2020

Application of permissioned block chain technology on population data consolidation and sharing

Omoka, Richard Siang'ani
Faculty of Information technology
Strathmore University

APPLICATION OF PERMISSIONED BLOCKCHAIN TECHNOLOGY ON POPULATION DATA CONSOLIDATION AND SHARING

Omoka Richard Siang'ani

Research proposal submitted in partial fulfillment of the requirements of the Degree of Master of Science in Information Technology at Strathmore University.

Faculty of Information Technology, Strathmore University, Nairobi, Kenya

2019

Declaration

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, the research proposal contains no material previously published or written by another person except where due reference is made in the research proposal itself.

Signature ……  …… Date:…… April 27th, 2020……

Student Name: Omoka, Richard Siang'ani

Registration Number: 114835

SUPERVISOR'S DECLARATION

This research proposal has been submitted for review with my approval as a university supervisor.

Signature ……  …… Date ……April 27th 2020……

Prof. Ismail Ateya Lukandu, D. Sc.

Director, Office of Faculty Affairs

Associate Professor, Faculty of Information Technology

Abstract

Population registers should provide the single source of truth for data regarding each resident of a jurisdiction of the register, over the lifetime of the individual. This data can then be shared and used by government agencies and private organizations regarding matters concerning the individual. In Kenya, however, data regarding an individual is collected by multiple government agencies resulting in duplication (of effort and data) and data inconsistency. The multiple collections of population data result in an individual having multiple valid identification documents. The use of relational database management systems, which have shortcomings in support for temporal data as well as no inbuilt security and auditability capability makes relational database management systems ineffective in the storage of population data. Lack of clear policy and standards; interoperability issues and data security are among the challenges affecting data sharing among government agencies. Blockchain technology, a shared, immutable, distributed ledger that facilitates the process of recording transactions and tracking assets in a business network, is a promising technology in the management of population registers. Blockchain technology has inbuilt capacity to solve most of the problems inherent in the current systems especially duplication, tampering, and sharing of data. This research, therefore, through the development of a prototype based on permissioned blockchain technology, explores the viability and validity of permissioned blockchain technology, in storing, securing, auditing, and sharing of population data to achieve the single source of truth of the population register. The prototype, implemented using a local installation of hyperledger fabric, enabled consolidation of data since all invited participants on the permissioned blockchain network were able to write data to the single blockchain. The invited participants were also able to read data off the chain based on defined access control rules therefore achieving a uniform standard for data sharing. Provenance, a key quality of blockchain was leveraged to track an individual's data changes over time, with the current block holding the latest records about the individual, yet still maintaining the historical chain of an individual's data changes. This was a key outcome especially because it solves the inability of relational database systems to support temporal data. This model for data consolidation and sharing was found to be simple in design and implementation since it provided a standard way of reading and writing data to the chain through the use of RESTful APIs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations/Acronyms

BEV	–	Blockchain-Enabled E-Voting
CIA	–	Confidentiality, Integrity, and Availability
CIS	–	Credit information sharing (CIS)
CISKenya	–	Credit Information Sharing Association of Kenya
CRBs	–	Credit reference bureaus
CRD	–	Civil Registration Department
DIS	–	Department of Immigration Services
DLT	–	Distributed Ledger Technology
FSDKenya or FSD	–	Financial Sector Deepening
HELB	–	Higher Education Loan Board
ICT	–	Information and Communication Technology
IPRS	–	Integrated Population Registration System
KLRC	–	Kenya Law Reform Commission
KRA	–	Kenya Revenue Authority
MIRP	–	Ministry for Immigration & Registration of Persons
MSP	–	Membership Service Provider
NCRAs	–	National Civil Registration Authorities
NDRS	–	National Digital Registry System

NEMIS	–	National Education Management Information System
NRB	–	National Registration Bureau
NHIF	–	National Hospital Insurance Fund
NIIMS	–	National Integrated Identity Management System
NSSF	–	National Social Security Fund
NTSA	–	National Transport and Safety Authority
ODIHR	–	Office for Democratic Institutions and Human Rights
OSCE	–	Organization for Security and Co-operation in Europe
PDTG	–	Presidential Digital Transformation of Government
PKI	–	Public Key Infrastructure
SMEs	–	Small and medium enterprises
SSOT	–	Single Source of Truth

Definition of Terms

Blockchain - A blockchain is a distributed ledger, structured into a linked block list. Each block contains a set of ordered transactions (Xu, Weber, & Staples, 2019).

Distributed Ledger - A distributed ledger is an immutable and append-only datastore of transactions running across many peer nodes (Xu, Weber, & Staples, 2019).

Membership Service Provider (MSP) - In Hyperledger Fabric, the membership service provider is a component that supports network participation operations such as user authentication through issuing and validating certificates by abstracting away cryptographic operations and related protocols.

Permissioned blockchain - Permissioned blockchains are distributed ledger systems in which the membership in some or all roles is on an invite-only basis (Falazi, Hahn, Breitenbucher, Leymann, & Yussupov, 2019).

Permissionless blockchain - Permissionless blockchains are public distributed ledger systems in which the membership is open to any participant and no participant is fully trusted (Falazi, Hahn, Breitenbucher, Leymann, & Yussupov, 2019).

Public Key Infrastructure - a technology for digital authentication where trusted parties issue digitally signed certifications certifying that a given digital cryptographic key belongs to a given user or device. This digital key can then be used as a digital identity for the user (SSH Communications Security Corp, 1995).

Chapter 1: Introduction

1.1 Background

In Kenya, the journey towards a single and comprehensive national population register has been a long and tortuous one. Many efforts and programs adopted to achieve a single source of truth about Kenyan residents have either done little or failed to achieve the desired goal. Despite numerous efforts and programs to establish a central database register of all Kenyan citizens, there are still cases of duplicate identification documents; forgery of identification documents; identification documents of deceased persons still in use; and the existence of undocumented citizens (KLRC, 2019).

Population registers are records of inhabitants of jurisdiction and have a goal of obtaining a precise identification of each individual. Governments or local authorities are required by law to keep a record of both citizens and foreigners residing in the country. (Poulain & Herm, 2013). In Kenya, several laws give provisions for some form of population registration. The Registration of Persons Act mandates the authorities to register people and issue identity cards. (Laws of Kenya, 1949). The Births and Deaths Registration Act mandates authorities to account for all births and deaths within the country and any other related matters. (Laws of Kenya, 1928). The Kenya Citizenship and Immigration Act mandates the authorities to account for matters relating to citizenship, issuance of travel documents, and immigration (Laws of Kenya, 2011).

Whereas the laws prescribe the mandatory population registration process, the forms of implementation have been numerous and varying. Of the numerous programs to manage the population registration activities, three initiatives that included substantial use of information and communication technology are the Integrated Population Registration System (IPRS) (KLRC, 2019), the National Digital Registry System (NDRS) (Omusolo, 2014), and most recently the National Integrated Identity Management System (NIIMS) (KLRC, 2019). Poulain and Herm (2013) have noted the big role information and communications technology has played in the centralization of population registers in many countries. Population registration involves activities that include collection, verification, collation, storage, monitoring, integration, harmonizing, securing, auditing, and sharing of population data. These activities are well suited

for the use of information and communication technologies. Blockchain technology is one such method.

Blockchain is a shared, immutable, distributed ledger that records transactions and tracks assets in a business network. The qualities of blockchain technology of being distributed, immutable, consensual, and transparent (Gupta, 2018) can be leveraged to address the shortcomings of current population registration systems. Regarding population data consolidation and sharing, blockchain's immutability is important because transactions cannot be tampered with by any participant once recorded onto the ledger. Blockchain's append-only nature ensures that only corrections to erroneous transactions are made by adding a new transaction to the chain instead of editing the erroneous transaction, in which case both transactions will be visible in chronological order.

The distributed nature of blockchain can easily facilitate data sharing between various entities such as banks that rely on population data. Provenance is also a key strength of blockchain technology because it guarantees historical transactions are maintained therefore ensuring auditability. The same cannot be said for existing Relational Database Management Systems (RDBMS) currently used in managing population data because they lack support for temporal data (Date, Darwen, & Lorentzos, 2003). RDBMS also lacks some key inbuilt accuracy and consistency checks, for example, referential integrity has to be explicitly set by the database administrator. RDBMS also have interoperability issues, especially in the case of different vendors.

For this research, we propose and use a variant of the blockchain technology called private or permissioned blockchain. Whereas permissionless or public blockchains such as Ethereum are public ledgers where anyone can read or write to, permissioned or private blockchains require authorization from existing network participants to join the network. This authorization ensures that all participants are known to each other. (Kakavand, Kost De Sevres, & Chilton, 2017). This specifically important because population data is sensitive information that needs protection and therefore only known entities can join the private network. Distributed Ledger Technology (DLT) and Blockchain intrinsically possess the qualities suitable for consolidation, securing, and distributing population data and this research tries to prove this theory.

1.2 Problem Statement

Population registers should provide a single source of truth for data regarding each resident of a jurisdiction of the register, over the lifetime of the individual (United Nations, 1969). This data can then be shared and used by government agencies and private organizations regarding matters concerning the individual (Verhoef & Van de Kaa, 1987). In Kenya however, data regarding an individual is collected by multiple (government) agencies resulting in duplication (of effort and data) and data inconsistency. The multiple collections of population data result in an individual having multiple identification documents (Kenya Law Reform Commission,2019). Yego (2014) notes that the lack of clear policy, standards, communication protocol, interoperability issues, and data security are among the challenges affecting data sharing among institutions. The use of relational database systems, which have shortcomings in support for temporal data (Date, Darwen, & Lorentzos, 2003), makes RDBMS ineffective in the storage of population data.

1.3 Aim

Based on these challenges, this study seeks to design a prototype of a population data management system based on the permissioned blockchain technology to explore the validity and viability of blockchain technology in consolidating, securing, auditing, and sharing of population data.

1.4 Research Objectives

i. To review the strengths and weaknesses of existing population data consolidation drives.
ii. To review existing consolidated data access and sharing frameworks.
iii. To propose the application of Blockchain Technology on population data consolidation and sharing.
iv. To develop a prototype for population data consolidation and sharing based on the Permissioned Blockchain Technology.
v. To test the prototype.

1.5 Research Questions

i. How effective are existing population data consolidation programs?
ii. How efficient are existing frameworks in data access and sharing of population data?
iii. How can a Blockchain-based approach to population data consolidation improve effectiveness?
iv. How effective is the Blockchain-based approach to population data consolidation?
v. How effective is the Blockchain-based prototype in addressing problems in population data consolidation and sharing?

1.6 Justification

The results of this research form the basis of the adoption of this distributed technology in the management of population data, especially in designing a central population database on a single source of truth concept. The distributed nature of the blockchain technology coupled with smart contract and consensus-building would form a framework for data sharing between government and third-party data consumers such as banks, insurance companies, education institutions, hospitals, and the like. The immutable nature of blockchain technology guarantees security and privacy of population data, as well as keeping track of all changes to population data over time.

1.7 Scope and Limitation

This study will be focused on testing the viability of blockchain technology as an option for population data management. The study is limited by time constraints as per academic requirements.

Chapter 2: Literature Review

2.1 Introduction

Population registers, their importance to governments and countries are reviewed. The concept of *Single Source of Truth* and its importance to the management of population data is reviewed. A review of existing laws in Kenya that govern population registration is done. The ICT-based systems and programs the Kenyan government has rolled out to achieve the Single Source of truth for Kenya population data are reviewed. A review of existing applications of blockchain technology in population data management is done. Finally, a case is made for the use of blockchain technology in population data management. A framework on how blockchain technology can be used to enhance population data consolidation and sharing is proposed.

2.2 The Population Register

2.2.1 Population Register Defined

Population registers are records of inhabitants of jurisdiction and have a goal of obtaining a precise identification of each individual. Governments or local authorities are required by law to keep a record of both citizens and foreigners residing in the country. (Poulain & Herm, 2013). In the report *Methodology and Evaluation of Population Registers and Similar Systems*, prepared by the United Nations, Population Register has been defined as a data system supporting continuous recording and/or organized linking of selected information about residents of a country such that it is possible to determine current information regarding the size and characteristics of that population at selected time intervals. (United Nations,1969).

2.2.2 Guidelines on Population Registration

According to the Organization for Security and Co-operation in Europe's (OSCE) Office for Democratic Institutions and Human Rights (ODIHR), an efficient population register system must include the following characteristics: Registration is mandatory for the entire population (citizens and legal residents); The population register is complete and includes citizens living abroad (excluding visitors and non-legal aliens). The efficient population register system is continuous and permanent. Different from a census, the population register collects personal data regularly and is permanent under the law. There are established data protection measures in place, data processing is legitimate, and data security is guaranteed. There is only one record per person. The

individual's data is registered in one place and one place only. Multiple uses are made of a single registration. There is a legal provision against multiple registrations (OSCE /ODIHR, 2009).

2.2.3 Role and Importance of Population Registers

Population registers exist to maintain a precise identification of an individual over time. This, therefore, means the population registration is a continuous process of updating an individual's data from birth to death including data about place and date of birth, sex, citizenship, marital status, language, ethnicity, date of arrival/departure, education levels, parity, occupation, activity status, marriages, divorces, migrations changes in address and place of death among other items. This information must be reliable because it is vital for administrative purposes in Government, especially in budget planning and tax planning. Other administrative areas where population registers are useful include, the establishment of personal identification systems, social safety nets, and social welfare, education, voting among others. Education institutions also rely on the population register information to issue admission documents to children joining these institutions (United Nations. Statistical Division, 2001). Population registers also serve another vital role as the database for demographic statistics where it is used to produce statistical overviews of its population (Verhoef & Van de Kaa, 1987). It can be used in estimating the size of a population and its socio-demographic structure at a time, as well as estimating the changes in population and its different components (Poulain & Herm, 2013).

POPULATION REGISTER CONTENT	
CAPTURED ENTRIES	Date & Place of birth Adoptions Sex/Gender Date of arrival/departure Citizenship(s) Marital status, marriages, divorces Language(s) Ethnicity Educational attainment Parity Activity status Occupation Migrations Changes in address

Figure 2. 1:Contents of a Population Register

India's ID system, known as *Aadhaar* (Hindi for "foundation"), is an example of the role population registers play to governments in the area of official personal identification and social welfare (Gates, 2019). The Aadhaar project arose from the need to weed out fake, duplicate, and fraudulent beneficiary records in the government's social welfare program for the underprivileged. Aadhaar project introduced the Aadhaar Number, which is a unique, 16-digit identification number for every Indian citizen. The number represents twelve identity parameters corresponding to the demographic information (Computer Weekly, 2011).

2.2.4 Population Register as the Single Source of Truth

The Single Source of Truth (SSOT) is the concept where certain data has a single primary and official central source to be referenced by dependant data consumers who can be humans or software, as the true and current version of that data. SSOT, therefore, seeks to mitigate the unnecessary duplication of data. This essentially forms the primary "source of truth" such that updates to the data elements in this primary location propagate to the whole system without the possibility of duplication enabling the data consumers to always have the current state of the data at all times. (Beck, Dumay, & Frost, 2015). The single source of truth trait is a desirable quality for a population register. Verhoef & Van de Kaa (1987) noted that a population register consists

of coordinated linkage in which notification of life events, often originally recorded in separate files, is automatically and continuously transferred to a central file throughout the lifetime of an individual. Figure 2.2, adapted from the Swedish Tax Administration report, illustrates the concept of a single source of truth.

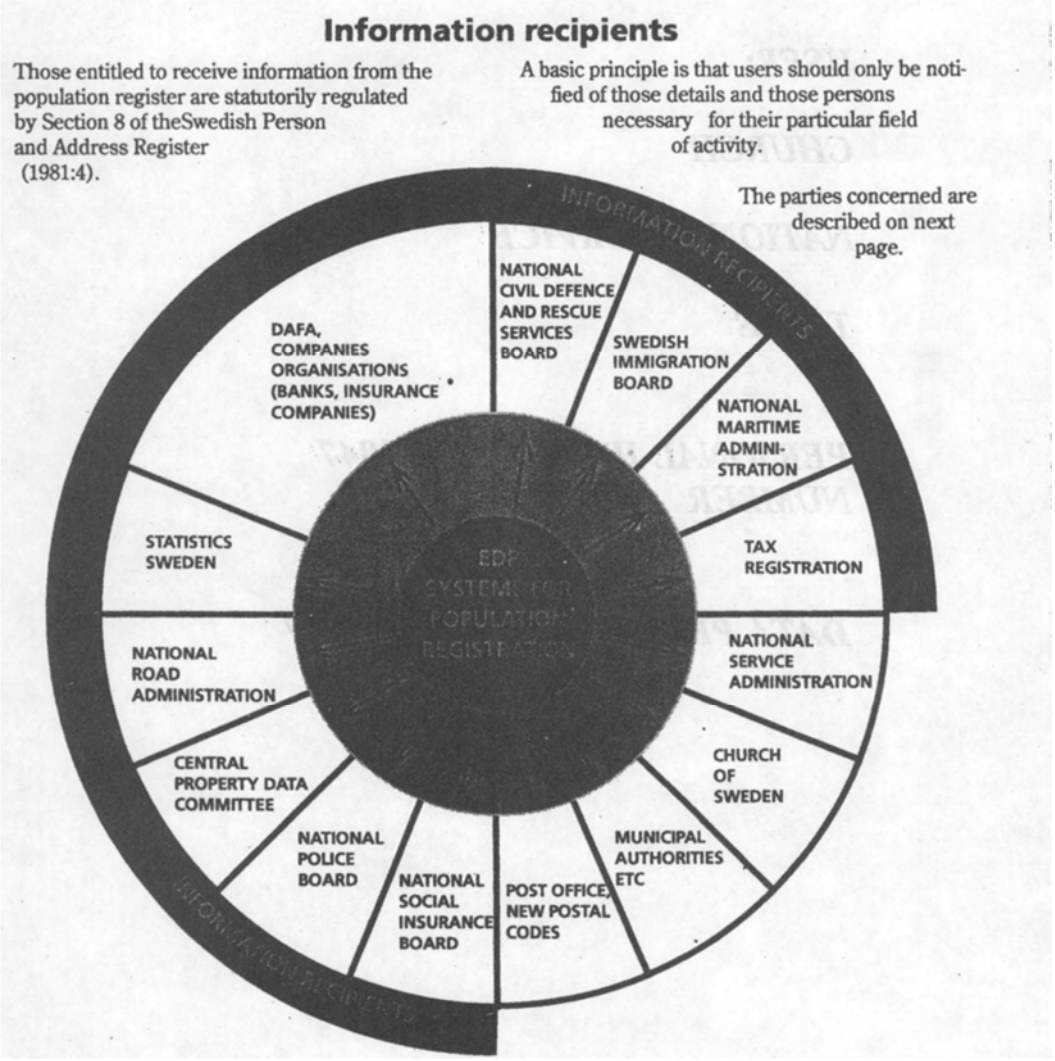


Figure 2. 2: Data recipients from EDP System

2.3 Population registration in Kenya

During the past two decades in particular "the number of national identification and similar programs has grown exponentially. . . to the point where almost all developing countries have at least one such program" (Gelb and Metz, 2018). Kenya, in particular, has had several such programs.

