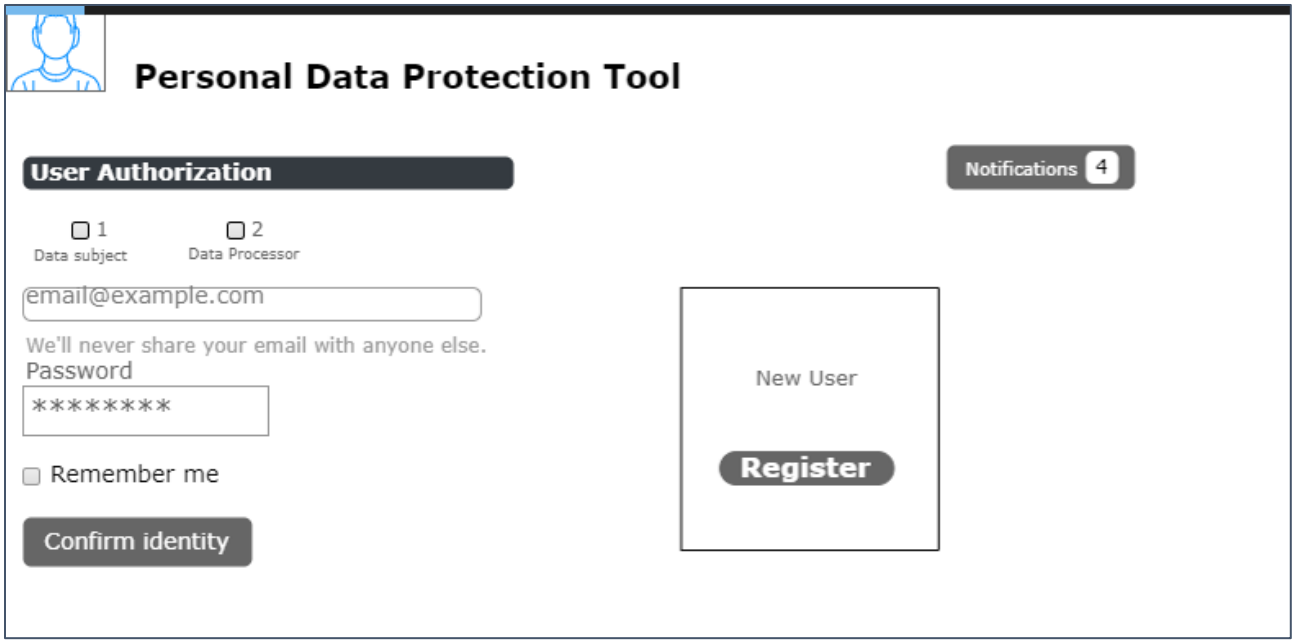


Figure 4.10:Wireframes Showing Authentication Page

Figure 4.11 illustrates the home page of the developed tool. Data processors and data subjects can navigate the dashboard to carry out various tasks like view and modify the collected personal data.

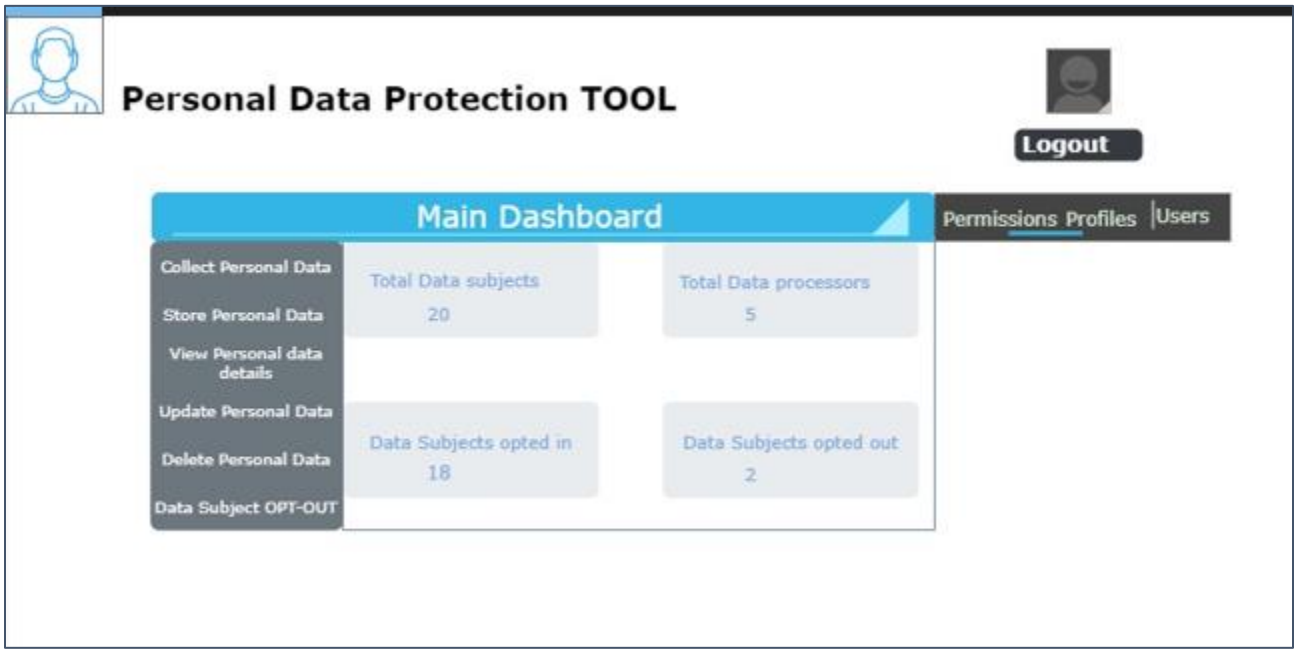


Figure 4.11:Dashboard

Figure 4.12 illustrates the personal data collection form and sample data that data processors collect from the data subjects. Data subjects are required to agree to the terms and conditions to show consent to personal data collection. In the same context, data subjects were provided with the option to opt-out from personal data collection and processing. In addition, the developed tool allows data processors to edit the personal data collected as well as add more information if needed. Once the collected data has accomplished its purpose, it is permanently deleted and discarded.

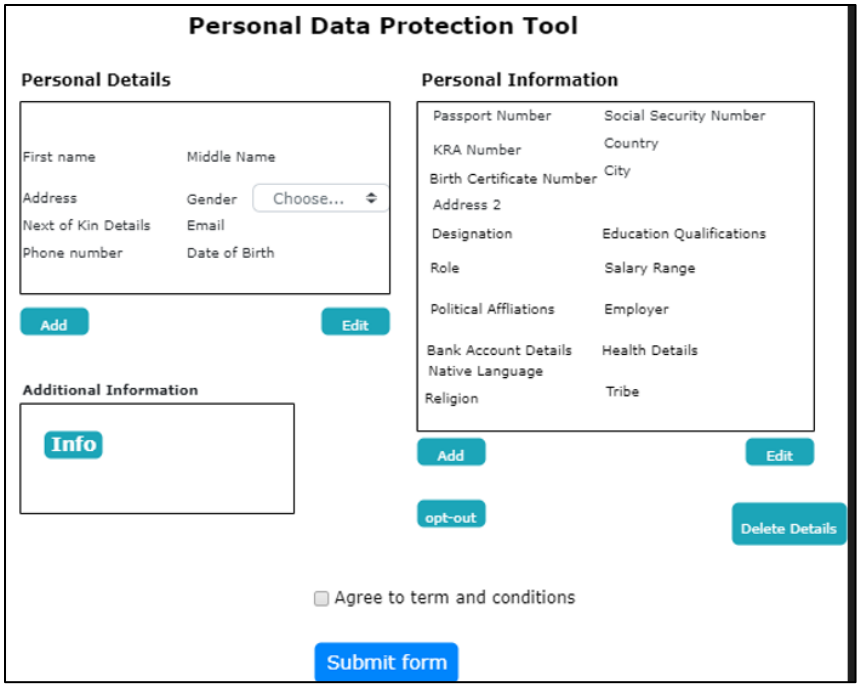


Figure 4.12:Personal Data

4.8 Conclusion

This chapter presented the system architecture of the integrated personal data protection tool which consists of five major components that include get personal data, data access, personal data storage, personal data processing and personal data deletion. Further, the system design tools used to model the tool were discussed in the chapter to indicate data flow and module interaction within the application, as well as the interactions of the actors within the system. To close the chapter, security design and network design were presented. Security design was used to illustrate the approach taken to protect the developed tool against possible exploitation and attacks. In the same context, network design illustrated a blueprint of the physical network.

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter covered the implementation of the integrated personal data protection tool and highlighted the hardware, software and network specifications required to build the tool. To close the chapter, the testing tools and procedures used to validate the performance of the tool were described in detail.

5.2 Development Environment

5.2.1 Hardware Environment

Table 5.1 illustrates the specific hardware requirements which were used to develop the application.

Table 5.1 Hardware Environment

Device Name	Wangui-PC
Processor	Intel Core i7-10510U
Device ID	F9F6CDDF-2696-49EC-816A-555E88FE10C5
Product ID	00329-10286-34367-AA700
RAM	16GB

5.2.2 Software Environment

The features and functionalities of the Software used to build the application are as follows.

a) Java programming language Version 8 is a high-level, object-oriented programming language, which allows applications to run on a variety of platforms. This language was used to build the APIs.
b) IntelliJ IDEA Community edition is the integrated development environment (IDE) that was used for coding the Java program. IntelliJ IDEA was a suitable development environment because it offers a wide variety of Java frameworks and instant code analysis.
c) Postman-This is an API platform tool used for designing, building, and using the APIs.
d) PostgreSQL version 13 was the database management system used for personal data storage.
e) PostgreSQL Server was used to add, access, and process the personal data stored in the PostgreSQL database.
f) Windows 10 Enterprise was the operating system used to run the application.

5.2.3Network Environment

a) JBoss webserver community edition which is an open-source platform developed by Red Hat, to implementing Java applications, was used to run, deploy, and manage the developed application.

b) Heroku is an open-source platform that was used to run and operate the developed application.

5.3 Application Programming Interfaces

The personal data protection tool developed had several modules that were in the form of application programming interfaces (API) that interacted in accordance with the functional and nonfunctional requirements of the tool.

5.3.1 Data Collection

Data collection was manually collected from the data subjects then it was input to an API. The collected data was filtered to ensure that minimum data is collected to achieve the quality and accuracy of the output. Different personal data that identified the data subjects was used in this study. Figure 5.1 illustrates the personal data variables collected.