2.3.1 Laws Governing Population registration in Kenya

In Kenya, several laws give provisions for population registration. The Registration of Persons Act mandates the authorities to register people and issue identity cards. (Laws of Kenya, 1949). The Births and Deaths Registration Act mandates authorities to account for all births and deaths within the country and any other related matters. (Laws of Kenya, 1928). The Kenya Citizenship and Immigration Act mandates the authorities to account for matters relating to citizenship, issuance of travel documents, and immigration (Laws of Kenya, 2011). The Kenya Citizens and Foreign Nationals Management Service Act created a state corporation comprising of several departments, namely: The Civil Registry Department (CRD), Department of Refugees Affairs (DRA), Immigrations Department (ID), Integrated Population Registry Service (IPRS), and National Registry Bureau (NRB) (Laws of Kenya, 2011).

2.3.2 Review of Digital National Population Registers in Kenya

Of the numerous programs to consolidate the population registration activities, four initiatives that included substantial use of Information and Communication Technology (ICT) are The Integrated Population Registration System (IPRS), The National Digital Registry System (NDRS), National Integrated Identity Management System (NIIMS) and The National Education Information Management System (NEMIS).

2.3.2.1 The Integrated Population Registration System (IPRS)

The IPRS was developed by the EDAPS consortium from Ukraine, linking together the main repositories of identification and citizenship status (Breckenridge, 2018): Civil Registration Department (CRD), Department of Immigration Services (DIS), and National Registration Bureau (NRB)s. The Kenya Law Reform Commission (2019) notes that the fact that IPRS sought to only consolidate data from these three primary population registration agencies, it was limited in capacity. The IPRS relied on information already collected by existing agencies, whose validity could not be guaranteed, instead of doing fresh registration of citizens. This essentially carried over potentially inaccurate population register information into the IPRS compromising its data integrity.

2.3.2.2 The National Digital Registry System (NDRS)

The National Digital Registry System (Omusolo, 2014) had three objectives: strengthening national security; enhancing efficiency, effectiveness, and accountability in service delivery; and provision of affordable and easy to access citizen-centric services. Breckenridge (2018) opines that, whereas this program had a legal backing in the Kenya Citizens and Foreign Nationals Management Service Act; and political goodwill as the foundation of Presidential Digital Transformation of Government (PDTG), it failed because of corruption and vested interests by involved parties (Breckenridge, 2018).

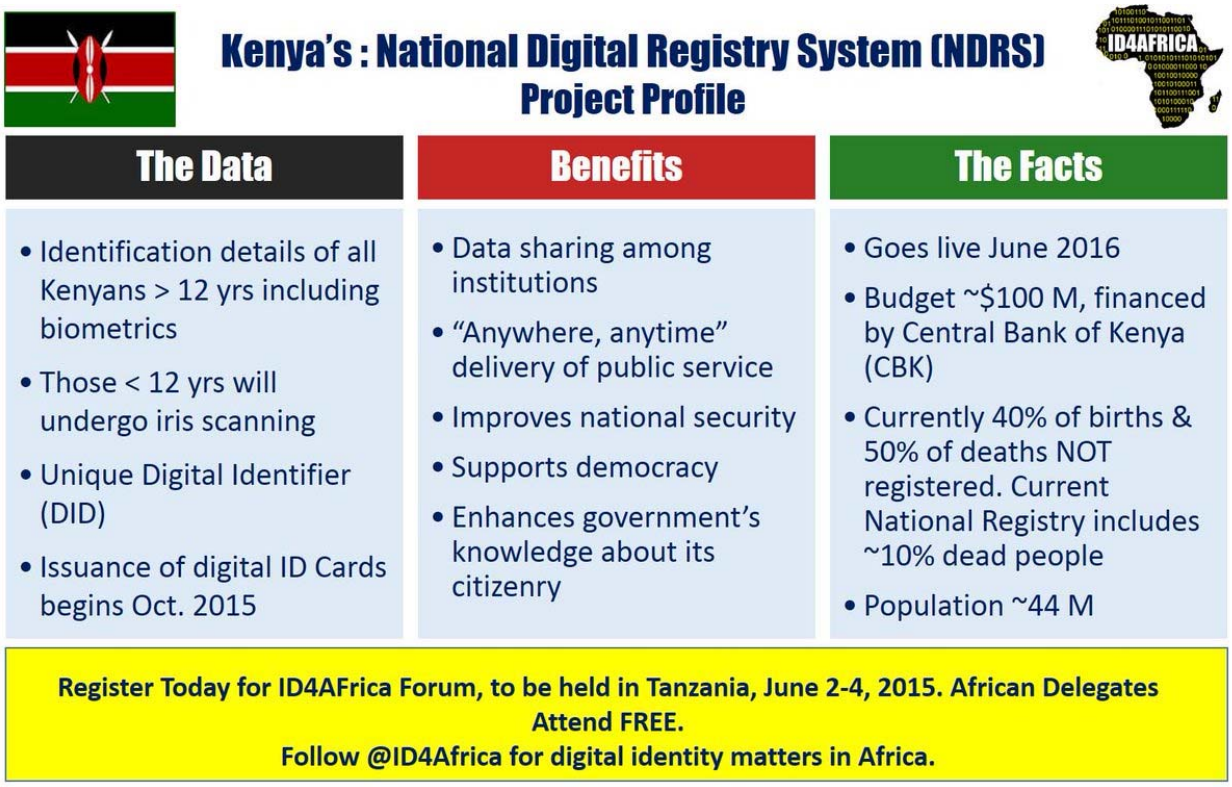


Figure 2. 3: The NDRS Project Profile

2.3.2.3 The National Integrated Identity Management System (NIIMS)

The Registration of Persons Act, Cap.107 created the National Integrated Identity Management System (NIIMS). This program was launched in early 2019. Some of the functions of NIIMS include: to create, manage, maintain and operate a national population register as a single source of personal information of all Kenyan citizens and registered foreigners resident in Kenya; To assign a unique national identification number to every person registered in the register; To

harmonize, incorporate and collate into the register, information from other databases in government agencies relating to registration of persons.

2.3.2.4 The National Education Information Management System (NEMIS)

The National Education Information Management System (NEMIS), is a digital registration platform for school going children which is used to register pupils and students from pre-primary to university and uses the birth certificate as the reference document for registration (*Digital Identification Document (ID) & Citizenship Consultative Meeting*, 2019)

2.4 Population Registration Empirical Models and Algorithms

2.4.1 Trends in Population Registration and identity management systems and processes

The report *Identity Management in 2030* (National Office for Identity Data, 2015), which explores directions for identity management by the year 2030, identifies the following trends in Identity Management Systems and Processes:

Identification is relative, quantitative, and dynamic - The Identity Management process will not only rely on official records but also on other sources such as public, restricted, structured, and unstructured sources, to quantify the evidence of identity from the information obtained. This will make identity management not only more accurate but also relative and dynamic, especially because data contained in identity management infrastructures may be incomplete or even outdated.

Decentralized System - The Identity Management solutions adopted by the National Civil Registration Authorities (NCRAs) will make use of cloud-computing technology to decentralize backups of data with compliance to the data protection and privacy regulations. Such decentralization may be in different forms. An example would be backing up data of one NCRA on the infrastructure of other NCRAs. Whereas the downside of such a dynamic backup system may be costly in terms of storage and processing, it has an upside of enabling ease of reconstruction of the database of any NCRA in case of catastrophic data loss.

Digital ID infrastructures take over the paper-based processes - The use of digital technology in identity management will replace the current paper-based processes. This will introduce efficient interoperability between ID infrastructure as well as between people involved in the identification processes and the systems. It will become easy to detect identity fraud and minor irregularities, a phenomenon called net-widening, thus drastically reducing the grey areas of tolerance between illegal and legal identity

Data integrity is managed and assessed – There will be standard and recommended practices including regulations for data protection and privacy which NCRAs of every country will be expected to comply with. The NCRAs will be also expected to offer efficient and trusted ID management services.

Unique Personal Number (UPN) - The enrolment of persons by NCRA will adopt the use of a unique personal number attributed to the person throughout their lifetime. This unique personal number will be linked to biometric and personal data containing evidence of the identity of the person.

Biometric technologies - With advances in technology, biometric technologies have become robust and adaptable to daily lifestyles and therefore provide a suitable option for implementation in enhancing the identity management processes.

2.4.2 The Aadhaar System - Review of India's identity management system

In 2009, India started the collection of demographic information including biometrics of its residents which would be used in initiating the identity management system. The demographic and biometric data such as name, gender, date, and place of birth, fingerprints, iris scans, and photos for every resident is collected and stored by the Unique Identification Authority of India (UIDAI). This creates a digital repository that is easily accessible for identity management and is linked to various government and financial services including banking. The result of this registration is the assignment of a unique 12-digit, biometrically verifiable personal identity number for every Indian citizen (Perrigo, 2018). The Aadhar number is used for biometric authentication wherein a person's iris scan or fingerprint is matched with their Aadhaar number against the central repository (Jain, 2019).

The Aadhar system appears to be the “Single Source of Truth” of citizens' data in India. Officially, the Aadhar number is not considered as proof of identity but rather proof of citizenship, however, there has been an increasing number of digital services linked to the Aadhaar number for identity verification. Some examples include: The DigiLocker service that enables downloading government-issued documents; various payment services for bank transfers; payment gateways for government-to-citizen transfers; Offline authentication service with QR codes; Mobile application lock and unlock service; Aadhaar verification history logs; VirtualID creation; Mobile OTP for authentication; electronic Know Your Customer (e-KYC); and eSignatures and document signed locker (Young, 2019).

The biggest concerns of the Aadhar system revolve around data security and privacy (Jain, 2019). Reports of data leaks, accidental exposures, duplication of Aadhar cards, insecure government apps, and third-party integrations have been widely reported, casting doubts on the security of the Aadhar system (Tech2 News, 2018).

Another concern revolves around errors of inclusion and exclusion. This is where either deserving cases of welfare programs are excluded (exclusion errors) or undeserving cases are included (inclusion errors) (Muralidharan, Niehaus, & Sukhtankar, 2020).

2.4.3 A Review of Identification, Civil Registration and Vital Statistics in Estonia

In Estonia, the activities involved in the identification, population registration, and vital statistics are carried out within two key subsystems. The first is the population registration system, which is aimed at recording and certifying births, deaths, and other vital events occurring in the population. The second is the identity management system, the purpose of which is to provide legal identification and associated documents to the population. These subsystems complement each other rather than duplicating their functions (Aguilar Rivera & Vassil, 2015). The core of the identity management system includes the Personal Identification Code (PIC) and the Electronic Identification card (eID). The Electronic Identification Card is not only a physical identity document but also has electronic chips that facilitate secure authentication and legally binding digital signatures.

According to Aguilar Rivera & Vassil (2015), the interconnection between the Population Register and identity management systems was operationalized by X-Road—an innovative data exchange

platform behind eGovernment. X-Road is, therefore, the core of integration; enabling data sharing between authorized state and non-state parties.

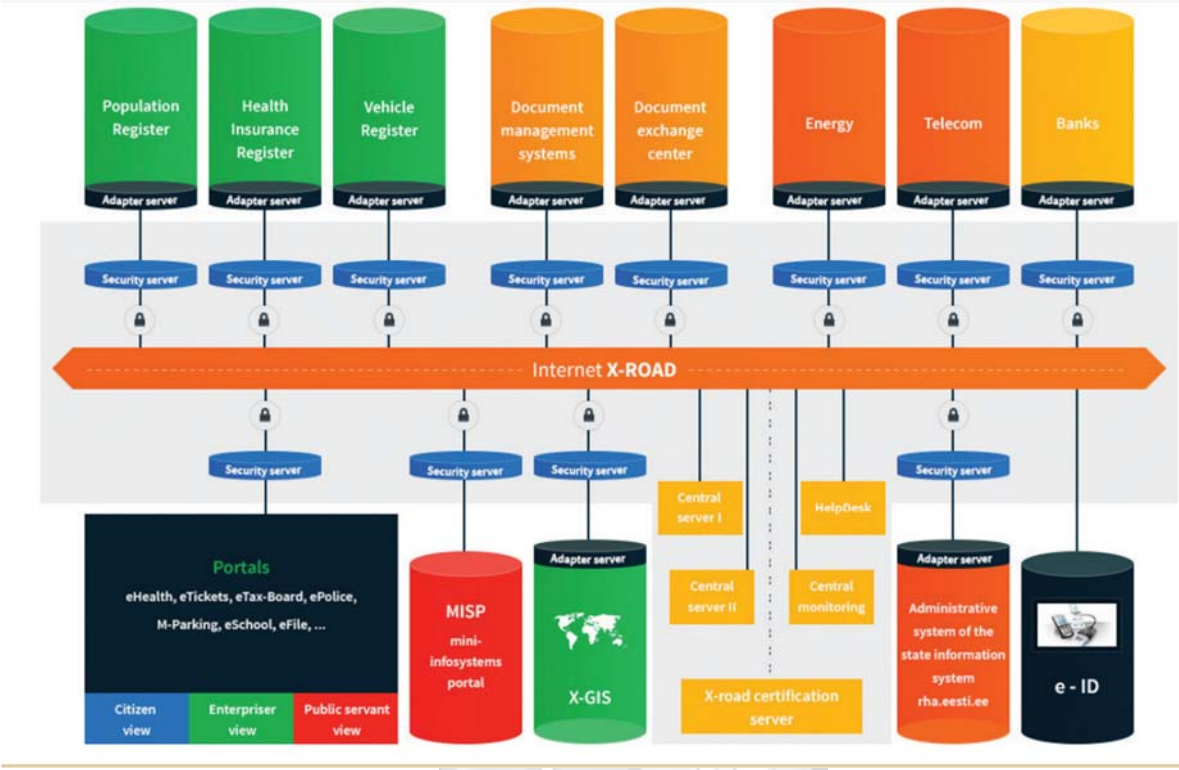


Figure 2. 4: X-Road Inter-Connectivity

2.5 Blockchain Technology

2.5.1 Blockchain Technology Defined

Blockchain is a shared, immutable distributed ledger that enables the process of recording transactions and tracking assets in a decentralized network (Gupta, 2018). A distributed ledger is a distributed database that is shared, replicated, and synchronized among the participants of a decentralized network (Brakeville & Perepa, 2019). The blockchain acts as a single source of truth, and participants in a blockchain network can view only those transactions that are pertinent to them.

There are two types of blockchain systems: Public or Permissionless blockchains such as Bitcoin (Bitcoin, 2009) and Ethereum (Ethereum, 2020); and Private or Permissioned blockchains such as Hyperledger Fabric (The Linux Foundation, 2009).

Permissionless blockchains have certain shortcomings that prevent their use for population data management. These shortcomings include the fact that they are public and therefore anyone can read from /write to them, and the consensus mechanism of Proof-of-Work is complex and energy-intensive. Permissioned blockchains, on the other hand, are cheaper and faster especially because of the less requirement of consensus building. They have a limited number of readers and writers who are preauthorized to join the network by a central authority.

Wust & Gervais (2018) have noted the similarities between private blockchains with centralized databases especially with the existence of a centralized authority. However, an offshoot of permissioned blockchain known as Consortium Blockchain drastically diminishes the existence of a central authority. In a consortium blockchain, the consensus-building process is controlled by a set of pre-authorized nodes. For example, in a consortium blockchain consisting of fifteen institutions, each running a node, for a block to be validated it can be set that ten of these institutions must sign the block (Kulhari, 2018). This research, therefore, proposes the use of Consortium Blockchain for population data management.

The basic principles underlying blockchain technology effectively making it an attractive option for transaction management include security, consensus, provenance, immutability, and finality (Gupta, 2018).

2.5.2 Do you need Blockchain Technology for Population Data Management?

Just like many other technologies in use today blockchain is merely a technology. Therefore, the decision of whether to adopt blockchain should entirely be based on the goal that is to be achieved.

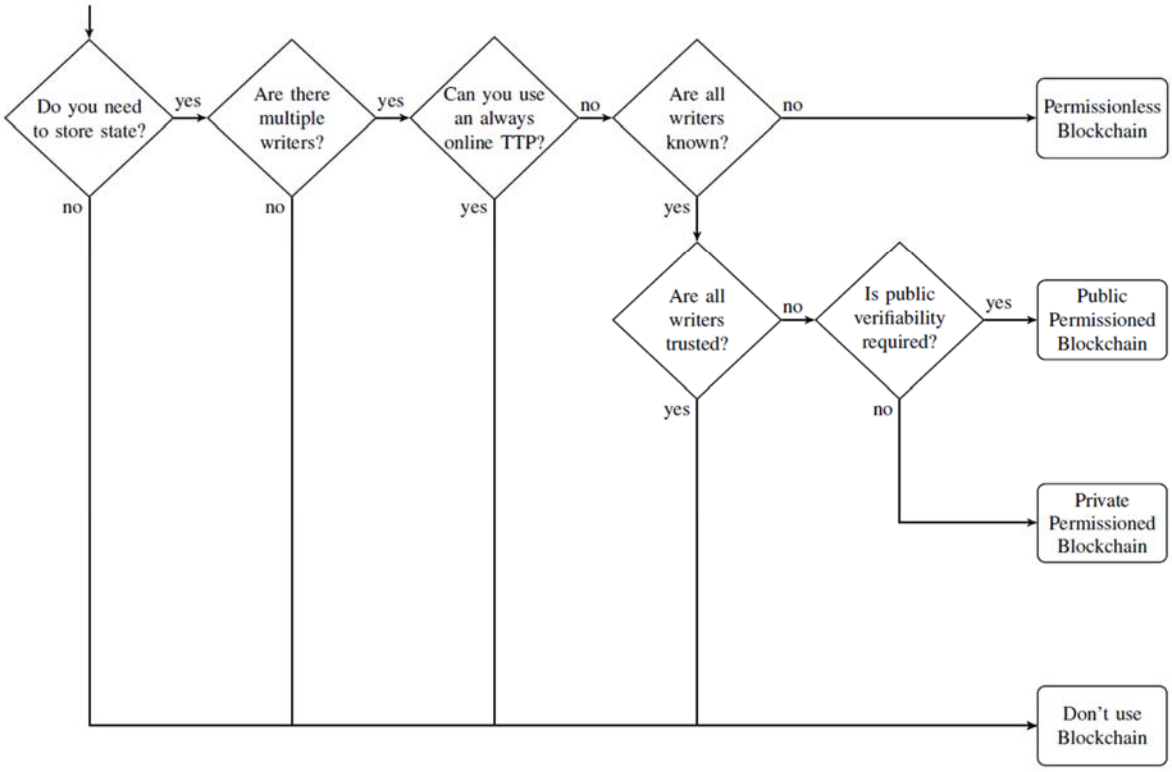


Figure 2. 5: Blockchain use case determination

Brakeville & Perepa (2019) propose a set of five questions to determine whether a use case is a good fit for blockchain. In the use case for population data consolidation and sharing, the questions can be answered as follows:

i. Is a business network involved?

   Yes. Population registers are created and managed by governments. The data needs to be available to other government agencies, for example, Immigration agencies for travel documentation. Financial institutions also need access to this information for credit purposes. Education institutions need access to this information to enroll students. These examples indicate the availability of a network through which population data is shared with third parties to help in the identification and verification of an individual.

ii. Is consensus used to validate transactions?

   Not necessarily. For a permissioned blockchain, participants are pre-approved to join the network and therefore participants involved in a transaction can validate it. In the case of population data management, the "government node" in the blockchain network is

automatically the validator of records; a process called selective endorsement (Jaeger, 2018).

iii. Is an audit trail, or provenance, required?

Yes. To maintain integrity for population data, provenance is a key quality. The data should be captured and continuously updated with key life events over the lifetime of the individual: from birth to death.

iv. Must the record of transactions be immutable, or tamper-proof?

Yes. Population data must be immutable to maintain integrity. Once records have been captured, they should not be modified, rather, new entries should be made to correct an error. In this way, avenues to temper with population data are blocked.

v. Should dispute resolution be final?

Not necessarily. For a permissioned blockchain network, participants can resolve disputes through means such as reversal.

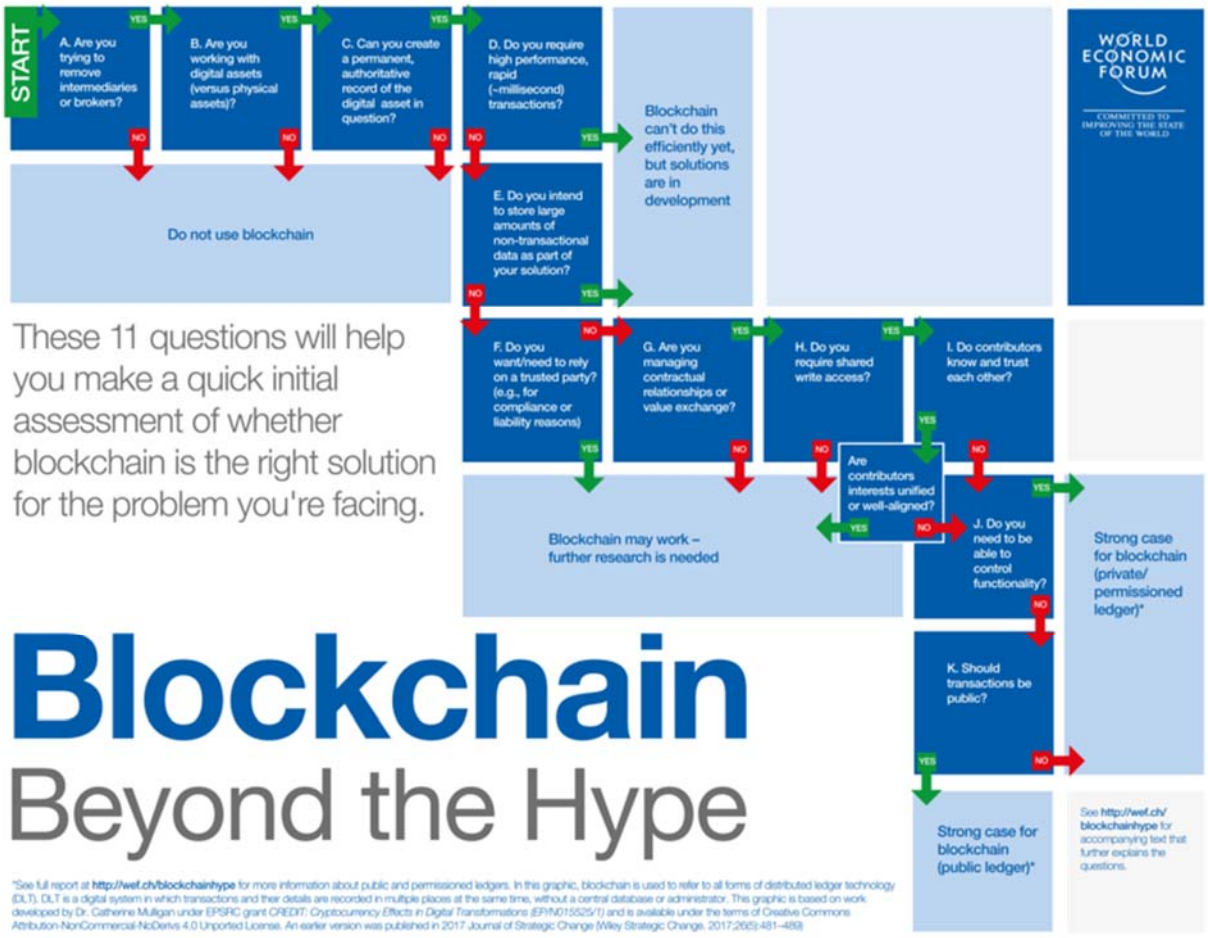


Figure 2. 6: The Blockchain adoption decision tree

The whitepaper *Blockchain Beyond the Hype A Practical Framework for Business Leaders* (Mulligan, Scott, Warren, & Rangaswami, 2018) underscores the importance of matching a technology such as Blockchain to the desired benefits, rather than thinking of the technology as an end in itself. In the case of population data management, there are clear benefits for the adoption of blockchain technology for data consolidation and sharing.

The World Economic Forum Strategic Intelligence identifies Identity and Persona management as one of the areas that can be highly impacted by the adoption of blockchain technology. Specifically, two technologies: Self-sovereign identity (SSI) and utilizing shared identity management infrastructure are seen to have the best potential for blockchain application. This is poised to address goal 16 of the United Nations Sustainable Development Goals, which has a specific target of providing a legal identity for everyone in the world by 2030; including birth

registration. Key challenges, however, include both the scalability of the blockchain and the need to maintain individual privacy while using the technology (World Economic Forum Strategic Intelligence, 2018).

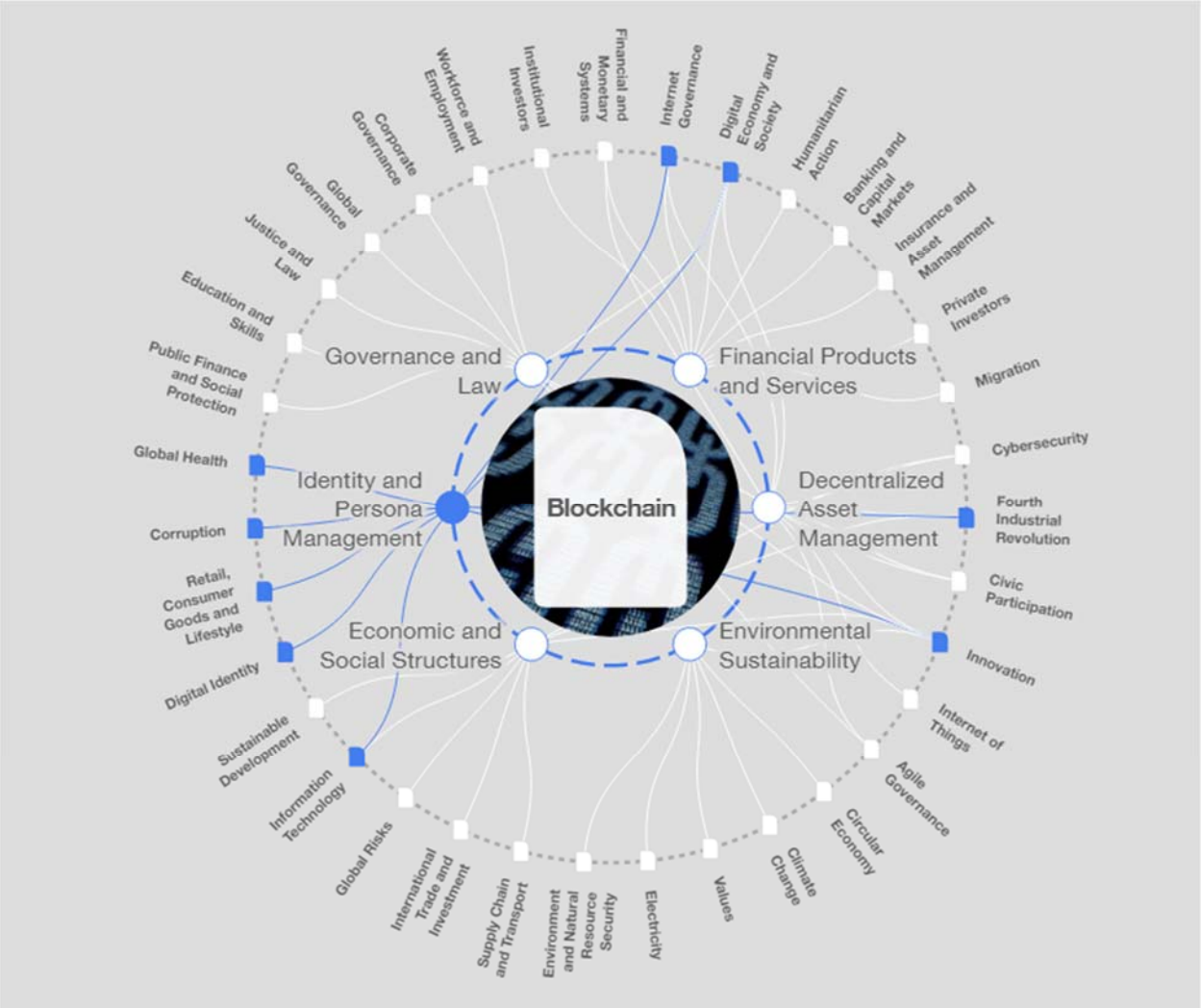


Figure 2. 7: Blockchain High Impact Areas

2.5.3 Review of Blockchain Technology Algorithms, Implementations, and Applications

Blockchain is a general-purpose technology: although it first gained prominence in Bitcoin cryptocurrency, the technology itself has the potential of being applied to virtually all areas of digital transaction processing. Some of the main application areas for the Blockchain include supply chain, real estate, healthcare, insurance, e-voting, Internet-of-Things (IoT), retail, public services, and digital property (Karafiloski, 2017). According to Karafiloski (2017), apart from cryptocurrencies, the adaption of Blockchain in real-world cases will arise where many institutions

need to share a database, but none of them should control or wholly own it and that adoption of private blockchains to businesses and governments may be the right way to start building the Blockchain world. This review identifies some notable application areas:

Figure 2. 8: Blockchain application areas

2.5.3.1 Blockchain Technology in Identity Management

The United Nations Sustainable Development Goals in target 16.9 addresses the issue of the provision of legal identity to everyone. The World Identity Network (WIN), through the "Blockchain for Humanity" challenge (UN-OICT, 2017) addressed this SDG by implementing a system based on blockchain technology to combat child trafficking in Moldova.

The report *Turning Invisible Children into Invincible Ones* (World Identity Network, 2018) discusses the advantages of blockchain in Identity Management. These advantages include general reduction of costs by automation and cutting out intermediaries who may be costly; automated and transparent processing; immutability making data tamper-resistant and, increased resilience through distributing transaction processing and record-keeping.

The report also notes the importance of blockchain technology in facilitating the adoption of self-sovereign identities: A self-sovereign identity (SSI) allows a person interacting with the digital

world, to control the sharing of their identification data by determining how much of it is shared, when it is shared, and with whom it is shared. The Veres One Project, which is a classic case of SSI, is not only working to provide SSI for individuals but also organizations by enabling people and organizations to control their identifiers and their identity data (Veres One, 2018).

The illustrations in figure 2.9 show the blockchain implementations presented by the winning contender in the “Blockchain for Humanity" challenge.

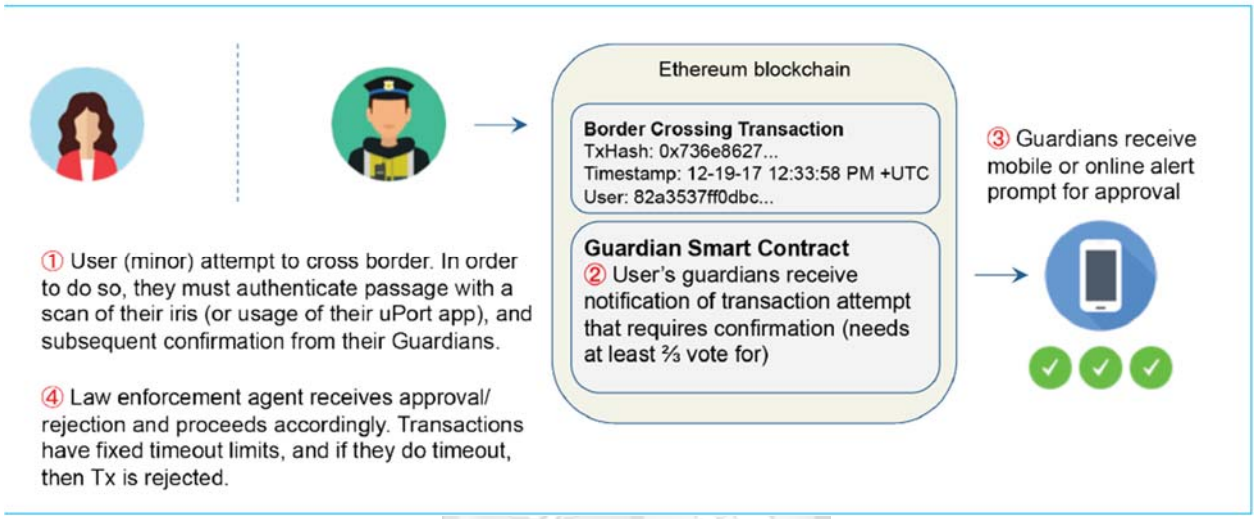


Figure 2. 9: ConsenSys Solution

Another notable project that implements blockchain technology is the World Food Program’s Blockchain for Zero Hunger (WFP, 2016). The World Food Program partnered with select stores as participants in the Building Blocks blockchain network which is is an accounting system that monitors and keeps records of cash-based relief to refugees so that they can buy food in participating stores. Before the purchase is authorized the beneficiary undergoes identity verification through iris scans, after which their account is checked for availability funds. Since all these transactions happen on the blockchain accounts can be easily and immediately consolidated (Blockchain4aid, 2018).

2.5.3.1.1 Personal Data Management System

The majority of personal data management systems proposed are focused on giving control of personal data to the person rather than third-parties. The model of these systems uses the blockchain as an access-control manager rather than storage- which is handled off-blockchain.

This has the advantage of autonomy such that it does not require trust in a third party. Therefore, personal data management systems on blockchain such as the one proposed by Zyskind, Nathan, & Pentland (2015) give the user data ownership, transparency, and auditability as well as granular access control to their personal information.

2.5.3.2 Blockchain Technology in Health Record Management

Patient records, which constitutes population data, is the medical history of a patient consisting of examinations and lab test results, diagnoses, and treatments, each of which can be handled by more than one physician or even more than one hospital, present an opportunity for adoption of blockchain technology especially for data consolidation and sharing among several parties. Several proposals have since been presented towards this end:

2.5.3.2.1 MedRec

MedRec is an Ethereum-based blockchain system for personal control of identity and distribution of personal information, implemented in the context of medical records (Nchinda et al., 2019). Whereas it addresses control of patient data by patients and sharing with third-parties, it has the disadvantage of requiring every node to verify and propagate all transactions and blocks. It also runs on a public Ethereum blockchain network causing security concerns because data can be publicly read outside the consortium of participating providers.

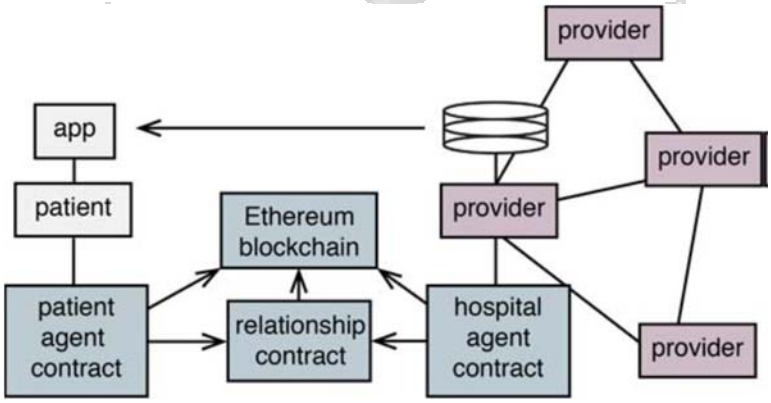


Figure 2. 10: The MedRec Network

2.5.3.2.2 FHIRChain

FHIRChain is a blockchain-based network that facilitates the safe and scalable sharing of clinical data to support collaborative decision making (Zhang et al., 2018). FHIRChain uses the public key infrastructure (PKI) to enable the sharing of data between clinicians and the blockchain is used to

maintain the public/private key pairs. While this PKI approach works for the intended use case, the fact that it is the clinicians controlling the sharing of patient records may be seen to contravene the protection of personal data. However, from the perspective of population data, the user (patient) has no control over what to share or not to share since it is a requirement by law. A key strength in the implementation is keeping the sensitive health records data off-the-chain and exchanging reference hashes on-the-chain thus avoiding storing encrypted health data in the blockchain. Such an implementation however requires maintaining two sets of chains which is not very efficient.