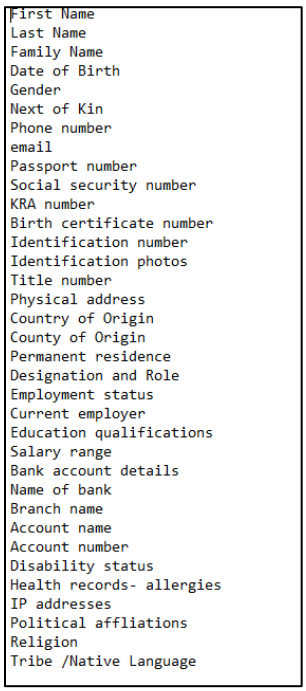


Figure 5.1:Personal Data Collected

In the same context figure 5.2 illustrates the personal data collection form and sample data that data processors collected from the data subjects.

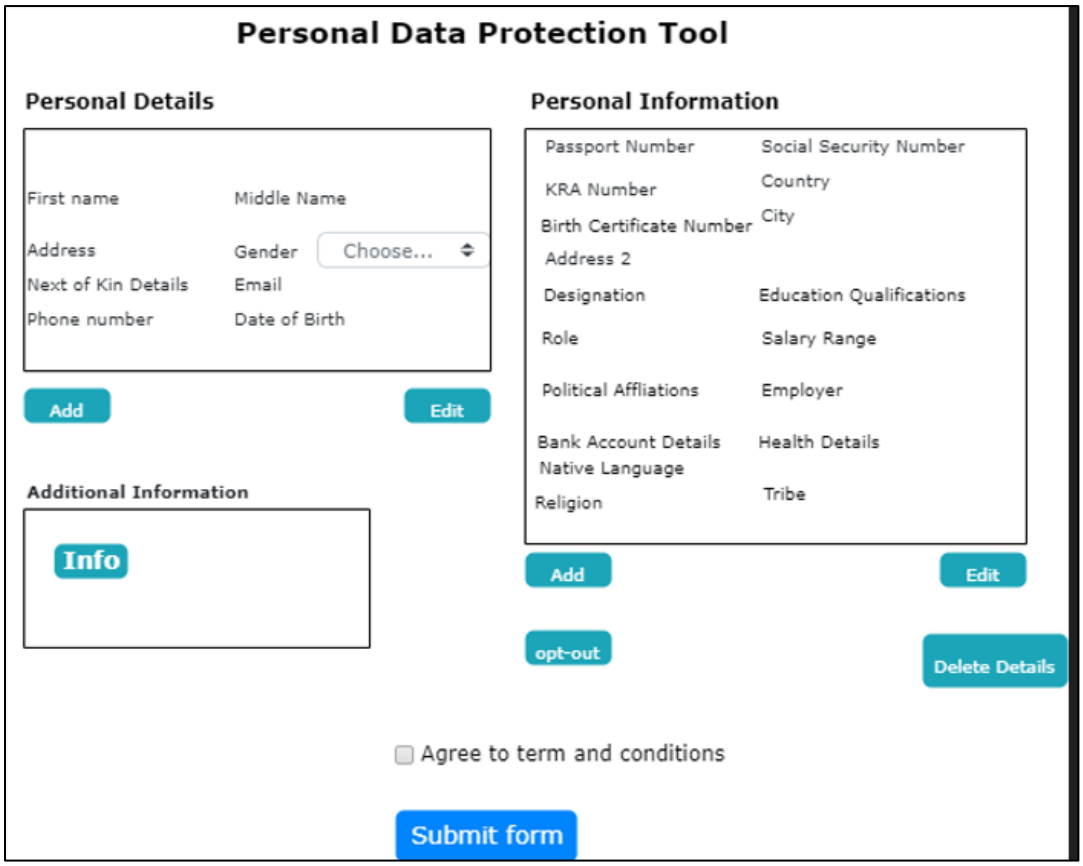


Figure 5.2:Personal Data Collection Form

Further, data classification was done to group sensitive data and non-sensitive data. Sensitive data required an extra level of security, so it was encrypted during data storage. Table 5.2 shows how some personal data was classified.

Table 5.2 Personal Data Classification

Sensitive Personal Data	Non-Sensitive Personal Data
• Social Security Number • KRA Number • Title Number • Account Number • Bank Details	• First Name • Last Name • Family Name • Mail • Designation and Role

Figure 5.3 illustrates how data was recorded in the API with actual personal data for further processing activities.

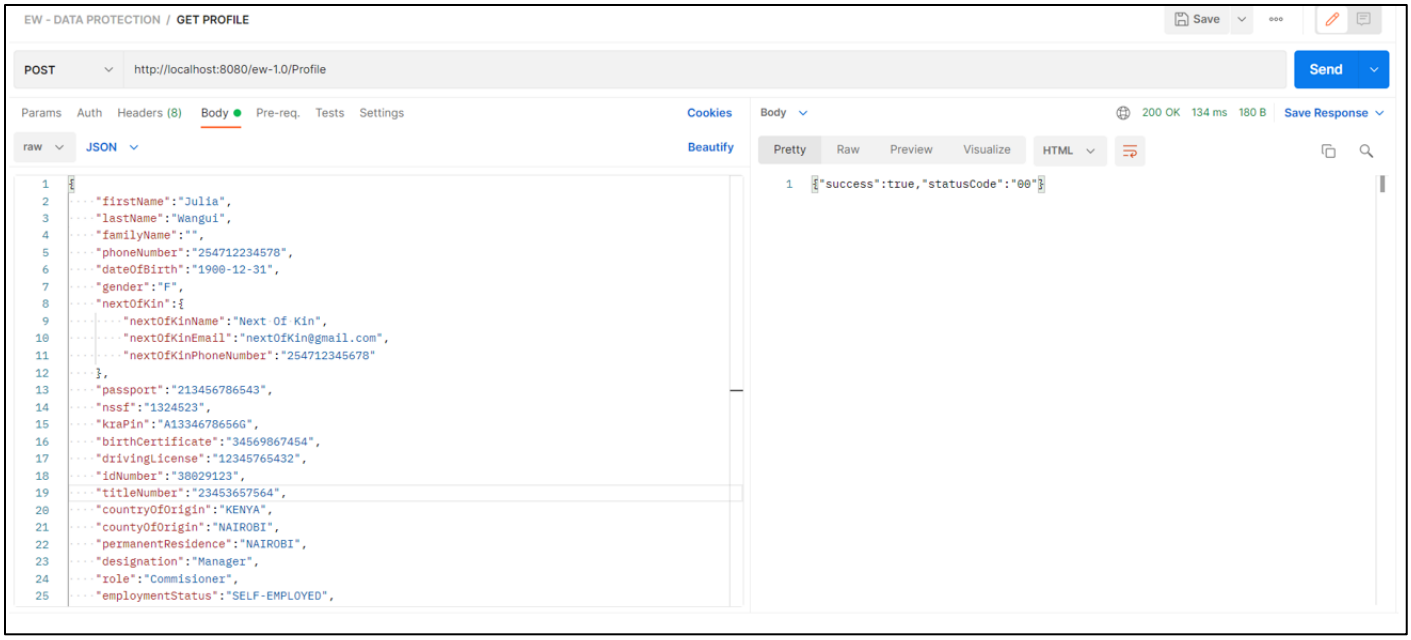


Figure 5.3:Sample Personal Data

5.3.2 Store Personal Data

The personal data collected was stored in a PostgreSQL database. Data was arranged in a table format which consisted of columns and rows to include the relationship the data. All the sensitive data was encrypted to ensure its confidentiality and integrity of data during storage and retrieval. RSA Encryption technique was used to encrypt and decrypt data.

Sharma (2018) notes that RSA is a public key encryption technique that provides security by encrypting and decrypting the data, so that only authorized users can access the data. In this study, the data processors used the public key whereas the private key was used by data subjects who originally owned the data. The encryption technique involved three steps:

a) Key Generation

This step involved generating two keys, public key E, and Private key D. In addition, two prime numbers (PR1 and PR2) were included in the process. Four prime numbers are multiplied and computed as N1. Further, two prime numbers are computed as N2 to increase the complexity of the encryption algorithm.

The basic steps of key generation process used for RSA encryption were as follows.

1. Selected two sets of large prime numbers. The first set included PR1 and PR2, while the second set included prime numbers P and Q.
2. The values of N1 and N2 were calculated as N1=P*Q*PR1, N2 =P*Q
3. Selected the public key E
4. The private key D was computed from D*E=1

5. The public key comprised of a pair of E and N1 and private key pair as D and N2

b) Encryption

This stage involved generating a ciphertext from the given plaintext. The process used the public key to generate the cipher text.

The formula for generating the cipher text from the given plain text was,

C=M mod(N1)

c) Decryption

In this step, the original plain text was retrieved by using the values of cipher text. The formula for generating the plain text was found by M=C mod (N2)

RSA encryption algorithm was used in this study because it is safe and secure for transmitting confidential data and it is very difficult for attackers to crack the message since the technique involves complex computations. Figure 5.4 outlines the working of RSA technique applied in the study.

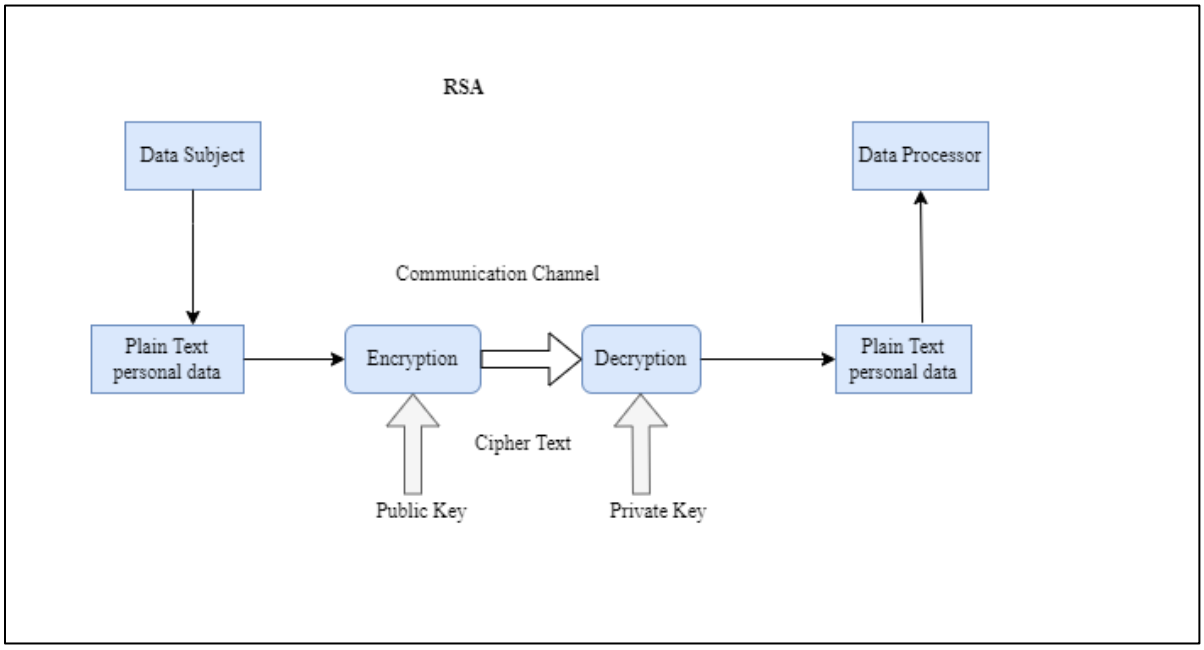


Figure 5.4:Encryption Technique

The personal data submitted by data subjects in plain text format is encrypted using the public key of the data processor. Further, the data processor decrypts the ciphertext using the private that is only known to the party.