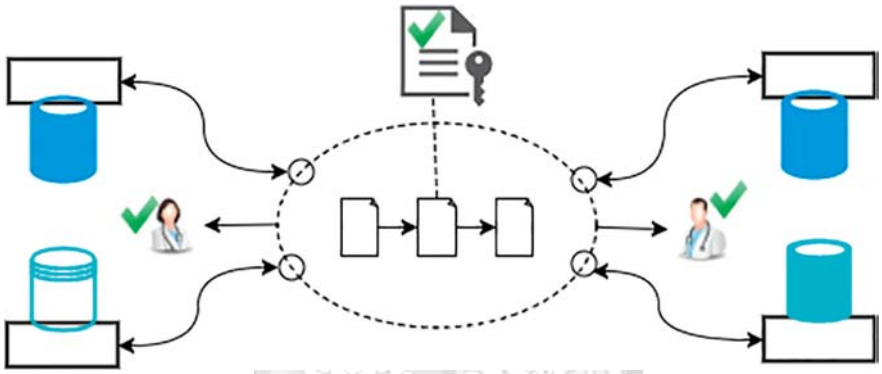


Figure 2. 11: Architectural Components in FHIRChain

2.5.3.2.3 E-health data access management with privacy protection

This blockchain solution uses two chains in the management of medical records: the mainchain which contains non-identifying information and is accessible by any third parties and the sidechain which is accessible only by the trusted nodes (Hirtan, Krawiec, Dobre, & Batalla, 2019). Communication between trusted nodes uses the Public Key Infrastructure (PKI) to authenticate thus ensuring the permissioned nature of the network.

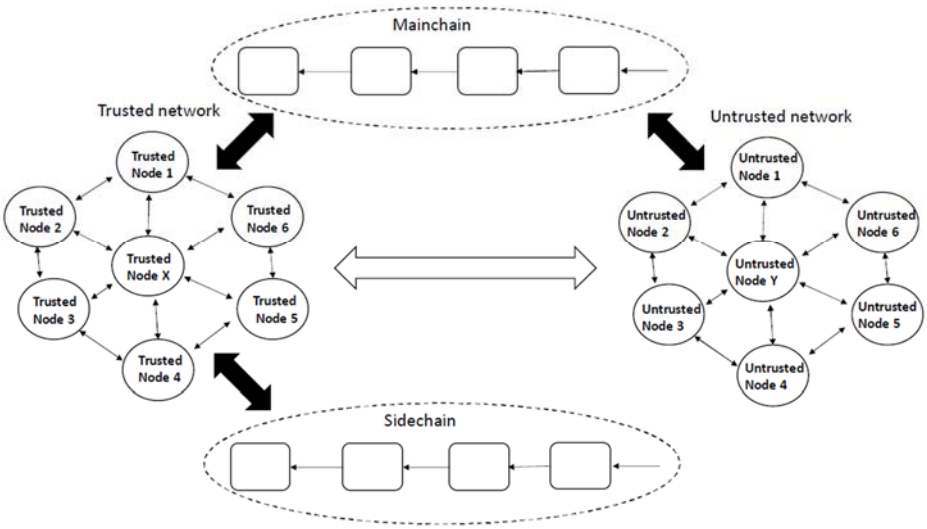


Figure 2. 12: Logical concept of the system design

2.5.3.2 Blockchain Technology in Supply Chain Management

Supply Chain Management refers to the management of the movement of goods, services, and information involving the acquisition, storage, movement of raw materials, infrastructure as well as full-fledged finished goods from source to destination. Blockchain technology has been applied to supply chain management especially in tracking products through the chain and by record keeping. Data sharing among the participants enables visibility and transparency of transactions, fostering trust, and ensuring faster transactions and reduced settlement periods. The IBM Sterling Supply Chain Suite (IBM, 2015) is an example of the implementation of Blockchain technology to supply chain management.

2.5.3.3 Blockchain Technology in Financial Services

Apart from cryptocurrencies such as Bitcoin and Ethereum, conventional financial institutions are yet to embrace blockchain technology yet they continue to face challenges including security, identity theft, and inefficiency in transaction costs (DeMuro, 2018). While blockchain technology does mitigate these challenges, it is yet to achieve faster transaction settling speeds which seems to be the only issue preventing the adoption of blockchain technology in the mainstream financial industry.

2.5.3.4 Blockchain Technology for Kenya

The report *Emerging Technologies for Kenya: Exploration and Analysis* (Distributed Ledgers Technology and Artificial Intelligence Task Force, 2019) identifies Distributed Ledger Technology as an emerging technology that could streamline many sectors. The report, in its recommendations, identifies the following areas where Blockchain technology may be impactful: public service delivery, agriculture, and food security, democracy, and elections (as shown in figure 2.13), eliminating corruption, improving health and drug safety, eliminating counterfeits, improving cybersecurity; among other areas.

Setting	The context	Remarks
The city of Moscow's Active Citizen program	In December 2017, the program started using a blockchain for voting and to make the voting results publicly auditable. Each question discussed by the community and put up for voting is moved to the e-voting system using a blockchain. After the voting is complete, the results are listed on a ledger containing all the previous polls.	The most popular polls were reported to have 137,000 to 220,000 participants.[10] In one such case on the Ethereum platform, citizens indicated their preferences for temporary relocation if the building in which they were living would be demolished and replaced by a better building. The platform reached a peak of approximately 1,000 transactions per minute. It's not clear whether the platform can handle the volume if a higher proportion of Moscow's 12 million citizens participate in the voting.
The South Korean province of Gyeonggi-do's community projects	The province used a blockchain-based voting system to gather votes on community projects. 9,000 residents voted.	The Korean financial-technology startup Block developed the blockchain platform.
The annual general meeting of the Estonian tech company LVH Group	Shareholders can log in using their verified national online ID and vote at the meeting.	The voting system issues voting-right assets and voting-token assets to shareholders. A user can spend voting tokens to vote on meeting agenda items if that user owns the related voting-right asset. Nasdaq designed the system.
Sierra Leone's March 2018 general elections	Swiss startup Agora carried out tallying in two districts. After the voting, a team of accredited observers from different locations manually entered approximately 400,000 ballots into Agora's blockchain system.	This test was considered a partial deployment of a blockchain.[11] The elections were only verified by blockchain, not blockchain powered. Agora provided an independent vote count, which was compared with the main tally.

Figure 2. 13: Blockchain-based Voting

2.6 Conceptual Framework

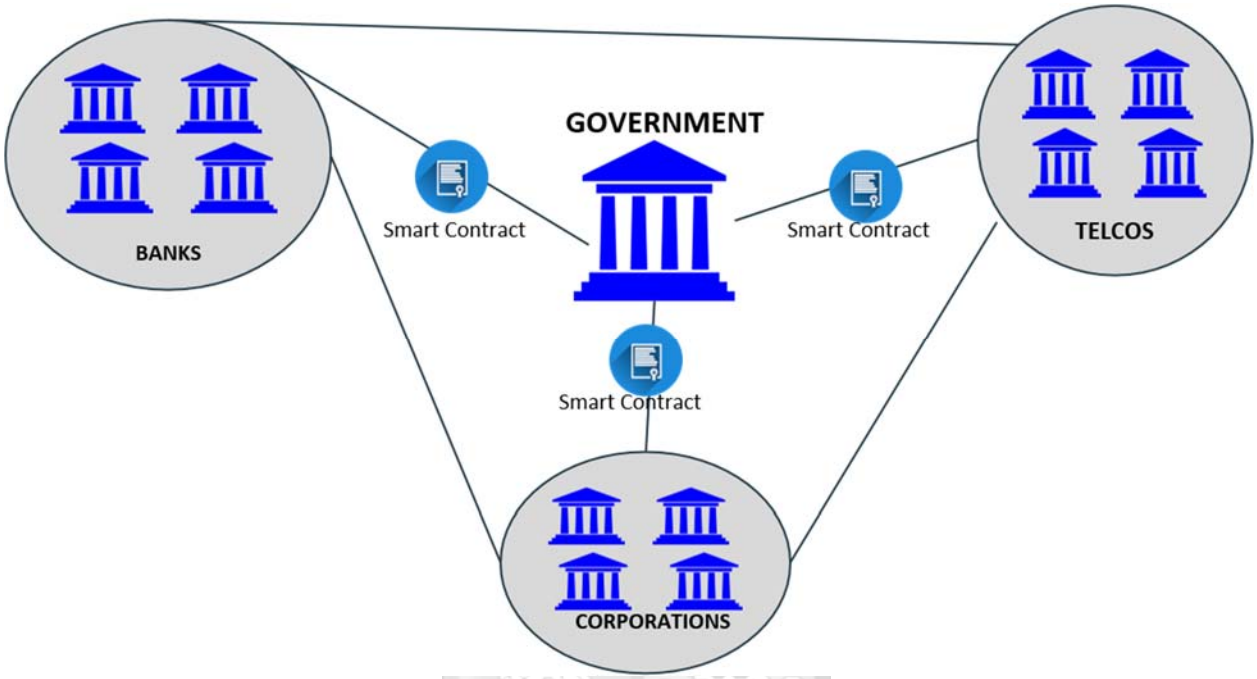


Figure 2. 14: High-level Conceptual Framework

Population data is controlled by a government in its jurisdiction, therefore, the partially centralized and partially decentralized approach of the permissioned blockchain is used. Permissions for fine-grained operations are maintained on the transaction level using the current Transaction Processing Systems (TPS). Deployment is done through consortium blockchain which is controlled by pre-authorized "Government" nodes. The rights to read the blockchain is assigned to the different participants. Participants can optionally create channels specific to their industries or relationships. Channels are private sub-networks of communication among a select group of blockchain network members enabling them to carry out private and confidential transactions. Channels, therefore, are the basis of the blockchain network since it is a “members club” of trusted blockchain peers.

Chapter 3: Research Methodology

3.1 Introduction

Research is the empirical and systematic search for relevant information related to a given subject (Kothari, 2009). Since no two types of research are similar, the researcher needs to tailor their research to address their specific problem. This approach the researcher adopts and the logic behind it is the methodology of the research. In this chapter, the research design is presented with guidance from the objectives of the research. The experimental nature of this research informs the choice of generating primary data which is specifically tailored to simulate personal information. The system development methodology is discussed and for the implementation and as the proof of concept, prototyping is presented as the system design approach.

3.2 Research Design

Research design is the structure that the research process follows from data collection to data analysis in obtaining answers to the research questions. Research design aims to match relevance to the purpose of research with the economy in the procedure. (Kothari, 2009).

Based on the shortcomings in the current system of management of population data, this research hypothesizes that the use of permissioned blockchain technology is better suited in the management of population data. This research is therefore designed and classified as experimental research. The experimental approach to research involves a controlled research environment where variables are manipulated to cause an effect on other variables. This effect is observed and recorded. The experimental approach is specifically fitting for this research because population data management is being replicated onto the blockchain platform, from the current relational database management model.

The blockchain network will be created on the Hyperledger Fabric platform. Two organizations representing the government (custodian of population data) and the registrar (Government's proxy for population data collection) are added to the network. Three user accounts, which contain user personal information, are generated by the registrar and committed to the blockchain as the "assets" to be tracked and shared. Two more organizations representing a third party population data consumers (a bank and a hospital) are invited into the network as participants. This "invitation" shows the permissioned nature of the blockchain. The permission to read/write to the

blockchain is granted to the third parties and smart contracts governing data exchange conditions defined. The registrar, the bank, and the hospital will be simulated by Laravel applications complete with their Relational Database management systems for their daily transaction processing. At predefined intervals, the third parties send a request to commit their transactions (in this case bank transactions and user hospital records) related to the three user accounts to the blockchain. In this way, the population data consolidation aspect is achieved. The sharing aspect is in the invitation of new participants into the network and providing read/write access rights and definition of smart contract conditions.

This controlled experimental environment ensures the achievement of the objective. However, as Kothari (2009) notes, the conclusions drawn based on experimental data are generally criticized for either faulty assumptions, poorly designed experiments, badly executed experiments, or faulty interpretations.

3.2.1 Target Population and Sampling

Population data is mandatory information collected by governments. Every documented citizen or is eligible to form the target population for this research. This research is not interested in the characteristics of the population data itself but rather how it is managed as a whole. Specifically, the study is keen on tracking the transition of a Kenyan citizen's personal statutory information from birth (when a birth certificate is issued) to death (when a death certificate is issued); including all life events in between where statutory information about the individual is collected. This information provides the primary data needed to simulate the use of blockchain technology in population data consolidation and sharing.

This, therefore, means that any subset of population data is usable. Therefore, The sampling method chosen is random sampling. Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population.s

3.2.2 Data Collection

This research will make use of primary data. Considering the huge volume of personal data required for the prototype, and since population data is sensitive information and access may be restricted, the primary data will be generated using a programmable tool called Faker (Fzaninotto,

2020), which can automate the process of generating relevant fields of any volume of data required. For example, the registrar participant will be tuned to generate personal information; the bank third-party will generate financial transaction data such as deposits and withdrawal; and the hospital third-party will generate health-related records such as appointments, diagnosis, and prescription information.

3.2.3 Data Analysis

Since this research is an experimental study with a focus on creating a prototype, controlled data generation is done using an automated tool called Faker (Fzaninotto, 2020). The focus is therefore not on the data but on the ability of the prototype to handle data in simulating consolidation and sharing of this data through blockchain technology.

3.2.4 Research Quality

Research quality is evaluated through the concepts of validity and reliability. The methods, techniques, and datasets used in achieving the research goals and how well these items fit into the research determine the research quality. These methods, techniques, or datasets are therefore evaluated through studying their reliability and validity. While validity is about the accuracy of a measure, reliability is about the consistency of a measure (Middleton, 2020). Population data is auto-generated in this study to be used in simulating the prototype. This makes the outcome reproducible should the same study be conducted under a similar environment with or without the real population data and therefore ensures consistency and therefore data reliability. On the other hand, data validity is guaranteed since the data is autogenerated and can be controlled to obtain only desired fields to match the population data fields required.

3.2.5 Dissemination of research findings

Dissemination refers to how research findings are communicated to the target audience in ways that enable uptake of the research in decision-making and eventually in practice (Wilson, Petticrew, & Calnan, 2010). This study proposes a secure, efficient, cost-and-time-effective way for the consolidation and sharing of population data through the use of blockchain technology. This is especially of interest to the Government of Kenya, which is the target audience for this research because it has a mandate in accounting for every Kenyan citizen through the collection and maintenance of personal information for its citizenry.

3.4 System Development Methodology

In agile software development methodology, the software is built incrementally over a specified time, and delivery is in bits instead of all at once. (Agile Alliance, 2019). It was designed to accommodate the rapidly changing environment and requirements thus enabling faster outcomes by improving upon previous models until the final product is achieved. The specific agile methodology to be used in the prototype development is Rapid Application Development. Rapid Application Development has the advantage of fast project turnaround times. It is therefore suited for the fast-paced environment of software development where requirements can be moving targets. The focus in RAD is on prototype development rather than planning to make outcomes visible quickly. With outcomes visible rapidly it becomes easy to track progress and communicate and adopt change in near real-time. As a result, there is effective communication, greater efficiency, and effective and faster development. (LucidChart, 2018).

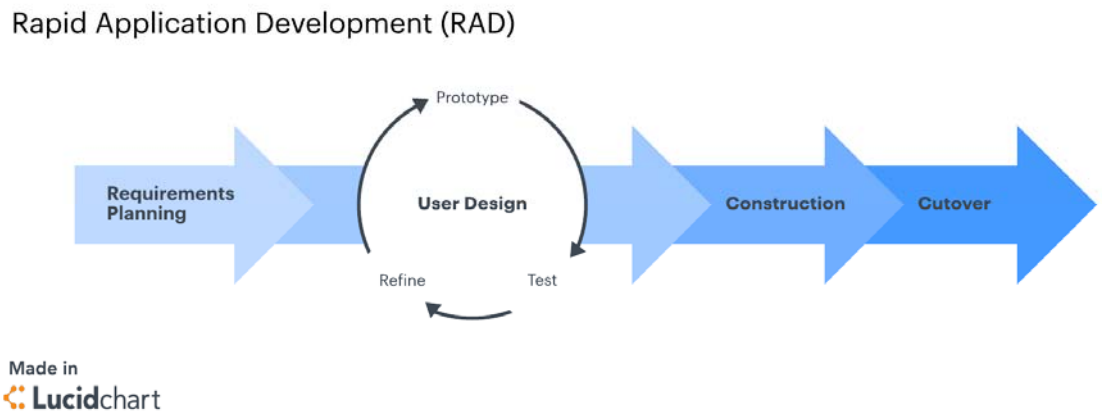


Figure 3. 1: The RAD Methodology

The blockchain prototype to be developed would require interactions with various existing systems to simulate data sharing. These legacy system interactions lay the ground for a rapidly changing environment considering the amounts of integration required. The incremental development approach presented by agile software development will be able to accommodate these changing requirements without greatly impacting the timelines. In agile development, requirements and design are done together, and execution is as a series of increments. This gives agile development an edge over a conventional model where requirements and design are developed as a series of steps, one after another.

The system analysis will involve, analysis of the need to manage contractual relationships and the type of contracts that exist; analysis of the need to track transactions that involve more than two parties; analysis of the need for immutability, transparency, auditability, and accountability in record-keeping; and, analysis of errors due to manual processes including fragmentation and duplication of effort and data.

3.5 System Implementation

The blockchain implementation uses the Hyperledger Fabric which is a blockchain framework implementation provided by Hyperledger. Hyperledger is an open-source community focusing on advancing cross-industry blockchain technologies. It is hosted by The Linux Foundation and includes support from leaders in various fields such as supply chain, manufacturing, finance, banking, Internet-of-Things, and technology (The Linux Foundation, 2009).

Hyperledger Fabric is an open-source enterprise-grade blockchain framework that introduces a modular architectural pattern in the development of block-chain based applications. The modular architecture is achieved through the use of container technology in the development of pluggable and interchangeable services. The modular architecture enables hyperledger fabric to support a wide variety of blockchain use cases with different requirements. Some notable attributes for the hyperledger fabric framework include support for permissioned shared ledgers; support for verified identities through the use of the Membership Service providers and confidential transactions through the use of channels. It also has better performance since it cuts out the proof-

of-work method as a consensus algorithm and instead uses Byzantine Fault Tolerance (BFT). The use of container technology also enables ease of scaling for hyperledger fabric.

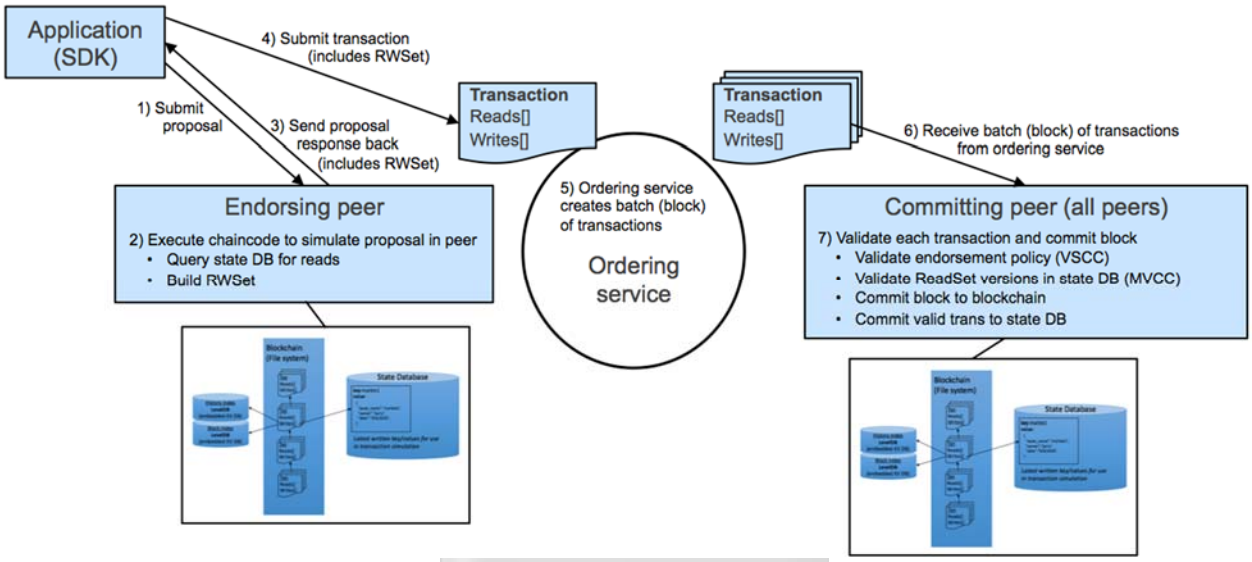


Figure 3. 2: Hyperledger Fabric Transaction flow

The Transactional Processing Systems, which will represent 3rd party partners to the whole systems, will be simulated using multiple Laravel application with an accompanying MySQL database running on docker containers.

3.6 Ethical Considerations

This study proposes a secure, efficient, cost-and-time-effective way for the consolidation and sharing of population data through the use of blockchain technology. This is especially of interest to the Government of Kenya, which is the target audience for this research because it has a mandate in accounting for every Kenyan citizen through the collection and maintenance of personal information for its citizenry. Since this research will be handling personal information, data security, protection, and privacy concerns should be considered. However, for the prototype and the proof of concept, dummy data will be auto-generated and used.

Chapter 4: System Design and Architecture

4.1 Introduction

The purpose of system design and architecture is to identify the different components of a system, and how to optimally combine them to achieve the required outcomes. Whereas the system architecture can be thought of as the superstructure of the system, system design is the plan of how to accomplish the requirements of the system. The Unified Modelling Language (UML) is the general-purpose, developmental, modeling language for visualizing software engineering designs and architectures.

This chapter details the system design and architecture of the prototype. The conceptual framework introduced in chapter 2.6 is broken down into its constituent parts to show how all the components interact to accomplish the goal of managing population data consolidation and sharing using blockchain technology. The system requirements are analyzed to inform the design. The system design diagrams are then modeled using the Unified Modelling Language (UML) - the general-purpose, developmental, modeling language for visualizing software engineering designs. The system design diagrams included are the use-case diagram, the system sequence diagram, and the data flow diagram.

4.2 Requirements Analysis

The aim of this research, as defined in chapter 1.3, is to design a prototype of a population data management system based on the permissioned blockchain technology to explore the validity and viability of blockchain technology in consolidating, securing, auditing, and sharing of population data. Based on this aim, the functional and non-functional requirements were deduced.

4.2.1 Functional Requirements

i. The system should allow participants to be added to the blockchain network.
ii. The system's access control lists should control access by blockchain network participants.
iii. The blockchain network participants should be able to write data to the blockchain.
iv. The blockchain network participants should be able to read data to the blockchain.

4.2.2 Non-functional Requirements

4.2.2.1 Security Requirements

i. Participants should be correctly authenticated to access the system.

ii. Access for reads and writes should be restricted to valid participants.

4.2.2.2 Reliability Requirements

The system should consistently perform correctly by outputting the correct results given the correct input.

4.2.2.3 Usability Requirements

The inner workings of the system should be transparent to the user.

4.3 System Architecture

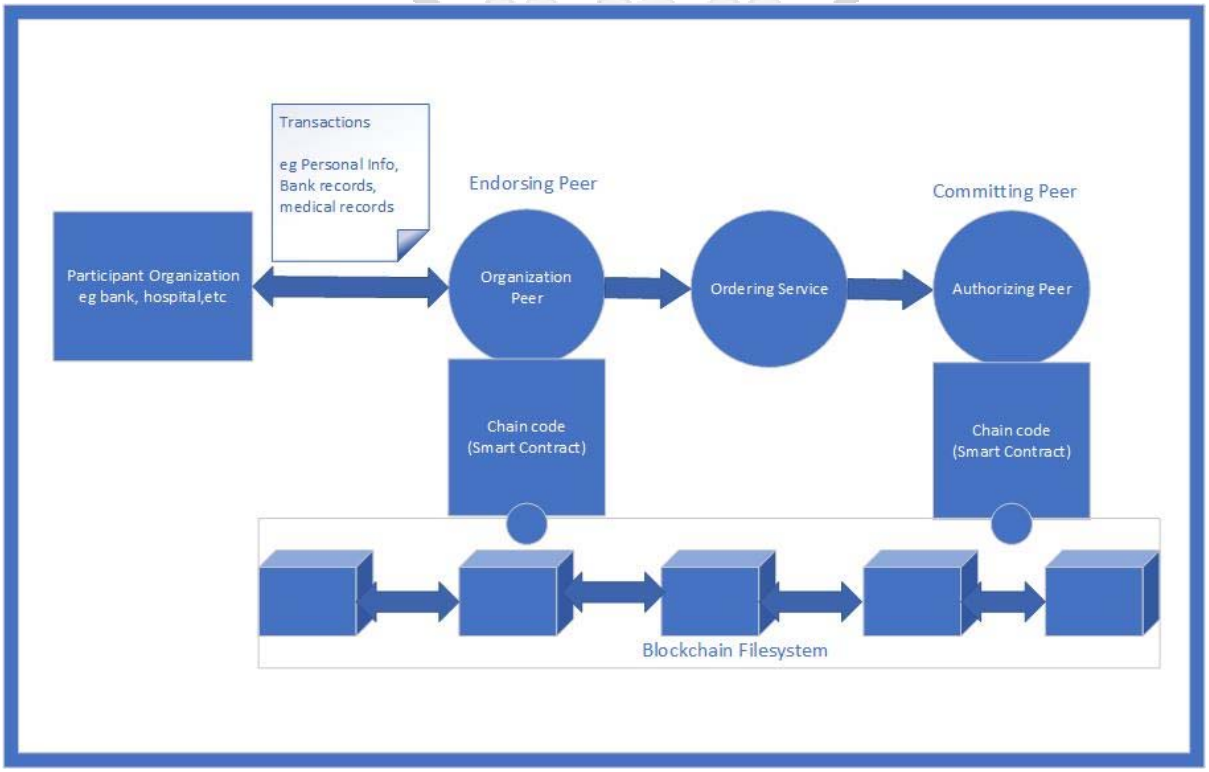


Figure 4. 1: General System Architecture

Figure 4.1 shows the general system architecture. The participant organizations - which represent any public data consumer/generator and specifically their transactional processing systems eg banks, hospitals, etc – are invited into the network channel by the membership service provider (MSP). Invitation involves the assignment of the digital signatures through public key infrastructure.

Once in the channel, the organization will submit a transaction proposal to its peer node, where a smart contract is run, querying the state database for reads and then creates the read-write set for the transaction. The transaction is then submitted to the ordering service where a block with ordered transactions is created from this and many other transactions from other peers. The ordering service then submits this block to the authorizing peer, which validates all transactions and commits the block to the chain. This in effect makes available (distributes) the new data to all connected peers who can read it as required.

4.4 Use Case Diagram

Figure 4.2 shows the use case diagram with four actors. The TPS represents whatever Transaction Processing system a participant may be using. The peer is the interface between the participant's organization and the blockchain filesystem. The administrator represents the MSP which may also be the authoritative peer. For illustration purposes, table 4.1 shows the detailed example of the Commit Transaction use case.

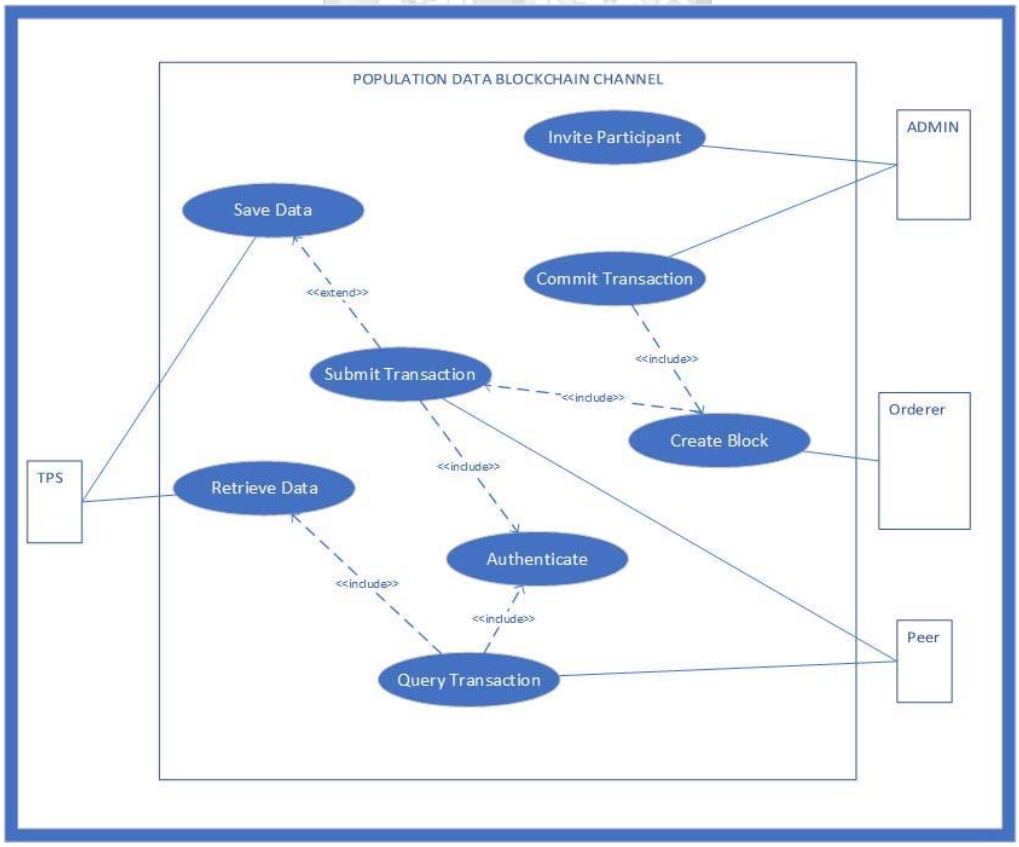


Figure 4. 2: The Use Case Diagram

Table 4. 1:Commit Transaction Use case

Use Case	Commit Transaction
Primary Actors	Transaction Processing System Peer Node Orderer Administrator
Preconditions	Peer is authenticated Pending Save Data request
Postcondition	Transaction is valid
Success scenario	1.TPS validates format and sends a request to Save Data 2.Peer executes the smart-contract to validate the transaction and builds a Read-Write Set 3.Peer passes the valid transaction to orderer for ordering and block creation 4.Orderer submits the block to admin for validation and authorization 5.Block is committed
Exception	The transaction fails smart-contract check

4.5 Data Flow Diagram

The data flow diagram (DFD) graphically presents the flow of data within a system. Figure 4.3 illustration shows the decomposed data flow diagram at level 2.

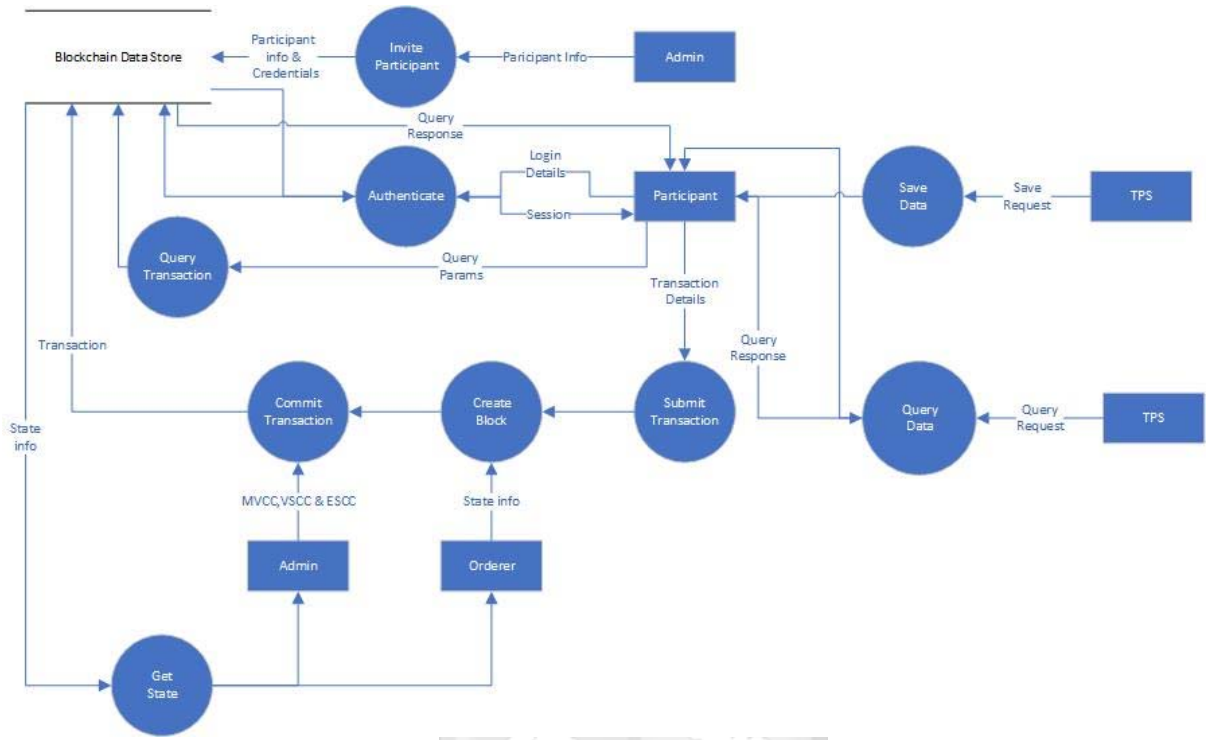


Figure 4. 3: System Data Flow Diagram

The admin, who is also a membership service provider (MSP), collects the participant information and invites them into the blockchain channel, creating digital certificate credentials for the participant to access the system. The participant, acting as the endorsing peer, then proceeds to authenticate whenever there is a request from the TPS. Query transaction is invoked and the transaction response is sent to the TPS whenever it requests data. Save data request from the TPS is validated by the participant. The participant then submits the transaction to the orderer to create a block. The orderer, using state information from the blockchain data store, creates the block and passes it on to the committing peer. Finally, the admin, in this case acting as the committing peer validates the block using statute information from the store and commits it to the blockchain data store.

4.6 System Sequence Diagram

The system supports interactions between itself and external entities through the use of events and feedbacks. Sequence diagrams are a type of UML diagrams that show how objects within a system interact with each other particularly by showing interactions and the order in which they occur, in other words, they show the sequence of events. Figure 4.4 shows this sequence of events.

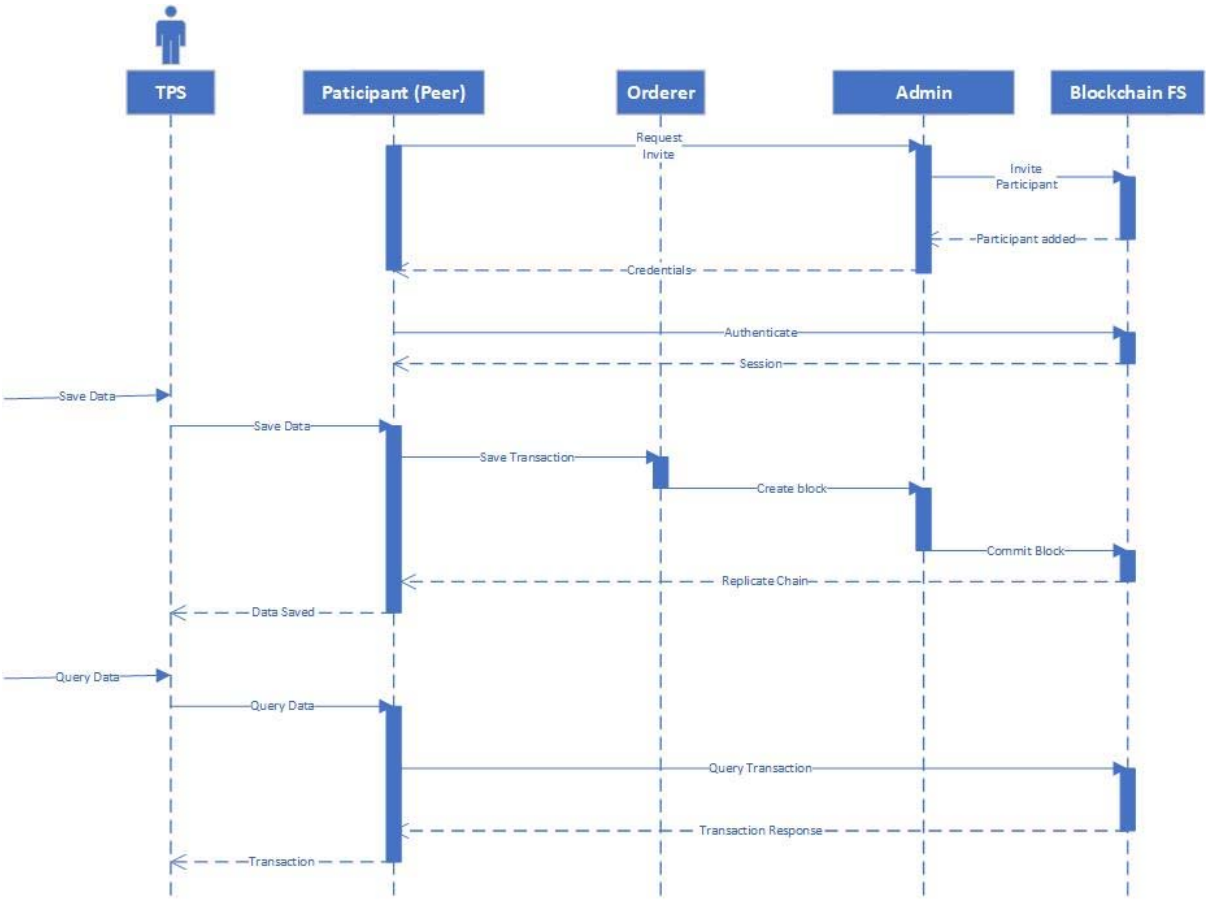


Figure 4. 4: System Sequence Diagram

4.7 Class Diagram

A class diagram is a UML representation of all classes in an object-oriented system, showing their attributes, methods, connections, interactions, and inheritances. For population data, the basic unit is the individual/person, and every other record relating to the person are included as attributes of the person. The class diagram in figure 4.5 models a sample person object. Birth and education objects are the only ones included for brevity, however many other objects describing the person such as financial, medical, tax, employment records etcetera can be included.