KRA PIN and account numbers were classified among the sensitive data. Figure 5.5 illustrates the encrypted format of the account numbers stored in the database.

phonenumber	lastname	salaryrange	accountnumber
254702376550	Wangui	50000-100000	5Sr4F6Cto+/B6XRCO21Z5i43TkUnx7gxC5j+kh7smHqAVTi1BBShol+6wnd1yMWaniryUJO++qyRl2Q5zaMRyXJUccdlG1kzrcaoKW8MNW4+UYkWEsvoGQWJaPxrX9SN0NqlVjZysj6+mu69Hvgu+Wd0C5i84foGVwQ0MFlnYyVdK/QdzDlGEeZswe8kjn6RUmkOKgxxq6pQUehSqlQimAiMpKjBaj8SlsD0SaNrV9N5p+olXUpWhzTL1BsYaEGSh5Wx7liwokWM3HuBGIJYcK+3YBH6XH0woOa0QxzL5jzbl/7rjdUu9kZSMrSHtIVBcyVPjQwy/fFxBCvtcRa4w==
254711195921	Vincent	50000-100000	G/xBMdldVkVVfENfMyYrcwB1ihb3+u7Nwa22uhrDgXzUzarAJgi6CUG2+rGXxpewgUzKtKpD9OraIAETIjq6YWpZvzSCeJa/j/TU3EQ8rk5KmRFqyyYTBtIFU9i5f5jJkSOtk9H27vZYHCULpvzwpel9Mk5CmuHMGmi5p9RPeeagZkevw+u6jA142SLLBz17eL/HSBCnWIbaUsyQ+PITPIHx590chnnI5ddPWWo015+mHkzOh0UBB0fNBKdIYYQo/37Va/DZIXJ/f4MBr5Yi1qezgiuOs6Uyee9E9ObY2KQj1ypGj9JAQrH1gkiYQMLFlrLrOwTamkjmqRHLnYMhw==
254712234578	Wangui	50000-100000	bUABDWQKGmOEPPufPinXpnBl9UqQO2nZYzzCvXWItXJlqLqX/W/C2KeMYL/CezUaZ8U6im7KDIbx0qcEdd6o2LR/oSXz/mgsbDZeXjAdDRfLKsSuyiWOcpRY8mt1DD0uNYrjguMEOFVC9NoJotugi6fEhcXyUa8AdWpcWUyaw9OJOPVfob1z0qIlfo/6IloZlAkBT6hAVsCuX+M7Ri4cBYo6wOLOaLmXE/b12VyoIkpyasF7tc7B5l4TZsp5hXqBgNfsD2HIpXvyUDRF7oQRYPlbyNbFYTHiXUvn/zFVJM/J9xvA3GUlZ6lhs/7WIx8TKJo6duNsNHRmt69cjtChyg==

Figure 5.5:Account Number Encryption

Figure 5.6 illustrates the encrypted format of the KRA PIN stored in the database.

firstname	krapin	passport	titlenumber
Evalyn	MVR8EXFciTqwp3QNft2KJDLlj0MYWvY5jBy3nIdcnI95Dr7rvGMazY4bjbOQIs0MKge2U7ZiZoXSsTPfT7KMCmINsNnFzMXHE0DpGYAdWlPRTJ3cDvOhh+WJbmgzoVX/Rx2ANJhyT/dZzjELzDWEsPrLBcz9hCH0izHZfWbO1yjGncBFkoHERdVALK3pqFZ2FYTenLK6KKIhYk29mWgqCcmnzE/R+O3z0Sf5vn+SXw3OcsDrue2iwY80LmS71huYmkFps6MMrTf22iHchoHKrkFa+/CHe4/DMhJww0Vq4GkKsY2u0liBuIApYK7Zd7dvWVKUB4dIUZt4acz4yl+V5A==	QUQFzq0n	SNo72dm9LpWN29dffNop5XKx773m6iNb4QHUhVkzo01
Kiplangat	gNsPUQ0i635NYFDvLUHZIar48BBfFsw6g3KiEgV5c8YnrbIgv9l0uY0P1UsuYhZ/Gl1L24gFiTrHeFEXQm+MDj3v4BoYIcc4FUeKpbG/qRuAaVmeiBv9s6/rouA+AFoz6ChxsdSJNTBDSJITjGgsc9o/U746P1NbAmRp0PNr1VZyeb/vxjLB2/hoiPGBialHV5yD+k+3JfCdmBNWnsQCYG/w9O9pIEJLdRFJIQof5MsnhBVemQXkoY/VaQ7A5L9B+s+OPRWuZKmauwGalCYS+jcjv4L1pazmIivgF+Q7P3sLg5L1y2wa3n00NidSixhfJbo6nGY0xRJdUkNTzcVRQ==	cchoK5NssI	RAFMWuybiDGw4MQhNTitQkUzzbE1o8UlrBMKlwq+lVtfE
Julia	ll/KcJ27QyXi/285awhSGZXcNxBDRi94Tk5lrSqUSC+j2Mgok/eTs2Daro8NvEXwKKbG4y4g6lOCoBvRdfQkGJie0G1dIziemBn5s5e/ASaqrI7uvNc7Sid042JtdJtRPkyTjfsemZXQFqowRmCydxE0FlNYprVePfenAoIxONQSB/YJwiCz5a8OKu96WxxTQyx5d5OiS2V50iXYpb60OtrXXsRRuu5qDY2GQ2UndvTOBIlEmW/01bcrCN9nXa1a9Yn30ChqZiZsgLMfU/9cyrQnpbiF6WcsZFoEXWe57onxwRIFBDKWuCmVdxofK61KDdhrsMDG2d9rRXTV2hA0w==	HDzmrwwjoZnzuucO4AtXP5Pfps2oFc7I3aBMxYaR7mU5y99ni5HqGx	

Figure 5.6:PIN Encryption

5.3.3 Opt-in /Opt-out

Opt-in and Opt-out mechanisms were used to obtain people's consent for the collection, use and processing of personal data. Figure 5.7 illustrates the opt-out option that data subjects have and by clicking on the terms and conditions, data subjects indicated their consent to personal data

collection. The opt-in process described an affirmative action by the data subject to offer their consent for data collection. In the same context, the opt-out option enabled the data subjects to withdraw consent for processing activities.

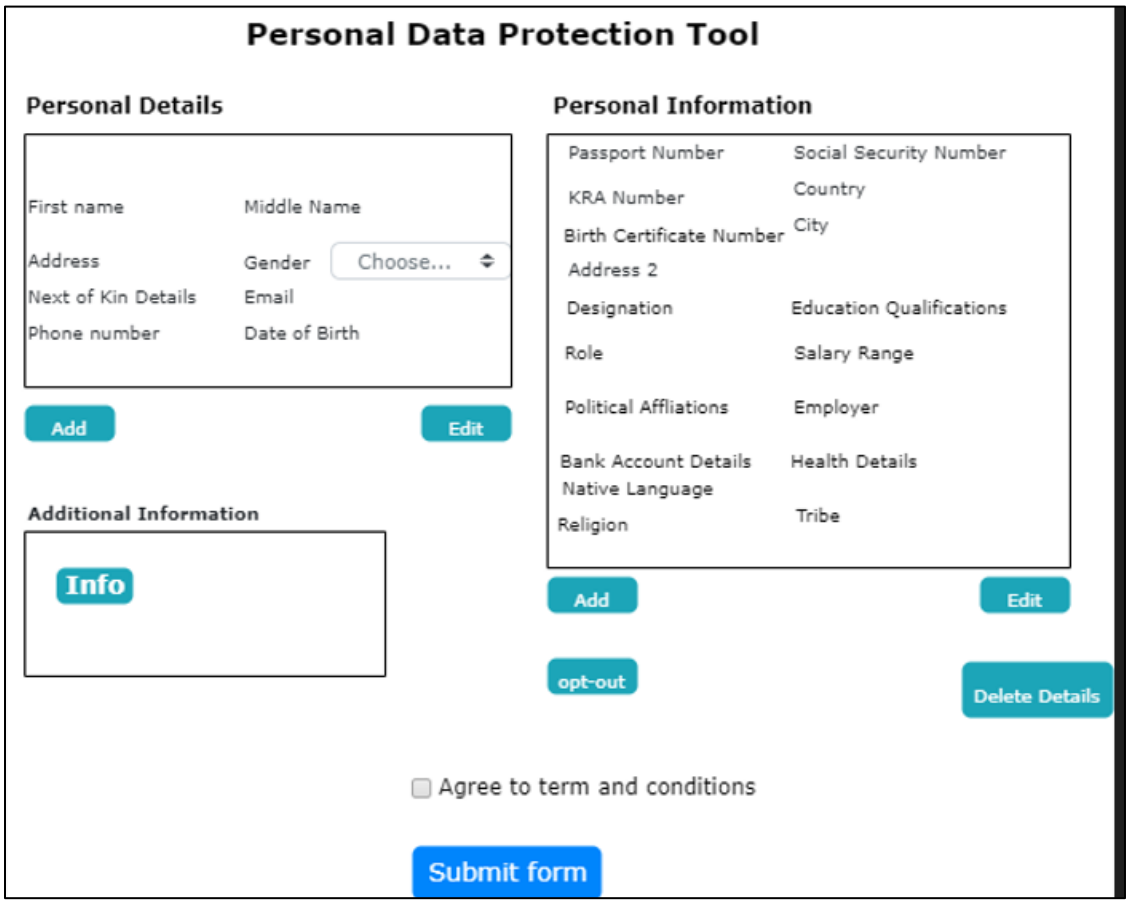


Figure 5.7:Page Indicating Consent

Figure 5.8 illustrates the opt-in method.

```
/**
 * update database
 */
public void opt() {
    query.append(Constants.TABLE_PROFILE_NAME)
            .append(" set accessible = ? where phoneNumber = ?");
    System.out.println(query.toString());
    System.out.println(Arrays.toString(params));
    int res = new JdbcTemplate(new Database().dataSource()).update(query.toString(), this.params);

    if (res == 1) {
        this.response.put("success", true);
        this.response.put("statusCode", "00");
    } else {
        response.put("success", false);
        response.put("statusCode", "57");
    }
}
```

Figure 5.8:Opt-in Method

Figure 5.9 demonstrates the database response when the data processors queried the database to check if a data subject had opted in. The database responded with the “True” Status code.