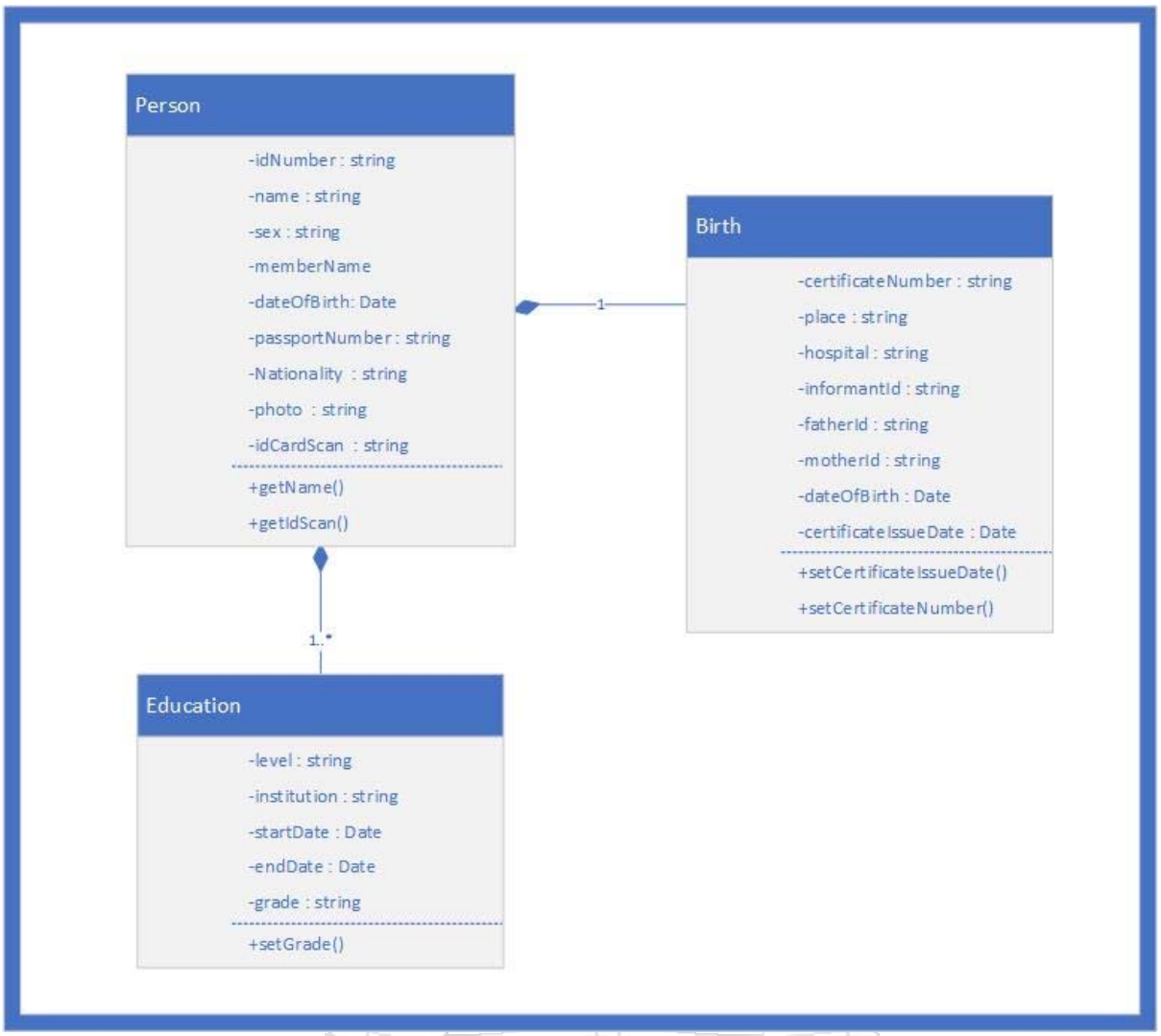


Figure 4. 5: Class Diagram

Chapter 5: Implementation and Testing

5.1 Introduction

Hyperledger Fabric is an open-source enterprise-grade blockchain framework that introduces a modular architectural pattern in the development of block-chain based applications. The permissioned and distributed nature of Hyperledger Fabric makes it suitable for the implementation of this prototype. It also includes a suite of other products such as the Hyperledger Composer which is an open-source development toolset and framework that simplifies the development of blockchain applications; and the Hyperledger Composer REST server which provides RESTful APIs to Decentralized Applications (DApps) for interaction with the blockchain network. Since system simplicity and transparency are desirous qualities of this prototype, the implementation seeks to effectively integrate existing population data management information systems with the blockchain without drastically affecting operations. This chapter, therefore, discusses the implementation and testing of the proposed prototype using these technologies.

5.2 Model Components

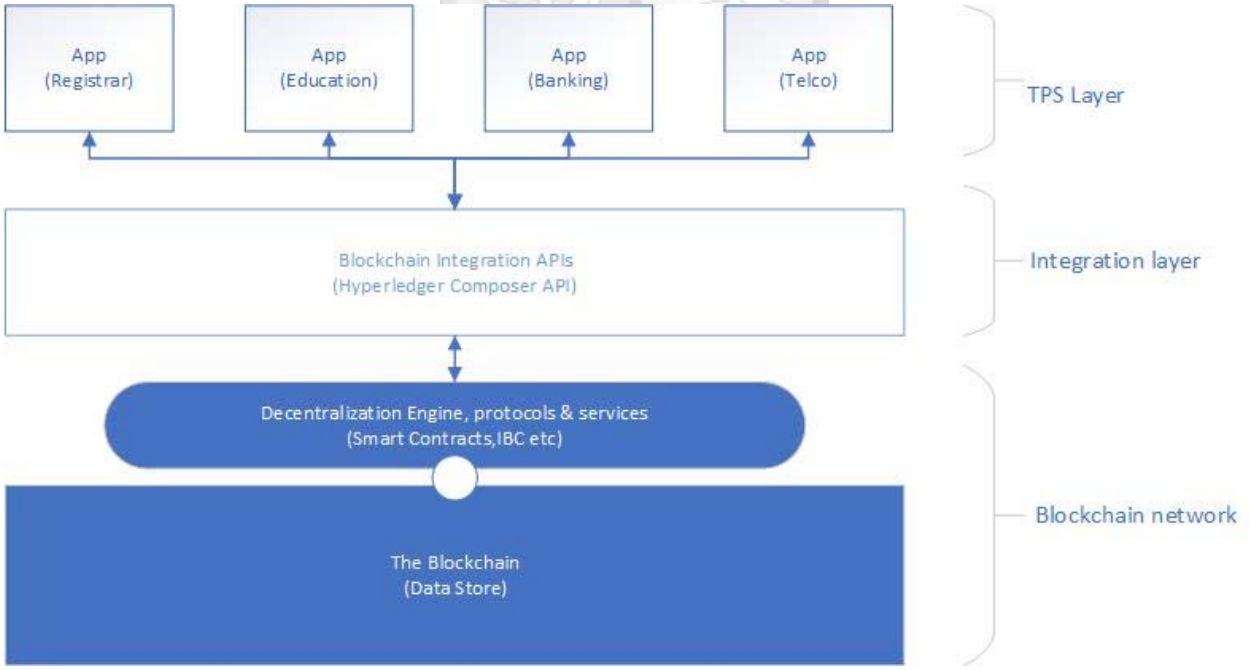


Figure 5. 1: Proposed Solution Design

5.2.1 The organization TPS (Peer/Participant)

Two separate organizations transactional processing systems were created, separated by models aptly named Education and Birth. These models represented the education institution and the registry of births, which will be the two participants in the blockchain network. The registry acted as the data producer, where a person's record is added to the blockchain and instantly becomes available to all the connected participants on the network, while the education model will be the consumer of this data for their internal processes such as school admissions.

5.2.2 The Blockchain business network

The blockchain business network used was the Hyperledger fabric through the Hyperledger composer toolset. Hyperledger Composer is an open-source development toolset and framework that simplifies the development of blockchain applications. In this business network, the asset is the population data, in this case, made up of individual personal records. The basic unit of the Asset is the person model which will store information based on data received from the TPS. Two participants were created on the network: the Registrar and the Educator. Transactions, which are the operations the participants invoke to manage the assets, were also defined. Roles were used to restricting what transaction a participant can access, essentially giving read-write access to the blockchain.

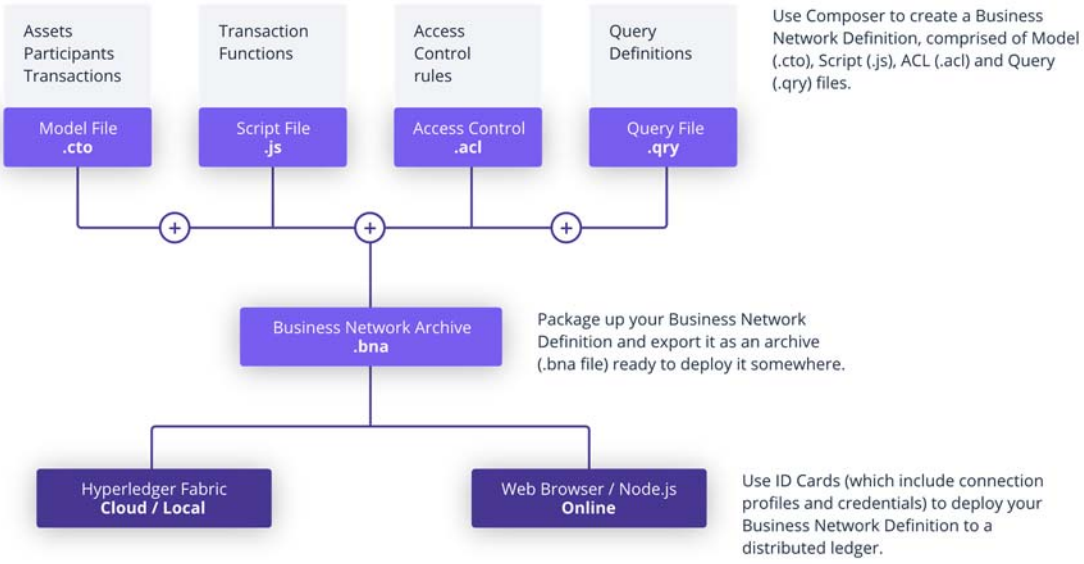


Figure 5. 2: Hyperledger Composer Components

5.2.3 The Integration API

The integration API used was the javascript Hyperledger Composer API which is made up of Client, Administrator, and Runtime javascript APIs. The administrator APIs are used to obtain a connection to the blockchain business network and perform administrative tasks. The Client APIs also obtain the connections, however, the functions are limited to performing business operations. Business operations include the creation of assets and invoking transactions in the network. The Runtime APIs allow access to the network for operations such as building and running queries, emit events, get registries and current participant information and post HTTP RESTful calls.

When an HTTP RESTful request is made from the Transaction Processing System to either save or query data, the appropriate client API is called, which interacts with the blockchain to add a new asset in the case of saving data or retrieve asset information in the case of querying asset data. Guzzle HTTP client API was used for making the API calls to run the APIs.

5.3 Model Implementations

5.3.1 Model TPS Implementation

The Transaction Processing Systems were implemented using the Laravel PHP framework. Two systems were created: the education institution TPS and the registry of birth TPS. The birth registry TPS is the custodian of data relating to birth and initial personal information of an individual. It is, therefore, the producer of initial personal information in the blockchain network. By committing the data to the blockchain network, it essentially makes it available for authorized participants in the blockchain network to access it.

The Education institution TPS is a consumer of the personal data from the blockchain network. It accesses the personal information of the individual from the blockchain and uses it internally for its operations. The Education TPS also commits some education-related information about the individual to the blockchain, therefore making it available to other authorized participants.

In essence, population data consolidation is achieved through saving data on a single blockchain, and sharing is achieved through reading from this single blockchain, and therefore the single

source of truth is maintained. The source code for this model TPS implementation, figuratively named the *Public Data Chain* (Omoka, 2020) can be found on GitHub.

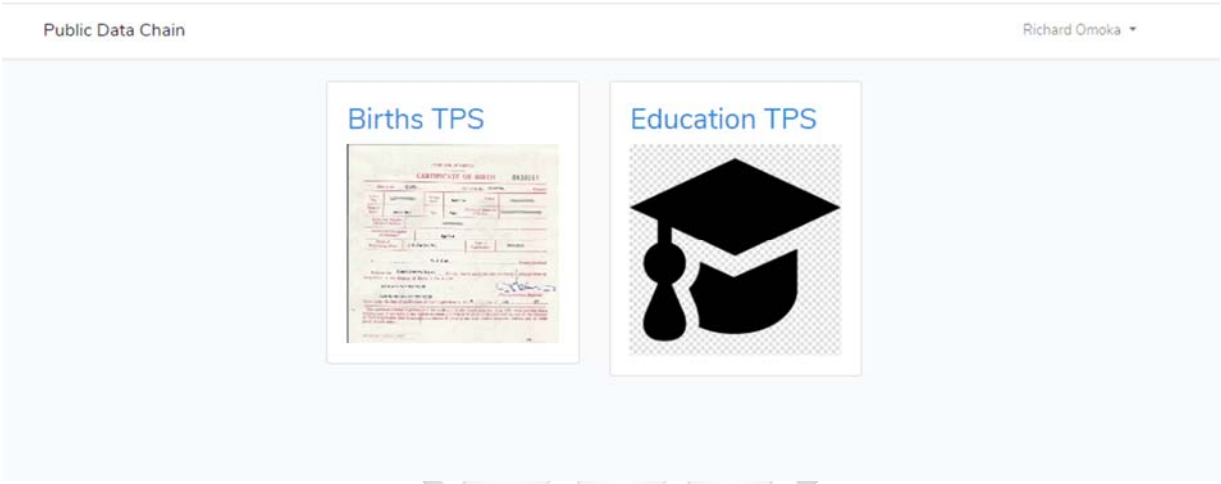
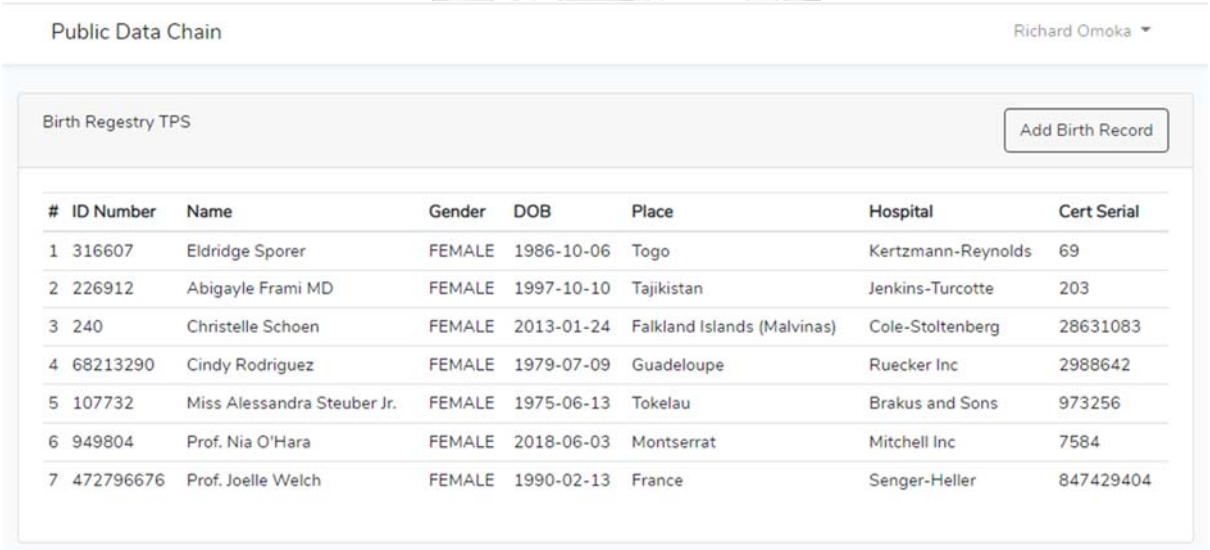


Figure 5. 3: Model TPS Implementation

5.3.2 Birth Registrar TPS Interface



#	ID Number	Name	Gender	DOB	Place	Hospital	Cert Serial
1	316607	Eldridge Sporer	FEMALE	1986-10-06	Togo	Kertzmann-Reynolds	69
2	226912	Abigayle Frami MD	FEMALE	1997-10-10	Tajikistan	Jenkins-Turcotte	203
3	240	Christelle Schoen	FEMALE	2013-01-24	Falkland Islands (Malvinas)	Cole-Stoltenberg	28631083
4	68213290	Cindy Rodriguez	FEMALE	1979-07-09	Guadeloupe	Ruecker Inc	2988642
5	107732	Miss Alessandra Steuber Jr.	FEMALE	1975-06-13	Tokelau	Brakus and Sons	973256
6	949804	Prof. Nia O'Hara	FEMALE	2018-06-03	Montserrat	Mitchell Inc	7584
7	472796676	Prof. Joelle Welch	FEMALE	1990-02-13	France	Senger-Heller	847429404

Figure 5. 4:Birth Registrar TPS Interface

Figure 5.3 shows the screenshots of the interface for the registrar of births. It has a button to add the birth record for an individual. For this prototype the operation for adding the involves two steps: first, the faker factory is instantiated using the faker Laravel package and used to generate

personal information as defined by the person object. This personal information is then stored in the local TPS database.

The code snippet shows the sample operation for the creation of a record of birth by the registrar of births. This operation happens on the registrar's transaction processing system.

```
Birth::create([ 'pid_number'=>$person->id_number,
        'certificate_number'=>$faker->randomNumber(),
        'place_of_birth'=>$faker->country,
        'hospital_of_birth'=>$faker->company,
        'father_id_number'=>$father->id_number,
        'mother_id_number'=>$mother->id_number,
        'informant_id_number'=>$informer->id_number,
        'registrar_id_number'=>$registar->id_number,
        'date_of_birth'=>$faker->date(),
        'date_of_issue'=>$faker->date()]);
    DB::commit();
$this->commitToBlockchain($person);
```

Immediately after the person record is commited to the registrars transaction processing system, the *commitToBlockchain()* method is invoked passing in the full person record as a parameter.

The *commitToBlockchain()* operation involves calling the public data chain, through the Hyperledger Composer RESTful API which invokes the *createPerson()* transaction on the chain. This effectively commits the person's record onto the blockchain and therefore making it available to all network participants.

5.3.3 The Education institution TPS Interface

This is the interface to manage education details for an individual. Apart from invoking the *addEducationRecord* operation and saving education records in its local database, it transparently invokes two other operations: *searchPersonDetails* and *commitToBlockchain.* The *searchPersonDetails* operation takes the ID number of the individual and queries the blockchain through the Hyperledger Composer RESTful API to find the details of the individual that were previously committed to the blockchain by the Registrar of Births. Once this detail is retrieved, the *commitToBlockchain* operation is then invoked, passing the education details that are associated with the person and committing this information onto the blockchain.

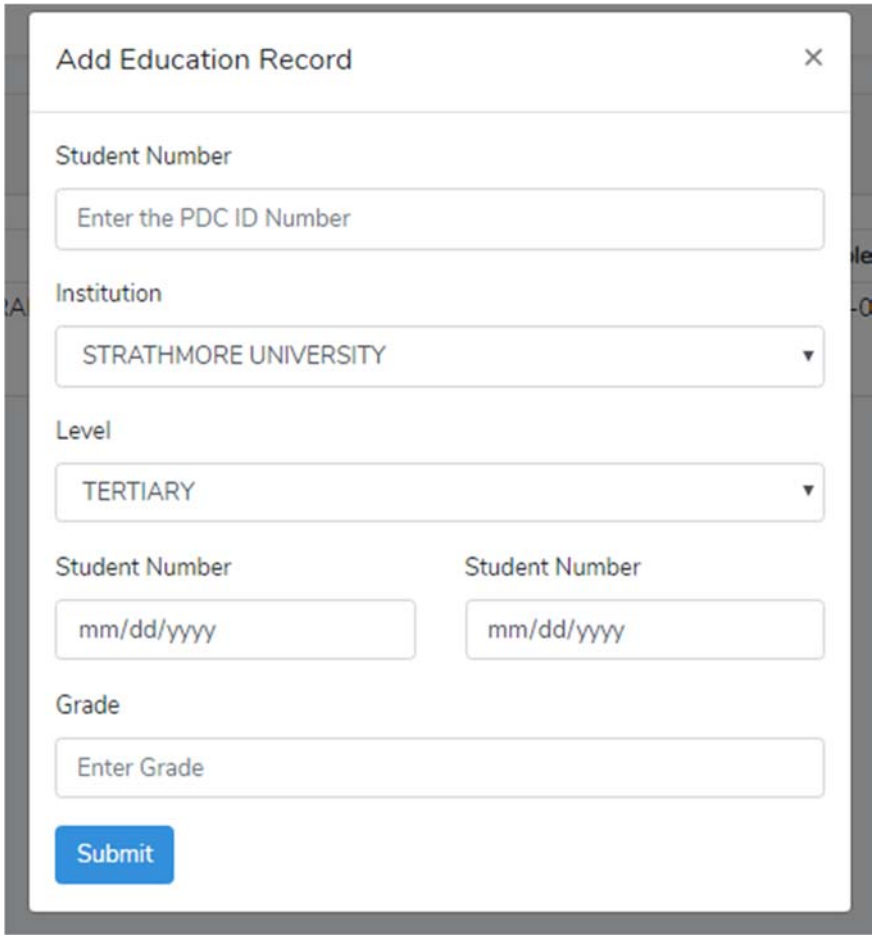


Figure 5. 5: Add Education Record Form

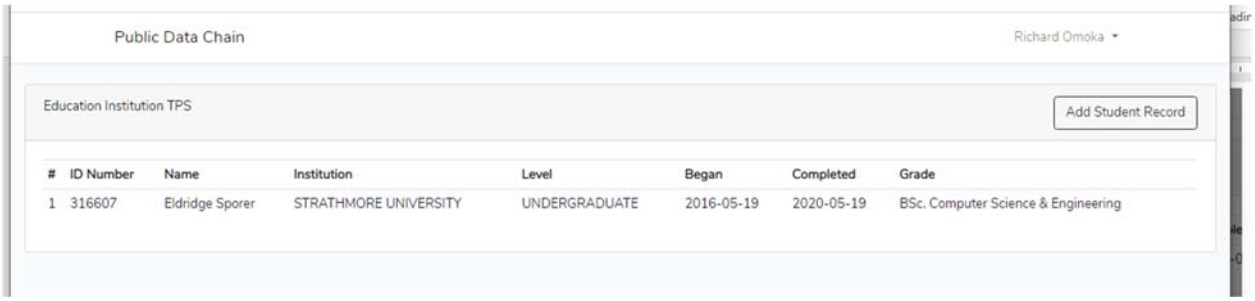


Figure 5. 6: Education TPS Interface

5.3.4 The Blockchain business network

The blockchain business network was implemented using a locally installed version of Hyperledger Fabric and Hyperledger Composer was used for interacting with this network. As explained in section 5.2.2, Assets, transactions, and participants were defined in the model file; the transaction functions were defined in the scripts file; the access control rules were defined in the ACL file and the data query functions were defined in the query file.

5.3.4.1 The Network Participants

In a blockchain business network, a participant is an actor and can be an individual an organization. Two participants were created for this prototype: The Educator and The registrar. The registrar is responsible for the creation of assets. The assets, in this case, refer to personal information maintained on the blockchain. The educator manages reads and writes education data onto the blockchain. Figure 5.8 shows the sample definition of the participants.

```
namespace net.pdc

abstract participant BusinessParticipant identified by registrarId {
    o String registrarId
    o String name
    o String email
}

participant Registrar extends BusinessParticipant {}

participant Educator extends BusinessParticipant {}
```

Figure 5. 7:Participant model definition

5.3.4.2 The Assets

In a blockchain network, assets are the items to be tracked and are stored in registries. In this model, the asset is the person/individual information. The person object is modeled such that it is extensible and can accommodate varied attributes about a person. Figure 5.9 shows the person asset model. The birth and eduction concepts are shown in the model indicate the modularity that

makes this prototype extensible such that should there need to track another attribute such as medical records for the individuals, it can simply be added into the person object.

```
concept Birth {
  o String certificateNo
  o String placeOfBirth
  o String hospitalOfBirth
  o String fatherIdNumber
  o String motherIdNumber
  o String informantIdNumber
  --> Registrar registrar
  o DateTime dob
  o DateTime doi
}

concept Education {
    o String level
    o String institution
    o DateTime startDate
    o DateTime endDate
    o String grade
    -->Educator educator
}

asset Person identified by idNumber {
    o String idNumber
    o String name
    o String sex
    o DateTime dob
    o String passportNo
    o String photo
    o String idScan
    o String address
    o Birth birthCertificate
    o Education[] educationRecords optional
}
```

Figure 5. 8: Asset Model Definition

5.3.4.3 The Transactions

Transactions are the operations through which participants manipulate with assets. The sample *createPerson* transaction is shown in Figure 5.10. The create person transaction is invoked by the registrar participant in the creation of an asset. The person asset includes the birth details.

```
/**
 * Create new person
 * @param {net.pdc.CreatePerson} person The new person asset
 * @transaction
 */
async function createPerson(tx){
   const factory = getFactory();
   const NAMESPACE = 'net.pdc';
   var newPerson = factory.newResource(NAMESPACE, 'Person', tx.newPerson.idNumber);
   newPerson.name = tx.newPerson.name;
   newPerson.sex = tx.newPerson.sex;
   newPerson.dob = tx.newPerson.dob;
   newPerson.passportNo = tx.newPerson.passportNo;
   newPerson.photo = tx.newPerson.photo;
   newPerson.idScan = tx.newPerson.idScan;
   newPerson.address = tx.newPerson.address;


   var birthCertificateConcept = factory.newConcept(NAMESPACE, 'Birth');
   birthCertificateConcept.certificateNo = tx.newPerson.birthCertificate.certificateNo;
   birthCertificateConcept.placeOfBirth = tx.newPerson.birthCertificate.placeOfBirth;
   birthCertificateConcept. hospitalOfBirth = tx.newPerson.birthCertificate.hospitalOfBirth;
   birthCertificateConcept.fatherIdNumber = tx.newPerson.birthCertificate.fatherIdNumber;
   birthCertificateConcept.motherIdNumber = tx.newPerson.birthCertificate.motherIdNumber;
   birthCertificateConcept.informantIdNumber = tx.newPerson.birthCertificate.informantIdNumber;
   birthCertificateConcept.registrar = tx.newPerson.birthCertificate.registrar;
   birthCertificateConcept.dob = tx.newPerson.birthCertificate.dob;
   birthCertificateConcept.doi = tx.newPerson.birthCertificate.doi;
   newPerson.birthCertificate = birthCertificateConcept;

   const personAssetRegestry = await getAssetRegistry('net.pdc.Person');
   await personAssetRegestry.add(newPerson);
}

```

s

Figure 5. 9: Transaction Function createPerson() implementation

5.3.4.4 Access Control Rules

Blockchain Business networks contain a set of access control rules. Access control rules define fine-grained control permissions that control what participants have access to; what assets in the business network and under what conditions. For example, the Educator participant may only be

allowed to write to the education node of the Person Asset but still be able to read the rest of the asset values.

```
rule Educator {
    description: "Educator transactcts for Education records"
    participant: "org.hyperledger.composer.system.Participant"
    operation: ALL
    resource: "org.hyperledger.composer.system.**"
    action: ALLOW
}
```

Figure 5. 10: Sample Access Control Definition for Educator Participant

5.3.5 The Hyperledger Composer RESTful API

This is the integration layer sitting between the Transaction processing System application and the blockchain business network. The RESTful API definitions were generated and exposed by the Hyperledger Composer REST server running locally on the development machine. The REST server users the localhost interface to convert and expose the business network model into API definitions. An authenticated client interacts with the blockchain by calling the exposed endpoints. Since each invoked transaction must be signed by a certificate, the REST server starts the identity certificate and signs all transactions with it. Depending on the nature of the API call and the transactions, the REST server implements the CREATE, READ, UPDATE and DELETE (CRUD) operations, manipulating the state of assets and participants.

5.3.6 The Hyperledger Composer Playground

The Hyperledger Composer Playground is a web interface of the Hypeledger Composer and it provides a user interface for the configuration, deployment, and testing of a business network. The business network for this prototype was configured and deployed using a locally installed version

of the Hyperledger Composer Playground. The business card public-data-chain was created and assigned to a user admin@public-data-chain who is responsible for managing the network.

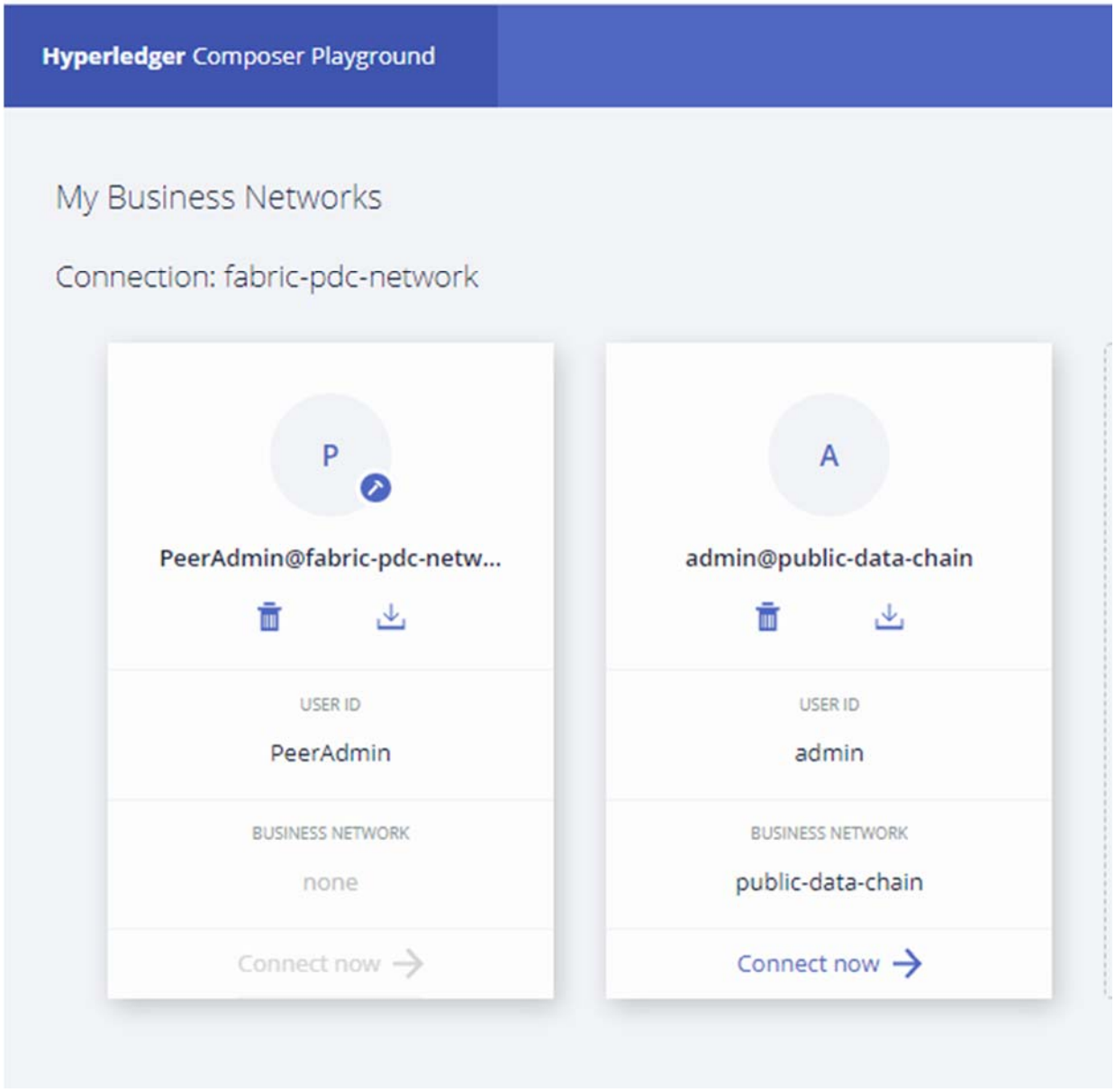


Figure 5. 11: Public Data Chain blockchain network as viewed on Playground

5.4 Software Flow

A web application simulating the internal systems for registrar of births and an educational institution was developed. The application simulated creation of personal records by various participants within their internal systems. Once the records are saved internally, the Hyperledger Composer RESTful API is called to persist the same data in the blockchain. Once the data is persisted on the blockchain, it automatically becomes available to all connected participants, who can then query based on personal record ID number. Participants can also persist in the blockchain personal data specific to them. All the access for querying and storage into the blockchain is controlled by the access control lists as shown in figure 5.11.

The registrar participant portal was used to create the personal record at birth. This record is persisted in the blockchain and therefore is available to all participants of the blockchain. The Educator participant can, therefore, query the personal details through the API using the person's ID number for use in their internal systems. The Educator can also persist education-related data in the blockchain thereby making it available to all the participants. All the interactions and Read/Write access to the blockchain is controlled by the access control list.

5.5 Model Architecture

Figure 5.6 illustrates the model architecture.