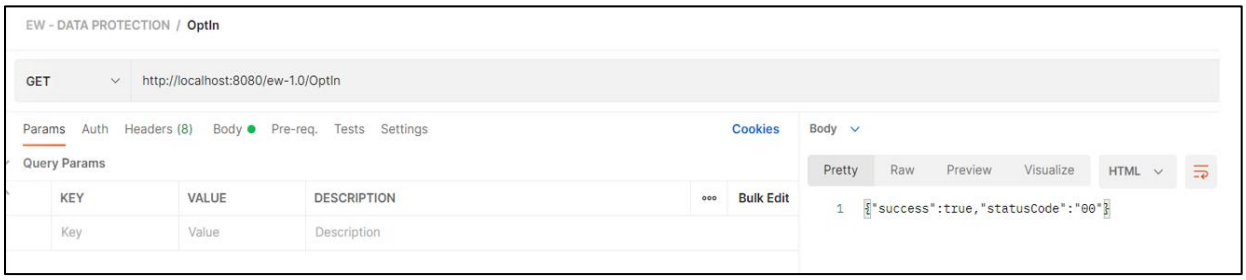


Figure 5.9:Opt-in Response

In the same context, phone numbers were used to check whether the data subject had opted out from personal data collection and processing activities. Figure 5.10 illustrates the opt-out method used to filter the data subjects who had opted out.

```
/**
 * @param phoneNumber
 * @return
 */
public OptProcessor optOut(String phoneNumber) {
    this.params[1] = phoneNumber;
    params[0] = false;
    return this;
}
```

Figure 5.10:Opt-out Method

Figure 5.11 demonstrates the database response when the data processor queried the data base The database responds with the “True” Status code.

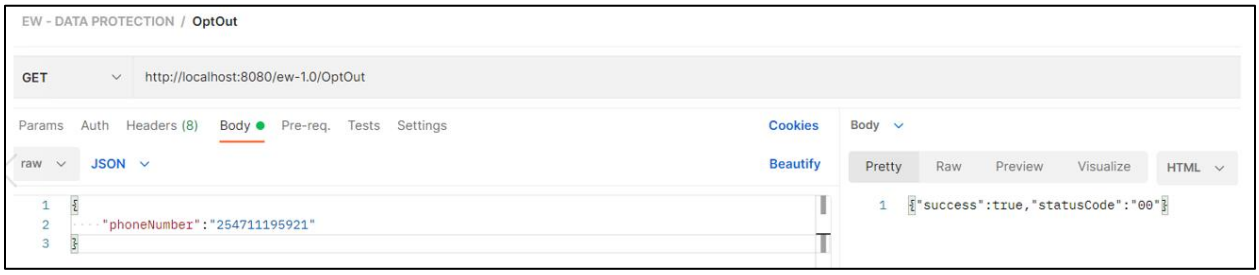


Figure 5-11:Opt-out Response

5.3.4 Personal Data Access

This study applied the use of OTP to authorize data processors to access personal data. This step ensured that only authorized users access the data. This step controls who accessed, modified, deleted, or performed any activity on the personal data. Figure 5.12 illustrates the OTP page.

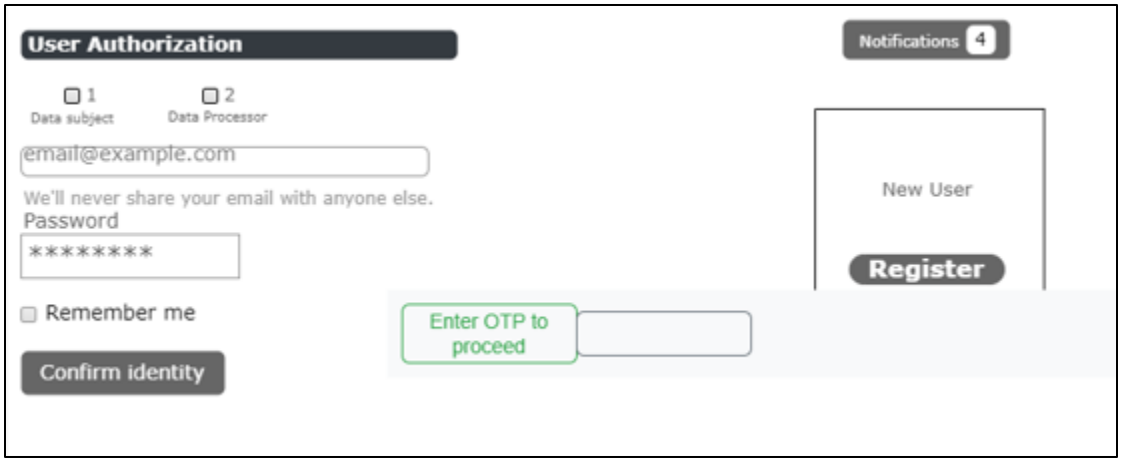


Figure 5.12:OTP Page

Figure 5.13 illustrates the OTP sent to authorize data access.

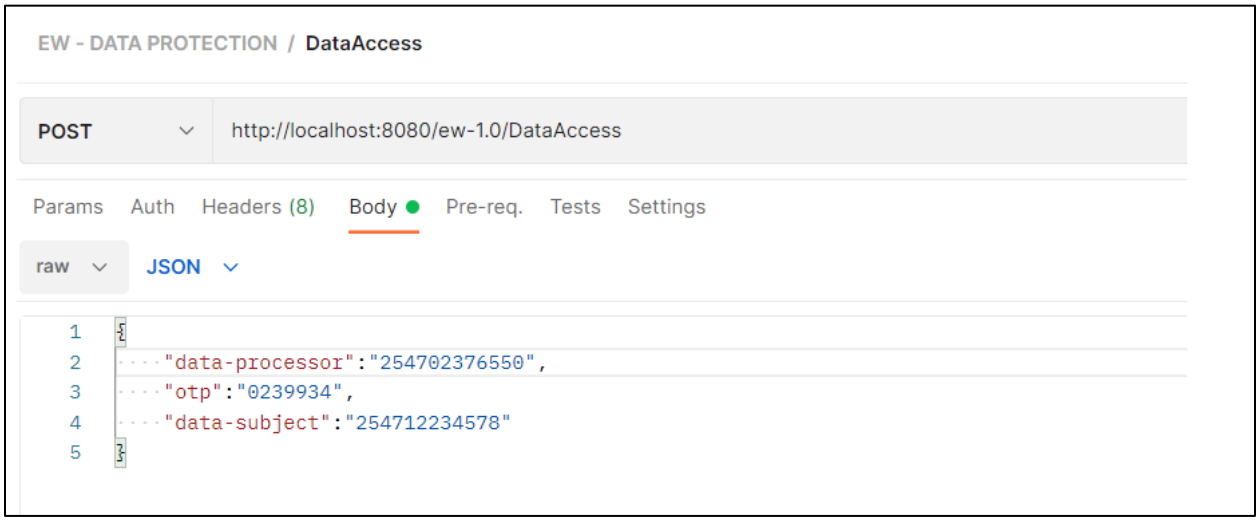


Figure 5.13:Data Access OTP

5.3.5 Personal Data Processing

In this stage, the data inputted during data collection was processed.it involved organizing, editing and dissemination of data. This section covered all the activities described in the study. Figure 5.14 illustrates all the data processing activities which include collecting, viewing, updating, and deleting personal data.

Figure 5.14: Data Processing Activities

Figure 5.15 illustrates how data processing OTP was generated.

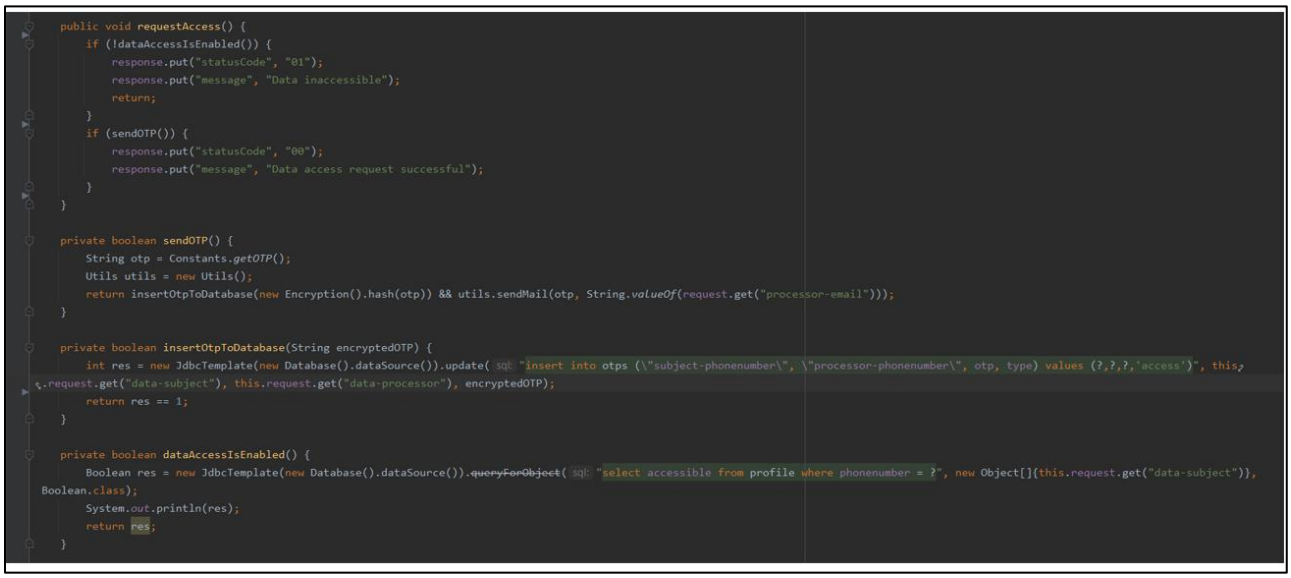


Figure 5.15: Data Processing OTP

5.3.6 Personal Data Deletion

Personal data deletion involved permanent erasure of personal data from the data storage upon data subject's request. This process was carried out under various conditions such as.

1. Personal data was no longer necessary in relation to the purposes for which it was collected.
2. Data subject withdrew consent of data processing.
3. Personal data have been unlawfully processed.

Figure 5.16 illustrates the data deletion request and approval process.

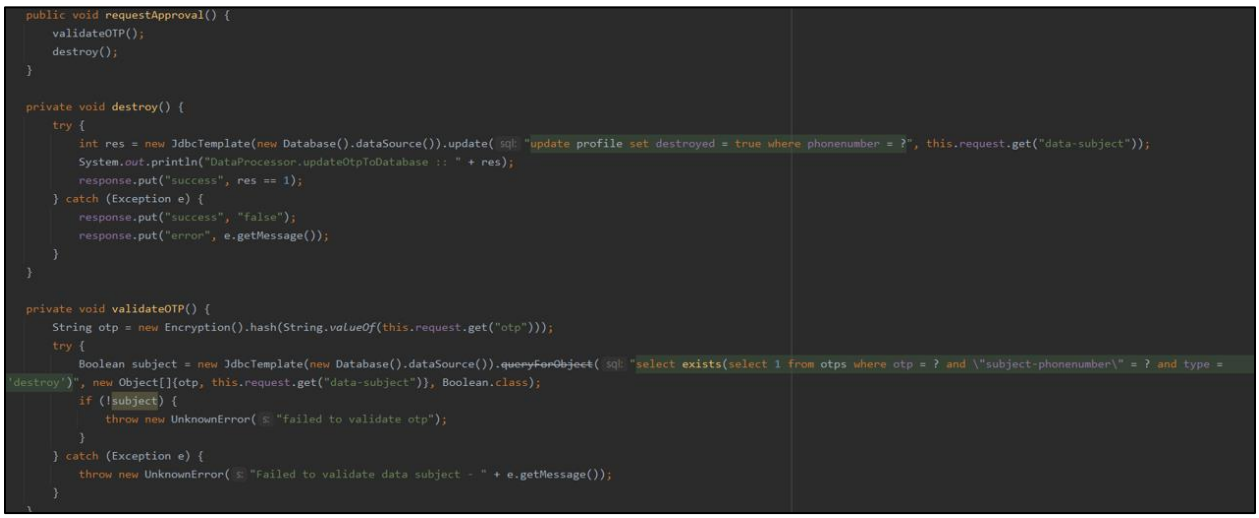


Figure 5.16: Data Deletion Request

5.4 System Testing

Manual and automated testing was carried out to check the functionality, reliability, performance, and the security of the developed tool. Manual testing was carried out to identify possible bugs, issues, and defects of the application.

Different categories of potential users were requested to carry out system testing. These individuals were employees of organizations that collect personal data, moreover some were human resource personnel who are mainly the custodian of personal data within an organization. Further, individuals who have previously shared their personal data with organizations also took part in testing the effectiveness of the developed tool. Table 5.3 illustrated the test cases generated during manual testing.

Table 5.3 :Test Cases

Test Done	Name	Test Description	Pass	Fail
5	Personal Data Encryption	Sensitive Data stored should be encrypted	5	0
3	Personal Data Access	Send OTP to authorize Data Access	3	0
5	Personal Data Processing	Send OTP to authorize Data Processing	5	1
3	Personal Data Deletion	Delete Data Request	3	0
2	Opt-in	Select opt-in field	2	0
2	Opt-out	Select opt-out field	2	0
Total			20	1

The test results were passed for 19 users which represent 95 % of the total tests conducted. A failure result was noted in one user who could not receive the one-time password. See Appendix Further, automated testing was also carried out to test the functionality and reliability of the application. Postman tool was used for automated testing and to sanitize and validate inputs. Figure 5.17 indicates the test results which are denoted by a success message.

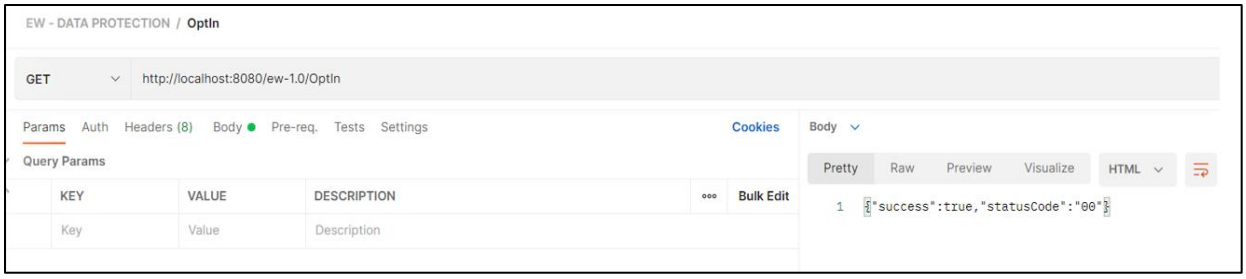


Figure 5.17: Test Results

5.5 System Validation

System validation was carried out to ensure the developed tool achieved the research objectives. This was done by getting feedback from potential users and recording their feedback. This helped in determining whether the developed tool was applicable for secure collection, secure storage and secure processing of personal data as well as ensuring confidentiality and integrity of personal data.

Potential users of the developed tool were randomly selected to carry out system validation. Same as in the system These individuals represented organizations that collect personal data. In the same context, individuals who have previously shared their personal data with organizations also took part in the system validation process. 70% of the users found the tool useful and would recommend it for protecting personal data. Additionally, 83% of the users were able to validate the developed functionalities of the tool. In summary, the results from validation show that the developed integrated tool can be effective in personal data protection. For more information on validation results refer to Appendix B

5.6 System Performance

In this stage, users also verified the performance of the developed tool to ensure that it fulfilled the research objectives. The performance of the developed integrated personal data protection tool was determined by its ability to encrypt personal data to ensure secure storage, the opt-in and opt out mechanisms that allowed data subjects to have control of their data, as well as its ability to send OTPs that regulated access to personal data.

Table 5.4 indicates the performance of the tool in achieving each functionality.

Table 5.4 System Performance

Functionality	Pass Rate
Send and Receive OTP	90.9%
Opt-in and Opt-out Option	100%

Data Encryption	100%
Average Performance	96.9%

5.7 Conclusion

This chapter covered the implementation of the integrated personal data protection tool and provided a detailed description of the hardware, software and network specifications used to build the tool. Further, the chapter provided a breakdown of the modules, interfaces and the processes involved in the tool. To close the chapter, the testing tools and procedures used to validate the performance of the tool were described in detail.

Chapter 6: Discussion of Key Results

6.1 Overview

This chapter analyses the key findings derived from the study and from the development, design, and testing developed integrated tool. The developed tool was reviewed to ascertain that it met the research objectives. The study results were also reviewed to examine whether they concurred with the literature review and that all the research questions were answered.

6.2 Objective One: Identifying Categories of Personal Data

The first objective of the study was aimed at identifying categories of personal data processed by organizations. This objective was achieved through systematic literature review whereby the literature revealed that organizations collect personal data for decision making and advertisement purposes.The common personal data collected to identify someone is the name. name is the most common means of identifying someone. However, a combination of identifiers is used to identify an individual. Such identifiers include but are not limited to personal identification numbers, address information, personal characteristics, and property information.

The literature further categorized personal data into sensitive data and non-sensitive data. Sensitive personal data is a category of personal data that requires extra protection. In today's technological world organizations are collecting, re-using, and processing personal data on an unprecedented scale, without observing data limits, this has led to an increasing concern about the effectiveness of the existing data protection law and the need for a comprehensive tool for personal data protection.

6.3 Objective Two: Reviewing Techniques, Tools and Approaches used by organizations to securely process personal data

The second objective of the study was to review the techniques, tools and approaches used by organizations to securely process personal data. This objective was achieved by systematic literature review methodology in that that the literature identified several techniques and approaches used to securely process personal data. Data encryption, data masking, data anonymization and data pseudonymization are the common tools that organizations use to protect personal data.

The literature reviewed the existing data protection tools are prone to errors and require constant improvement for them to recognize the emerging data threats. Other data protection tools are rigid

because and can be exploited remotely by users. These gaps can be addressed by the development of an integrated tool for protecting personal data.

6.4 Objective Three: Designing, Developing and Testing an integrated tool for Personal Data Protection

The third research objective was to design, develop and test the developed tool for protecting personal data. The objective was achieved by employing the agile development methodology that guided the design, development, and testing of the integrated tool. The tool was designed and modelled using various tools that included flowchart, use case, sequence diagram and ERD diagrams.

Development of the personal data protection tool was done using IntelliJ and Java Object oriented programming language. The APIs were setup using Postman API testing tool. Before delivery and deployment, manual and automated testing were carried out to test the functionality of the tool. User feedback was also collected using standard google forms.

6.5 Objective Four: To validate the effectiveness of the developed tool in securing the processing of personal data

The fourth and final objective was aimed at validating the effectiveness of the developed in securing the processing of personal data. This objective was achieved by collecting personal data from potential users and allowing them to test the tool as well as recording their feedback.

The results from the validation survey showed that an average of 70% of the users found the developed tool relevant and would recommend it to others. The results also showed that the developed tool had a high accuracy rate and can reliably be used to protect personal data.

6.6 Conclusion

This chapter discussed how the research objectives were actualized by identifying the common personal data collected by organizations to identify individuals. Further the chapter pointed out the challenges faced by organizations in securely processing personal data. The chapter also discussed the tools used to design, develop, and test the developed system.

To close the chapter, the tools used to validate the developed application as well as its accuracy were described.

Chapter 7: Conclusions, Recommendation and Future Work

7.1 Introduction

This chapter summarizes the main purpose of the study, which was to develop an integrated tool that helps organizations to securely store and securely process personal data. Through conclusions, this chapter also summarizes the research objectives and how they were achieved through literature review, design, the, development and testing of the integrated tool. Further, this section gives recommendations on the developed tool as well as suggestions for future work that can be done to extend this study.

7.2 Conclusion

This study focused on developing an integrated tool for personal data protection, and applied an integrated approach to ensure secure collection, secure storage, secure access, and secure sharing of personal data. One-time passwords were used to ensure secure access and sharing of data, and RSA encryption technique, was employed to protect all the stored sensitive data. Further, the tool provided opt-in and opt-out options that allow data subjects to have more control of their data.
The literature review revealed the gaps in the existing tools and approaches used to protect personal data prompting the need to develop an integrated tool to address the deficiencies. The developed tool met the functional and non-functional requirements as well as the stated objectives. Further, validation and testing were done where the results proved that the tool can be used to protect personal data.

7.3 Recommendation

The researcher recommends that, while using the developed tool, individuals and organizations should collect and process personal data lawfully to ensure data quality and that they must maintain security safeguards to protect the data. Any personal data collection and processing with malicious intentions is highly discouraged and the researcher will not bear any responsibility. Moreover, the use and application of this tool should be in accordance with the Data Protection Act 2019.

7.4 Future Work

To extend the results from this study, and make the developed tool more robust, future work can focus on creating a backend application that consumes the data sent by the APIs and implements more backend components to make it dynamic.

Moreover, future work can extend the data encryption technique used in the study, to include two different types of encryptions, for instance combine RSA and AES encryption techniques to exploit the advantages of each technique to build a high security encryption algorithm.

References

Absolute. (2019). *Endpoint Security Trends Report.* USA: Absoulute.

Ahmad, N. (2017). Cloud Storage Encryption. *Data Hiding using Encryption Techniques*.

Asamba, M. (2020). Huduma Number to replace national ID from December 2021. *Nairobi*.

Ashford, W. (2020). Mobile Banking Malware Surges in 2019. *Computer weekly*, 3.

Assolini, F. (2015). Benign Feature,Malicious Use. *Securelist*, 1.

Awareness, A. (2018). *The Cyber Attack Cycle.* Washington: Army Strong.