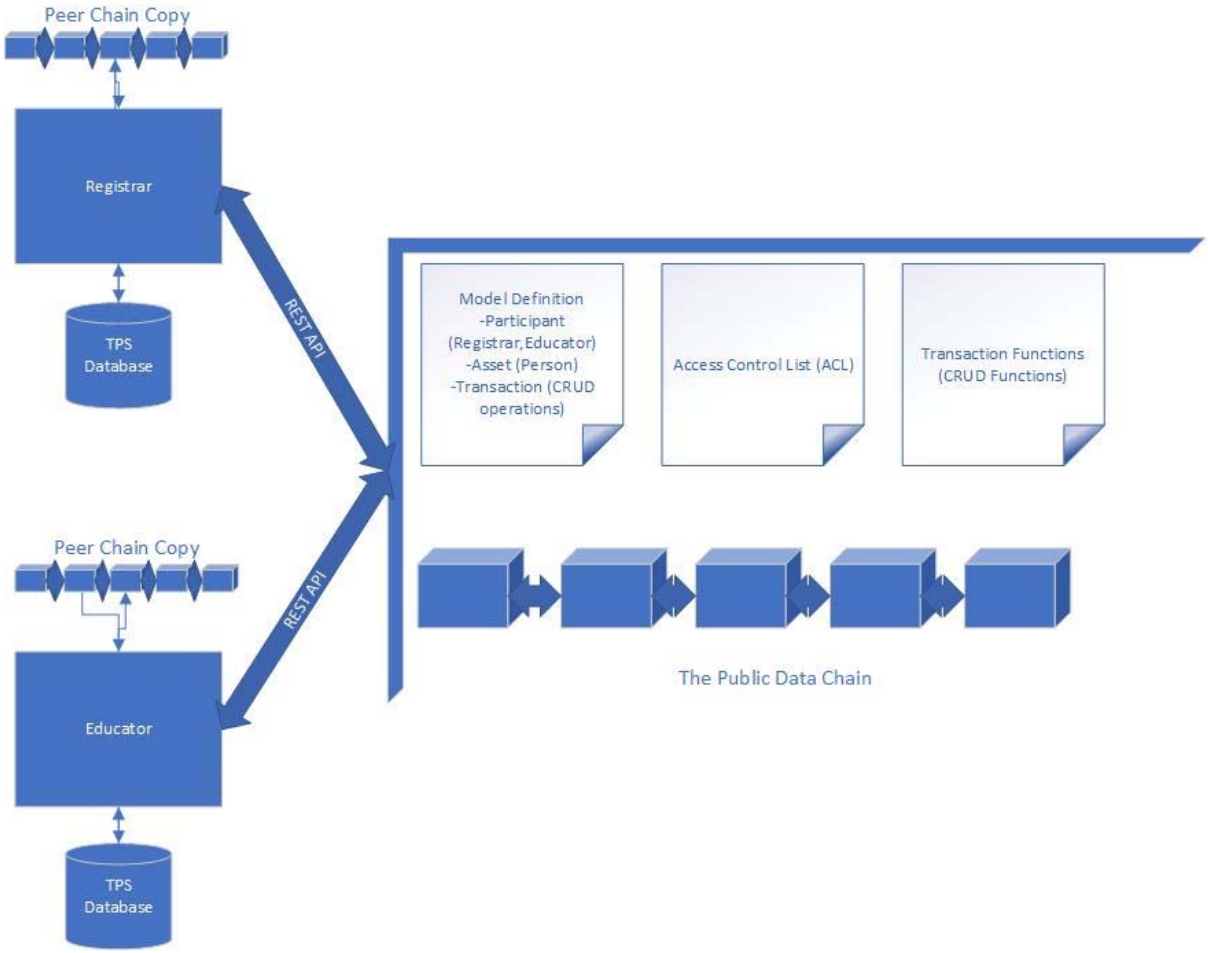


Figure 5. 12: Model Architecture

5.6 System Testing

System testing was done to determine the behavior of the system based on the requirements defined in chapter 4.2. To ensure the system meets the requirements, several tests were carried out as outlined in table 5.1:

Table 5. 1: System testing cases

ID	Case	Priority
1.0	Participants can join the network	High
2.0	Participants' agreement on the rules of engagement	High
3.0	Participants can write to the blockchain	High
4.0	Participants can read from the blockchain	High
5.0	Participant Authentication	High
6.0	Participant Access restriction	High
7.0	Reliability of input and output	Medium
8.0	System transparency	Medium

5.7 System Testing Cases and Process

5.7.1 Participants can join the network

The successful installation of the business network archive (.bna) on each participant node served as an invitation of the participant into the public data chain network. The *composer network ping* command was run to test the connection of the participant to the blockchain business network.

5.7.2 Participants' agreement on the rules of engagement

Successful installation of the chaincode or Smart Contract to all the peer nodes constituted the agreement for rules of engagement by the participant. The chaincode which was inbuilt in the business network file (.bna) serves as the contract between participants.

5.7.3 Participants can write to the blockchain

Invoking the *commitToBlockchain* method on the TPS and passing the Person object as a parameter, caused the passed asset object to be stored in the public data chain asset registry. This was observed also when education details were passed to the *commitToBlockchain* operation.

5.7.4 Participants can read from the blockchain

Invoking the *searchPersonDetails* method on the TPS and passing an individual's ID number as a parameter, a search was initiated on the public data chain. For valid ID numbers, the Person object was returned to the TPS.

5.7.5 Participant Authentication

At the point of adding peers, a digital certificate and a private key are assigned to the peers. These credentials are then used in signing every API request made to the blockchain via the Hyperledger Composer RESTful API. This ensured that all access requests are authenticated.

5.7.6 Participant Access restriction

Several ACL rules were defined in the permissions.acl file. In one of the rules, for example, the educator participant was unable to update a person's object after an ACL rule was included restricting the educator to only accessing the *updatePersonEducation* blockchain transaction.

5.7.7 Reliability of input and outputs

A comparison of personal data on the chain and in the local MySQL databases of the Registrar TPS and the Education TPS was done and the data was found to be similar. Although the blockchain includes some extra metadata such as the hash, this does not in any way affect the integrity of the data received from the TPS.

5.7.8 System transparency

The underlying operation of the system was transparent to the user. The users, in this case, the registrars and the educators would not tell that they were running on a blockchain network. The system design achieved transparency through the use of the REST APIs which were added as modular plugins to the transaction processing systems; connecting the TPS to the public data chain blockchain network.

5.8 System Testing Results

The system met all the requirements as outlined in the requirements analysis. The requirement for participant authentication and access restriction were adequately handled by the access control list definitions. Read/Write operation requirements passed with the ability to commit data to the blockchain and the ability to query using the REST API and the transaction function definitions. Reads and writes were also reliable since they were being done on the same distributed ledger thus ensuring a single source of truth. System usability was achieved through transparency since participants worked entirely on their internal systems while the blockchain plugin worked entirely in the background.

Table 5. 2:System Testing Outcomes

ID	Case	Priority	Outcome
1.0	Participants can join the network	High	Passed
2.0	Participants' agreement on the rules of engagement	High	Passed
3.0	Participants can write to the blockchain	High	Passed
4.0	Participants can read from the blockchain	High	Passed
5.0	Participant Authentication	High	Passed
6.0	Participant Access restriction	High	Passed
7.0	Reliability of input and output	Medium	Passed
8.0	System transparency	Medium	Passed

Chapter 6: Discussion

6.1 Introduction

Permissioned Blockchains offer an ideal solution to the problems inherent in population data management. The append-only nature of the blockchain guarantees provenance, a key quality of the blockchain technology, and a requirement for population data where an individual's information is constantly updated and maintained throughout the life of the individual - from birth to death. This enables the use of one unique identifier for the person, who is the asset in this context, thereby cutting out the need for multiple identification numbers and documents as well as duplication of effort and data in the identification of persons.

The use of Blockchain technology for population data management provides a single source of truth about an individual's information, effectively consolidating all population data in a single, trusted, and, secure data store. This data can then be made accessible to third-party users such as banks and educational institutions who apart from using the data for identity verification will further enrich an individual's data through committing industry-specific records such as financial and educational histories of the individual respectively.

The use of Permissioned Blockchains, where only authorized participants can join the network, coupled with smart contract and access control lists ensure the security required for population data, preventing misuse and unauthorized access that may lead to issues such as fraud and identity theft.

Permissioned Blockchain is therefore effectively suited for the task of population data consolidation and sharing.

6.2 Findings and Objectives Analysis

This prototype, by using the permissioned blockchain network, eliminates the need to maintain both an on-chain and an off-chain system such as the one proposed by Zyskind, Nathan, & Pentland (2015). All the participants joining the network will be pre-approved and therefore are trusted. Directly storing the data on the blockchain, therefore, does not pose a risk to the data because the unauthorized parties are not on the network. This approach of storing the data directly on the blockchain not only enhances consolidation and reduce fragmentation but also has the advantage of making the system design and architecture simpler.

This prototype is also simple in design and architecture and also transparent to the user. The use of APIs in the interaction between the blockchain and existing population data systems does not require drastic changes in existing systems or operating procedures. It has the benefit of being able to be plugged into the existing system without causing major changes.

This solution, however, poses a challenge to data ownership. Should the personal information of an individual collected over time from different sources belong to the government or the individual? Who should control access to this data by third parties? Zhang et al. (2018), Zyskind et al (2015) and Nchinda et al. (2019) have all proposed solutions where data privacy is maintained and controlled by the individual rather than the organization collecting and aggregating the personal information. In the case of population data, data ownership may not arise because governments must collect and maintain population databases of their citizens and residents. However, it poses a challenge during sharing where participants may be invited into the blockchain network, gain access to the data without the knowledge of the individual. Whereas it is lawful for the government to have the data, it may not be lawful for third party participants to have access to this data, and therefore a mechanism of the individual being informed before it is shared with third parties is required.

With this prototype, we have proven that indeed consolidation of population data is, in fact, possible through the permissioned blockchain technology. By using an individual's personal information as the asset on the blockchain, changes became available immediately to all participants within the network. As with the education records node within the person asset in the study, other participants also could update the person's information related to the participant domain.

Essentially, the structure adopted for storage of personal information with the person as the asset and therefore the root allows it to be extensible such that any new partner joining the network can have their domain-specific nodes created within the person object, for example, medical records, financial records, tax records, etc.

With the consolidation part where every participant further enriches the person object, sharing becomes easy such that all participants will be reading from the same chain, hence achieving the elusive single source o truth for population data.

Security and auditability were achieved largely through innate blockchain capabilities. Smart contracts and access control lists ensured business rules were defined and assigned correctly and access was restricted to only authorized parties. The transaction history is maintained all through starting from the genesis block thus ensuring auditability.

This solution further solves many of the issues with current systems of population data management such as data fragmentation and duplication, out-of-date non-synchronized data, difficulty in sharing data, lack of support for temporal data among others.

Interoperability costs will also go down since there will be no need for customized middleware integrations for data sharing but rather a standard API for participants once they join the public data chain network.

6.3 Model Contribution to Research

This study introduces a new way for population data management by governments through the use of a permissioned blockchain network. It shows how to design a national population register based on the consolidation of all population data and a unified way of sharing this data with authorized parties, thereby achieving the single source of truth principle; a key quality of population registers(OSCE /ODIHR, 2009). One unique ID assigned to the person at creation time is used throughout the lifetime of the individual. More importantly, it effectively makes the government the authoritative custodian of population data with control and oversight, making it easy to enforce penalties for misuse by network participants. The new mode of data sharing through the blockchain opens up opportunities in the development of population data-related systems such as real-time identification among others.

This study also proposes a better way for data sharing among government institutions as well as between government and non-government players. As observed by Yego (2014), challenges in data sharing among government institutions are due to a lack of clear policy, standards, communication protocol, incompatibility (interoperability issues), and data security issues. The National Audit Office of Great Britain also identifies data sharing as one of its three substantive issues in using data across government (National Audit Office, 2019). The proposed model in this study addresses the issue of data sharing through having related and data-dependent organizations become participants in the same blockchain network. In this model, issues of lack of policy and

standards are automatically addressed by the use of smart contracts which predefine terms and conditions for participant access to the data.

Security concerns are also addressed not only implicitly by use of the permissioned blockchain where only authorized users are invited into the network; accountability and auditability enhanced through the blockchain qualities of provenance, immutability, and finality; but also explicitly by use of smart contracts and access lists which collectively manages access to and manipulation of the consolidated data.

The proposed model is also effective in time, operational, and financial cost savings. Implementation and operation of this model are cheaper because it cuts out interoperability concerns since there is a unified way to access data on the chain. As observed by Choi, Ae Chun, Kim, & Keromytis (2013) consolidated data and unified sharing causes citizens to no longer have to visit different government offices to present documents, it saves the costs of multiple collections of citizen records and the resulting data fragmentation and provides as a central location where data security and privacy enforcement can be focused.

6.4 Limitations of the Model

Whereas Governments by law are mandated to collect and maintain population data and population registers, other third party users such participants in this network should not get access to this information without explicit permission from the person whose data is being shared. This system does not address this issue.

Chapter 7: Conclusions and Recommendations

7.1 Conclusion

The existing systems in Kenya for population data management continue to be plagued by various limitations considering there are still cases of duplicate identification documents; forgery of identification documents; identification documents of deceased persons still in use; and the existence of undocumented citizens. Besides these issues, there are multiple agencies and other stakeholders independently collecting personal data for their internal usage without a single authority to validate the correctness of the collected data.

The numerous drives to leverage information and communications technologies in the management of population data and identity management has seen the creation of various systems such as The Integrated Population Registration System (IPRS), The National Digital Registry System (NDRS), National Integrated Identity Management System (NIIMS) and The National Education Information Management System (NEMIS) which have all collected some form of personal information about the citizens. However, there is little to show that these systems are anywhere near achieving the proper consolidation and sharing of population data. Whereas the expected success scenario would mean the eradication of multiple identification documents for citizens; all government agencies still independently register and issue their identification documents to citizens. From the National Social Security number to the National Health Insurance number to the Kenya Revenue Authority PIN, every citizen is identified by three separate identification numbers, indicating the disconnect in achieving the single source of truth for a citizen's information.

This research aimed to provide a solution that addressed these shortcomings in the management of population data. Using the permissioned blockchain technology for data consolidation and sharing, this research has shown the many advantages of solving the problems plaguing the current systems and achieving the single source of truth for population data. Through the development of a public data blockchain network as the repository for all population data, and giving access to third party participants who need access to this data for their operations, the system lays the structure for the design of a proper way to a secure population data consolidation and sharing.

The proposed solution takes advantage of the distributed nature of blockchain technology for easy sharing of data, tracking data changes over time, ensuring security, immutability, and auditability of the population data.

The proposed solution takes advantage of the distributed nature of blockchain technology for easy sharing of data, tracking data changes over time, ensuring security, immutability, and auditability of the population data. The blockchain or the Distributed Ledger Technology is uniquely suited for management for population data because it is by nature an append-only data store. This quality suited in tracking the changes in an individual's life events over time, which by definition is temporal data, can be easily achieved on the blockchain yet it is a demerit of current relational database systems in use. Besides tracking temporal data, the append-only quality of the blockchain does enforce auditability of all the data in the system. Provenance, a key quality of the blockchain technology ensures the history of an asset, in this case, the personal information being tracked on the chain, can be accounted for from the genesis block.

Assets on the chain are tracked by unique identifiers. The model uses the individual as the asset to be tracked on the chain and assigns them a unique identification number. It is this unique identifier that is used to identify the person throughout the system. Any deviation from this identifier would mean the existence of an entirely new person, avoiding duplication.

Apart from the use of a unique identifier for every individual, another value proposition for the proposed model is data consolidation. In the proposed data chain, all institutional third parties that would need access to the data are invited as participants to the public data chain. With the conditions for participation set using smart contracts and the permission for access defined using the access control lists, these participants not only read data from the chain but are expected to also enrich an individual's information through writing institution-specific data to the chain. An example as demonstrated in the model would be the registrar of birth commits details of the birth of an individual to the chain. An education institution, being a participant in the blockchain network, would have access to this information which they will use for the student admission process. When the student completes the course, their academic achievement is committed to the chain making it available to the other institutions such as employers who may need to verify the academic credentials of their potential employees. This form consolidation is not only cheap but also achieves the much needed single point of reference for all data regarding an individual.

7.2 Recommendations

In light of the findings of the implementation and testing of the prototype, the following recommendations are made:

i. To ensure the single source of truth for population data, population registers should be stored in a blockchain network on which all updates and are made and all other dependent systems should read from this chain. To achieve this, an individual's information should be stored in a document-like structure with a unique identifier identifying the person document. Every other attribute about the person can be modeled as an attribute of the whole person document thus consolidating all information into one unified document.

ii. To ensure proper sharing of data, the adoption of "integrator network participants" is highly recommended. It will be impractical to have all institutions that depend on personal data to be included in the population data chain. The integrator network participants would act as the gateway for other non-critical third-parties to gain access to the chain. Their main role would be to control access and compliance, develop standard APIs and middleware components that give other organizations access to the chain.

iii. To ensure ease of sharing of data with dependent third-parties, standard APIs, and middleware should be developed to enable authorized dependent third-parties to easily integrate their systems to this central repository.

iv. To ensure no duplication and effective tracking of the person in the registry, only one unique number should be used in identifying the person and their records throughout all systems for the entire existence of the person.

v. Privacy and security are important for the proposed model. While privacy relates to the rights an individual has to control their personal information and how it is used, security refers to how the personal information is protected. Whereas security is guaranteed in our model, privacy is not guaranteed especially because data is shared with third parties who may misuse the data. To ensure privacy we recommend the implementation of a privacy policy implementation where a person is informed before their data is shared with third parties. Once the person consents that their data can be shared with a third party only then should that data be shared.

7.2 Future Work

In light of the current trend of personal data management; especially the General Data Protection Regulation (GDPR) and Self-sovereign Identity (SSI), personal information security is key. To improve personal data security and privacy, before sharing personal data with dependent third parties, consent must be sought from the owners of this information. Therefore, to further improve this system, further research needs to be done on the integration of the owner's consent management into the system. This will ensure no misuse of personal information by third parties as well as give control to the personal information owner to share or revoke access to their data by third parties.

References

Agile Alliance. (2019, August 24). What is Agile Software Development? Retrieved August 26, 2019, from https://www.agilealliance.org/agile101/

Aguilar Rivera, A. M., & Vassil, K. (2015, November 1). *Estonia - A successfully integrated population-registration and identity management system: delivering public services effectively*. Retrieved October 20, 2019, from http://documents.worldbank.org/curated/en/873061495178335850/Estonia-A-successfully-integrated-population-registration-and-identity-management-system-delivering-public-services-effectively

Andrade, N. N., Chen-Wilson, L., Argles, D., Wills, G., & Zenise, M. S. (2014). *Electronic Identity*. Basingstoke, England: Springer.

Beck, C., Dumay, J., & Frost, G. (2015). In Pursuit of a 'Single Source of Truth': from Threatened Legitimacy to Integrated Reporting. *Journal of Business Ethics*, *141*(1), 191-205. doi:10.1007/s10551-014-2423-1

Bitcoin. (2009). Retrieved March 23, 2020, from https://bitcoin.org/en/

Blockchain4aid. (2018). Building Blocks: an analysis. Retrieved February 26, 2020, from https://blockchain4aid.org/analysis/building-blocks/

Brakeville, S., & Perepa, B. (2019, June 1). Blockchain Basics: Introduction to Distributed Ledgers. Retrieved September 16, 2019, from https://developer.ibm.com/tutorials/cl-blockchain-basics-intro-bluemix-trs/

Breckenridge, K. (2018). The failure of the 'single source of truth about Kenyans': The NDRS, collateral mysteries and the Safaricom monopoly. *African Studies*, *78*(1), 91-111. doi:10.1080/00020184.2018.1540515

Choi, J., Ae Chun, S., Kim, D. H., & Keromytis, A. (2013). SecureGov. *Proceedings of the 14th Annual International Conference on Digital Government Research - dg.o '13*. doi:10.1145/2479724.2479745

Cocco, S., & Singh, G. (2018). Top 6 technical advantages of Hyperledger fabric for blockchain networks. Retrieved April 16, 2020, from https://developer.ibm.com/articles/top-technical-advantages-of-hyperledger-fabric-for-blockchain-networks/

Computer Weekly. (2011, September 23). Understanding the UID Aadhaar project and IT's role in its success. Retrieved August 25, 2019, from https://www.computerweekly.com/tutorial/Understanding-the-UID-Aadhaar