Baguma, R. (2018). Policy Definition. *Information Systems Security*.

Bank, R. (2019). Seven ways of banking spell convenience. *Open a savings account*, 1.

Bank, T. W. (2022). Monitoring COVID-19 impact on households in Kenya. *The World Bank*.

Banker, I. (2020). The rise of digital banking brings fresh security concerns. *International Banking*.

Banking, P. K. (2019). *Three risks presently asscociated with banking o the web.*

Barrett, M. (2019). Cyber Attack Vectors. *Eight Common cyber Attack Vectors and How to Avoid Them*, 1.

Basvala, S. R. (2017). Mobile Application Vulnerabilities. *Mobile Applications-vulnerability assessment through static and dynamic analysis*.

Bearer. (2022). *You should be automating your data flow map*. Retrieved from Bearer: https://www.bearer.com/resources/automatic-data-flow-mapping

Becta. (2019). Redstor Protector. *Good Practice in Information Handling*.

Belal, A. (2017). *Emulation based technique.*

Berdardino, J. (2020). Pseusonymization vs Anonymization. *Analysis of data anonymization techniques*, 236.

Berners, T. (2017). *Personal data report.* World wide web foundation.

Biega, A. (2021). Data minimization. *Operationalizing the legal principle of data minimization for personalization*.

Brandvoice. (2021). T-mobile foe business. *Five reasons why hackers target mobile devices*.

Bronlee, J. (2020). Data Reduction. *Introduction to dimensionality reduction for machine learning*, 1.

Brownlee, J. (2019). Feature selection for machine learning. *Machine Learning Mastery*, 4.

Budek, B. O. (2018). What is reinforcement learning? *Reinforcement Learning*, 3.

Canedo, E. D. (2020). Software Requirements classification using machine learning. *Entropy*.

Casey, K. (2018). Top 7 vulnerabilities in Android applications. *Codersera*.

Centre, W. a. (2018). Definitions. *Personal Data Protection Policy*, 3-4.

CFI. (2018). *What is Mobile Banking?* Corporate Finance Institute.

Cherednychenko, M. (2020). *How to Prevent Cybersecurity Threats on Mobile Banking App.* Softensy.

Chi, C. (2021). A beginners guide to data flow diagrams. *Hubspot.*

Chopra, A. (2018). Introduction. *Security issues of Firewall*, 4.

Christiansen, L. (2020). The importance of data gathering to improve business intelligence. *Data gathering*.

Cisomag. (2021). Client/Server-Side vulnerabilities. *Half of mobile banking apps are vulnerable to fraud Data Theft*.

Coker, J. (2020). Widespread Security in Mobile Banking. *Insecurity Magazine Home*, 7.

Columbus, L. (2019). 10 Ways AI and Machine Learning are Improving Endpoint Security. *Forbes*, 5.

Commission, D. p. (2019). *Guidance on anonymization and pseudonymization* . Data protection commission.

Contributor, T. (2021). *Usecase Diagram(UML Case Diagram)*. Retrieved from WhatIS.com: https://www.techtarget.com/whatis/definition/use-case-diagram

Corley, E. (2019). Information and network security is keyy to corporate counter espionage. *Cybersecurity education guides*.

Creately. (2020, March 2022). *Creately*. Retrieved from Creately Blog: https://creately.com/blog/diagrams/sequence-diagram-tutorial/

Crocker, P. (2021). *PADRES:Tool for Privacy,Data Regulation and Security*.

Cuckoo. (2020). Introduction. *What is Cuckoo?*

Dai, H. (2020). Public key cyptography. *Asymmetric key cryptography*.

Dakota, N. (2020). Seven types of cybersecurity threats. *Seven types of cybersecurity threats*.

Damodaran, A. (2018). Dynamic Analysis. *A comparison of Static,Dynamic and Hybrid analysis for Malware Detection*.

Dasgupta, S. (2019). Phishing. *Ransomware,Phishing and Endpoint Attacks*, 1.

Datarobot. (2020). What is data preparation for machine learning. *Data Preparation*, 1.

Dataworks. (2021). *Integrity in the Data lifecycle*. Retrieved from Data works: https://www.dataworks.ie/5-stages-in-the-data-management-lifecycle-process/

Dearmer, A. (2022). Data Security. *Intergrate*.

Defense, C. D. (2020). Whichh are the OWASP Top 10 mobile app vulnerabilities. *Cypress data defense*.

Diceus. (2019). How banks can overcome mobile banking security risks. *Mobile banking security risks*.

DNV. (2017). The seven phases of a cyber attack. *Digital Solutions*, 1-3.

Dosal, E. (2020). Top 5 computer security vulnerabilities. *Compuquip Cybersecurity*.

Edition, E. (2021). Definition of Systems. *The Economic Times*.

Education, D. o. (2022). *Protecting Student Privacy*. Retrieved from U.S Department of Education: https://studentprivacy.ed.gov/content/personally-identifiable-information-education-records

EFD. (2020). Data anonymization concepts. *EFD data anonymization guidelines*, 1.

Enisa. (2021). *Data pseudonymisation:Advanced techniques and usecases.* Enisa.

Ensight. (2019). Data protection in Kenya:What you need to know. *Ensafrica*.

Eriksson, U. (2018). What is the difference between the functional and non functional requirements importance? *REQtest*.

Esteve, A. (2017). Privacy issues regarding processing of personal data by Google and Facebook from the US and EU respectively. *The business of personal data:Google ,Facebook,and privacy issues in the EU and the USA*, 40.

F.Hayata, F. &. (2017). Performance of classification models. *Supervised Learning Based Detection of Malware on Android*.

Faisandier, A. (2020). *System Architecture.* Stevens Institute of Technology.

Faith, L. (2018). Privacy tools. *Privacy tools*.

Feldhuhn, G. (2020). ECC. *Information Systems Protection*.

Forcepoint. (2020). Heuristic Analysis. *Forcepoint*.

Fowler, K. (2016). Hacktivists. *An Overview of Data breaches*.

Fox, N. (2021). 11 best malware analysis tools and their features. *Threat Detection*, 5.

Fraley, J. (2018). The promise of Machine Learning. *The promise of Machine Learning*, 1.

Ganzer, M. (2020). Vulnerability management challenge :Identifying vulnerabilities. *5 OT Vulnerability management challenges(and how to overcome them)*.

Garjrani, J. (2018). Vulnerability scanners. *Vulnerability scanners -a proactive approach to assess web application security*, 116.

Gavrilova, Y. (2020). What is the goal of ML testing? *Machine Learning Testing*, 5.

GDPR. (2018). *GDPR data controllers and data processors.* GDPR.

Githaiga, J. (2021). Key considerations in the insurance sector. *Data protection in the insurance industry*.

GMS. (2018). The architecture of an SMS malware fraud. *How to address the threat of SMS malware*, 3.

Golchha, P., & Lunia, R. D. (2019). A review on network security threats and solutions. *Types of attacks*, 20-24.

Governance, I. (2022). Six tools to help you manage your organisations security measures and GDPR compliance. *IT Governance*.

GOVUK. (2019). *Data protection.* London: GOVUK.

Gray, B. (2018). Types of data collected. *Digital Farmer profiles:Reimagining Smallholder Agriculture*, 11.

Green, D. (2017). Top 10 vulnerabilities in mobile applications. *White hat security*.

Griffith. (2022). Aim. *Systematic style literature reviews for education and social sciences*.

Guru. (2018). Agile Methodology. *Agile Methodology*, 2.

Hagen, K. (2020). Seven security risks of mobile banking:How to avoid them. *The Ascent*.

Hamid, Y. (2017). Machine Learning. *Machine learning techniques for intrusion detection*.

Hersey, F. (2019). 'Prone to hacking':expert witness in Kenya's Huduma Namba hearings first round. *Biometric Update*.

Hofmann, L. (2021). What is sensitive data? *What is sensitive data and how is it different to personal data?*

Hone, K. (2022). Introduction. *Information security policy what do international information security standards say?*, 402.

Humadi, A. (2020). Data Encryption. *Encryption*.

Hussain, F. (2017). Information Security Policy. *Developing an Information security policy: A case study approach*, 692.

ICO. (2019). What is personal data? *Guide to the General Data Protection Regulation(GDPR)*.

Initiative, O. S. (2019). *Kenyas National Intergrated Identity Management System.* Open Society Justice Initiative.

Insights, K. (2021). Safaricom faces class action suit over massive data breach . *Kenyan Insights*.

International, P. (2018). The keys to data protection. *A guide for policy engagement on data protection*.

Irwin, L. (2022). The GDPR:What exactly is personal data? *IT Governance*.

John, S. K. (2019). Kenya:Lost in transit. *IFLR*.

Jupyter. (2018). The Jupyter Notebook. *Jupyter*, 3.

Kallam, S. (2017). Diffie-Hellman. *Diffie-Hellman:Key exchange and public key cryptosystems*.

Kaspersky. (2018). SMS attacks definition. *SMS attacks and SMS mobile threats*, 1.

Kaspersky. (2021). What is heuristic analysis. *Kaspersky*.

Kazmi, Z. (2019). Introduction. *Cyber Security Analysis of Internet Banking in Emerging countries:User and Bank Perspectives*, 1-2.

Keerthiga, R. (2019). IOT data anonymization. *Analysis of data anonymization techniques for securing sensitive data in IOT environment*.

Khraisat, A. (2019). Intrusion detection systems. *Survey of intrusion detection systems,techniques,datasets and challenges*.

KICTANET. (2019). *Data protection in Kenya.* Nairobi: KICTANET.

Knudsen, J. (2021). Security vulnerabilities are common in bank mobile apps. *Powering smart decisions*.

Krugel, G. T. (2017). Mobile banking stakeholder. *Mobile banking technology options*, 5.

Kulshrestha, S. (2016). What is E-banking? *Internet-Banking -Benefitsand Challenges in an Emerging Economy*, 3.

Kumar, C. (2017). Diifie-Hellman Algorithm. *Enhanced diffie-hellman algorithm for reliable key exchange*, 3.

Kumar, D. (2015). *Network Security Attacks and Countermeasures.* USA: Information Science Reference.

Kumire, J. (2020). How are banks dealing with a rise in cyber attacks? *Dealing with a rise in cyber attacks*, 1-6.

Kyriazoglou, J. (2020). Reasons for protecting personal data. *Why protect personal data*, 2-7.

Lal, N. (2019). Diffie-Hellman Algorithm. *A review of encryption algorithms-RSA and Diffie-Hellman*.

Lavalle, D. (2020). The most common types of network vulnerabilities. *Redteam Security*.

Lonetti, F. (2018). Issues and challenges. *Issues and challenges of access control in the cloud*.

Lord, N. (2018). What is Endpoint Security. *Data Insider*, 3.

Lucidchart. (2021). What is a flow chart? *lucid Diagram*, 2.

Lukie, A. (2015). *Benefits and Security Threats in Electronic Banking.* ISSN.

Luvanda, A. (2014). *Proposed Framework for Securing Mobile Banking Applications from Man in the Middle Attacks* . Nairobi: IISTE.

Majority. (2019). How does mobile banking work? *Majority*, 2.

Malgieri, G. (2020). Data protection and research:A vital challenge in the era of COVID-19 pandemic. *National Library of Medicine*.

Martin, K. (2019). Understanding customer data vulnerability. *Data privacy:Effects in customer and firm performance*, 37.

Matplotlib. (2020). *Matplotlib:Visualization with Python.* Matplotlib.org.

Mbanaso, U. (2018). Personal data privacy. *Personal data privacy and security*, 1.

Melnick, J. (2018). Top 10 most common types of cyber attacks. *Netwrix Blog*, 5.

(2018). *Modern Endpoint Management.* USA: The Entreprise Strategy Group.

Monteclaro, A. (2018). What is Riskware? *Malware Fox*, 4.

Mostert, M. (2020). Introduction. *Bigdata in medical research and EU data protection law:challenges to the consent or anonymise approach*, 1-5.

Muruganandam, K. (2018). *Data Encryption and Decryption by using triple DES and perfomance analysis of crypto system*, 25-31.

Muthoni, K. (2022). *Court dismisses plea to roll out Huduma Namba over data safety.* The Standard Newspaper.

Newbanking. (2021). *All you needed to know about personal data.* Newbanking identity.

Ng, C. (2020). A guide on the data lifecycle:Identifying when your data is vulnerable. *Inside out security blog*.

NIST. (2019). Overview. *Data Encryption Standard*, 144.

Niswonger, B. (2021). Disadvantages of antivirus. *Seven disadvantages of a free antivirus*.

Njana, A. (2021). Kenya cracks down on digital lenders over data privacy issues. *Startups*.

Odamah, S. (2020). Blueborne. *Top 10 android vulnerabilities*.

OECD. (2020). *`Personal data use in financial services and the role of financila education.* OECD.

Olufohunsi, T. (2019). Encryption. *Data Encryption*.

Orcutt, M. (2019). *Once hailed as unhackable,blockchains are now getting hacked.*

Organization, M. (2015). *Overview of How Cyber Resiliency Affects the Cyber Attack Lifecycle.* USA: Mitre Corporation.

P.Sangster, H. (2018). Network Endpoint Assessment. *Endpoint Security*, 4.

Paganini, P. (2013). Phishing. *Modern Online Banking Cyber Crime*, 1.

Paloalto. (2021). *Zero day attacks.* Paloalto networks.

Park, T. (2019). Network Security vs Endpoint Security. *Evantage Technology*, 5.

Paro, A. (2021). Is your business in danger from the bad actors dropping 22 million revords on the dark web in 2020? *Data Security*, 1.

Phua, C. (2018). Risks. *Protecting organisations from personal data breaches*, 14.

Poremba, S. (2021). What is Personally Identifiable Information? *First Savings Bank*.

Rabi, J. (2022). Data Masking process. *A study on dynamic data masking with its trends and implications*, 19.

Returns, G. (2020). Advantages of Mobile Banking. *Good Returns*, 3.

Rodriquez, A. (2016). The Big challenge. *Malware Analysis and Detection on Android*, 1.

Roh, Y. (2018). Data Discovery. *A survey on data collection for machine learning*.

Salim, P. S. (2018). Integrated and dynamic method. *Integrated static and dynamic analysis for malware detection*.

Satori. (2022). How does access control protect data. *Satori*.

Scikit. (2020). *Mchine Learning in Python.* Scikit-learn.

Seaborn. (2020). *Seaborn:Statistical Data Visualization.* Seaborn.

Security, H. (2020). Organizations Struggle with Patching Endpoints against Critical Vulnerabilities. *HelpNet Security*, 1.

Sedletsky, V. (2019). Five Tools for a Defense in Depth Strategy for Endpoints. *The CyberArk Blog*, 2.

Senan, F. (2019). The top Ten Machine Learning for EDR . *Endpoint Detection and Response:Why Machine Learning?*, 3.

Seymour, J. (2020). History of Electronic Health Records. *Electronic Health Records*, 202.

Sharma, S. (2018). Proposed Methodology. *Enhanced RSA Algorithm for Data Security in Cloud*, 65.

Sheldon, R. (2021). Data lifecycle management. *Tech Target*.

Shen, J. (2018). Literature Review. *Examining the risks of mobile banking applications through blog mining*, 3.

Shree, K. (2017). Introduction. *Detection of malware Applications in mobile devices using supervised machine learning*, 3.

Simmons, G. (2018). Introduction. *Symmetric and Assymetric encryption*.

Singletons. (2017). Introduction. *Object-Oriented Security*, 3.

Slate, S. (2018). What is endpoint security. *Endpoint Security:An overview and look into the future*, 2.

States, O. o. (2022, April 25). *Data Protection*. Retrieved from Organizations of America States: http://www.oas.org/dil/data_protection_public_private_sector.htm

Stealthlab. (2021). SQL injection. *Cybersecurity threats and attacks:All you need to know*, 5.

Sturt, C. (2021). Systematic literature review. *Literature Review:Systematic Literature Review*.

Sturt, C. (2022). Systematic literature reviews. *Literature reviews:Systematic Literature reviews*.

Svajcer, V. (2016). The Smartphone as an Emerging Threat. *When Malware goes Mobile*, 2.

Target, T. (2021, April). *What is considered Personal Health Information*. Retrieved from Tech Target: https://www.techtarget.com/searchhealthit/definition/personal-health-information

Taylor, H. (2020). *The Missing Report.* USA: Prey.

Technologies, P. (2019). H. *Vulnerabilities and Threats in Mobile Applications*, 5.

Technology, S. (2016). Methodology. *Development Methodology*, 3.

Technopedia. (2021). Network Design. *Technopedia*.

Thengade, A. (2018). Integrity Checking. *Virus detection techniques and their limitations*, 4.

Tierney, M. (2021). Security and compliance. *Data security explained:challenges and solutions*.

Trust, B. (2017). Endpoint Security. *Beyond Trust*, 7.

Tunggal, A. T. (2021). What is a vulnerability? *Upguard*.

Udaidullah, M. (2016). Classification of cryptography. *A review on Symmetric key encryption techniques in cryptography*, 44-45.

Verma, P. (2019). Endpoint Virus vs Endpoint Security. *Endpoint Security*, 3.

Villanove. (2021). What is cyber espionage? *Information security management*.

Vizard, S. (2019). Black friday, data concerns,online sales. *Excellence in marketing*.

Wallace, F. (2019). Cybersecurity and natural disasters. *How natural disasters affect cybersecurity*, 5.

Webroot. (2019). *Understanding Endpoints and Endpoint Security.* USA: Webroot.

WFP. (2018). *WFP guide to personal data protection and privacy.* Rome,Italy: World Food Program.

Whitney, L. (2019). Cyberattaks against endpoints rising. *TechRepublic*, 3.

Wigmore, I. (2020). Security by Design. *Whatis.com*.

Williams, S. (2021). *Mobile Device Biggest Entreprise Security Threat.* Australia: Techday Limited.

Wills, T. (2019). *Mobile Banking:Emerging Threats ,Vulnerabilities and Counter-measures.* Information Security Media Group.

Wireshark. (2021). About Wireshark. *Wireshark*, 1.

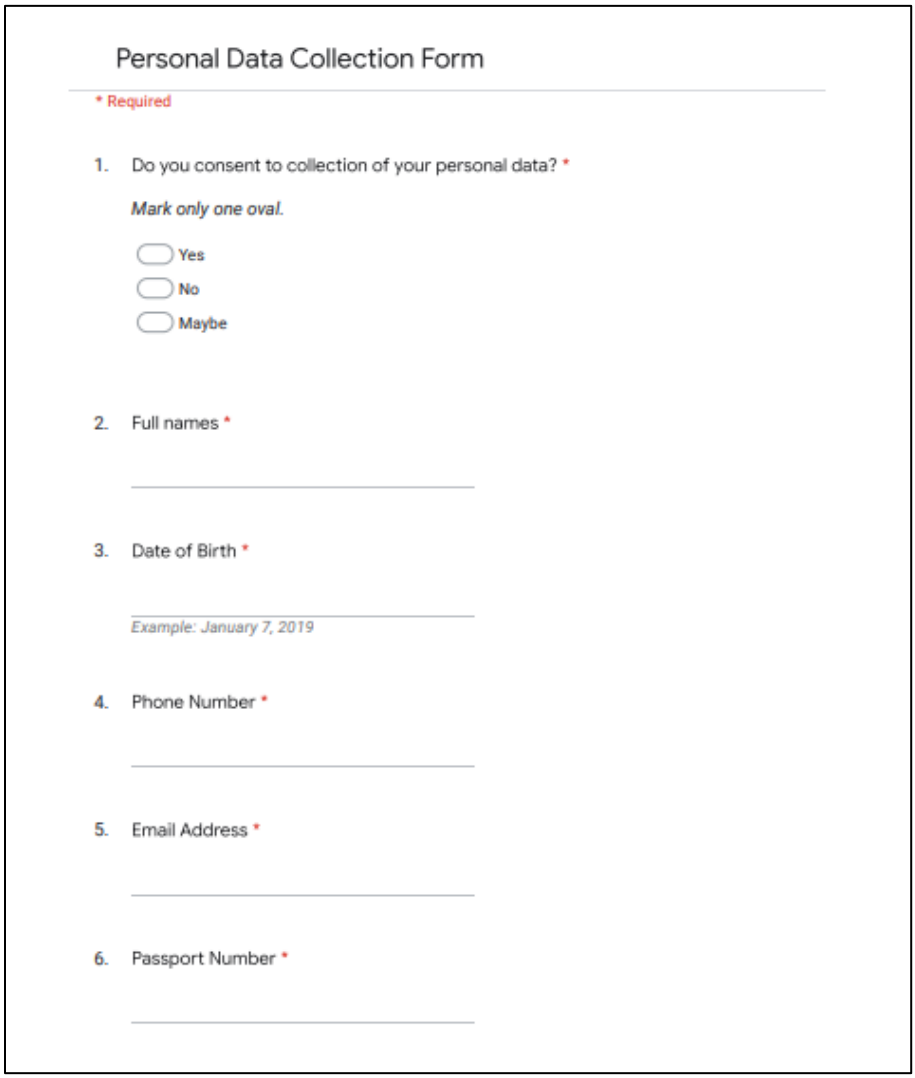
Yehya, D. (2020). Definition. *AES encryption: Study and evaluation*.

Zavrak, S. (2015). *Adware:A Review.*

Zeeshan Ahmand, A. K. (2020). Introduction. *Network intrusion detection system:A systematic study of machine learning and deep learning approaches*, 3-5.

Appendices

Appendix A: Personal Data Collection Form



Personal Data Collection Form

* Required

1. Do you consent to collection of your personal data? *

   Mark only one oval.

   - ◯ Yes
   - ◯ No
   - ◯ Maybe

2. Full names *

3. Date of Birth *

   Example: January 7, 2019

4. Phone Number *

5. Email Address *

6. Passport Number *

7. KRA Number *

_____

8. Identification Numbers *

_____

9. Employment Status *

Mark only one oval.

- ◯ Full time
- ◯ Part time

10. Bank Details *

_____

11. Bank Account Name *

_____

12. Bank Account Number

_____

13. Gender

Mark only one oval.

- ◯ Female
- ◯ Male
- ◯ Other: _____

14. Disability Status *

15. Addresses *

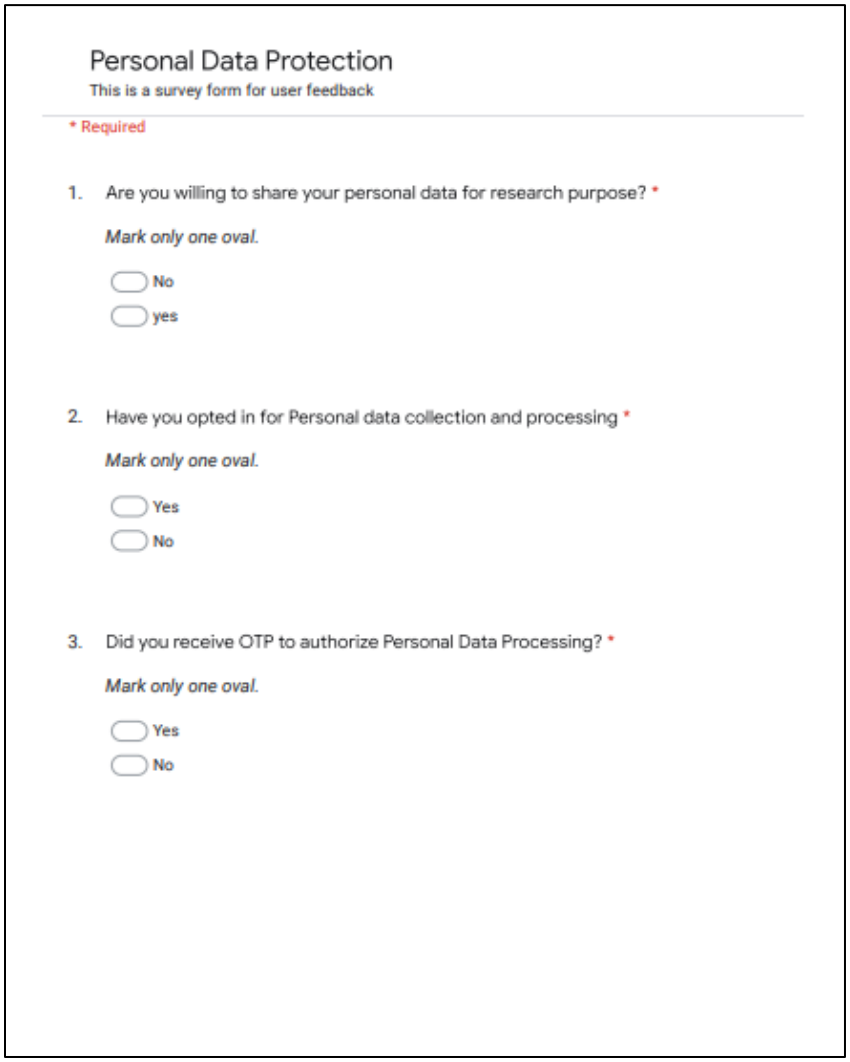
16. Profession *

Mark only one oval.

- ◯ Government Official
- ◯ Consultant
- ◯ Domestic Worker
- ◯ IT Specialist
- ◯ Teacher
- ◯ Student
- ◯ Financial Sector
- ◯ Other: _____

17. Comments

Appendix B: Personal Data Protection User Feedback



Personal Data Protection

This is a survey form for user feedback

* Required

1. Are you willing to share your personal data for research purpose? *

Mark only one oval.

- ◯ No
- ◯ yes

2. Have you opted in for Personal data collection and processing *

Mark only one oval.

- ◯ Yes
- ◯ No

3. Did you receive OTP to authorize Personal Data Processing? *

Mark only one oval.

- ◯ Yes
- ◯ No

4. I would recommend this tool for personal data protection *

Mark only one oval.

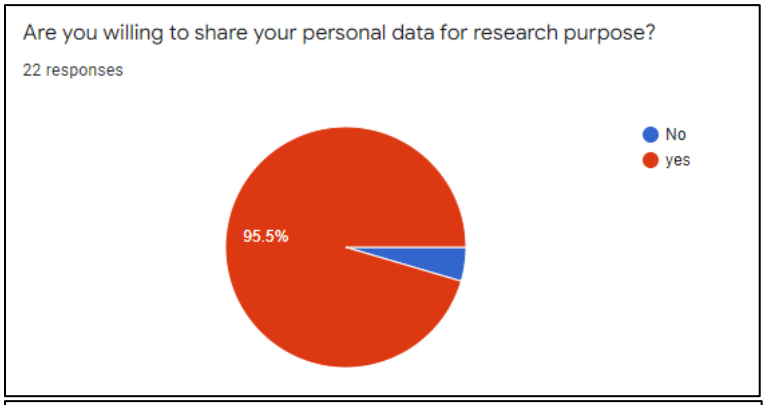
- ◯ Agree
- ◯ Neutral
- ◯ Strongly agree
- ◯ Disagree
- ◯ Strongly disagree

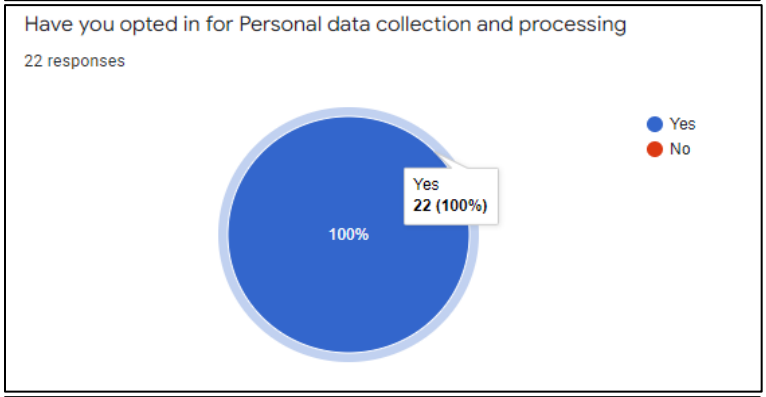
5. How effective is the tool in protecting your personal data *

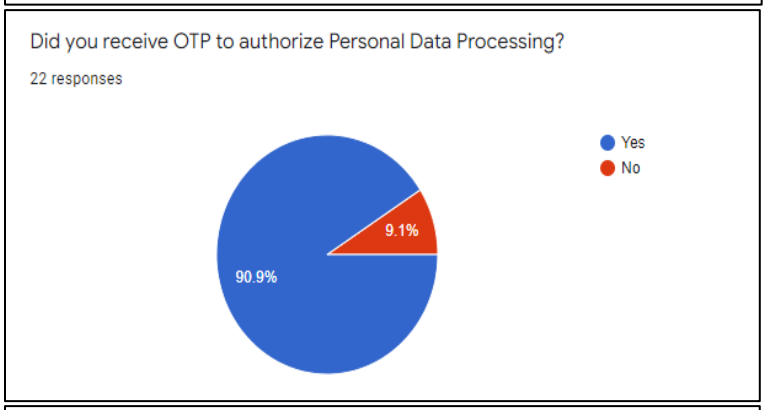
Mark only one oval.

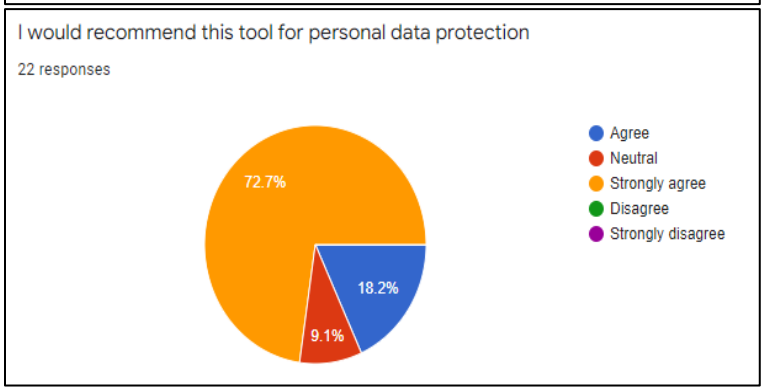
1	2	3	4	5	6	7	8	9	10
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

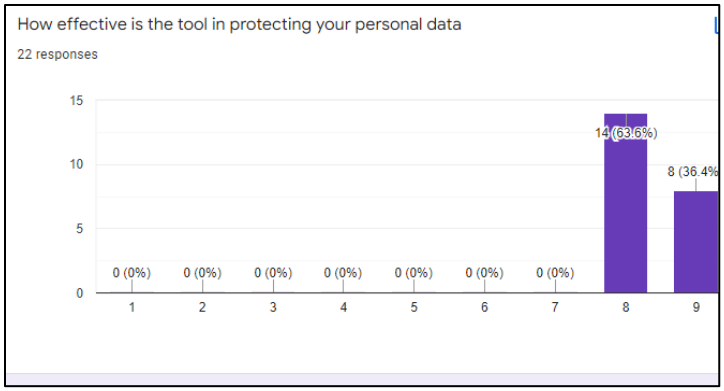
Appendix C: Validation Test Responses










How effective is the tool in protecting your personal data
22 responses
15
10
5
0
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
14 (63.6%)
8 (36.4%)
1
2
3
4
5
6
7
8
9

Appendix D: Similarity Report

Document Viewer

Turnitin Originality Report

Processed on: 17-Apr-2023 4:26 PM EAT
ID: 2067189931
Word Count: 17739
Submitted: 1

INTERGRATED PERSONAL DATA PROTECTION TOOL.pdf By Evalyn Wangui Wanjiru

Similarity Index	Similarity by Source	
24%	Internet Sources:	19%
	Publications:	9%
	Student Papers:	13%

Appendix D: Ethical Approval

Strathmore UNIVERSITY

25th July 2022

Ms Wangui Evalyn,
ewangui84@gmail.com

Dear Ms Wangui,

RE: Integrated Personal Data Protection Tool

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU- master's** research proposal. Your application reference number is **SU-ISERC1443/22.** The approval period is **25th July 2022 to 24th July 2023.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-ISERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-ISERC

Cc: Prof Fred Were,
Chairperson; SU-ISERC

STRATHMORE UNIVERSITY INSTITUTIONAL SCIENTIFIC AND ETHICAL REVIEW COMMITTEE
(SU-ISERC)
25 Jul 2022
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA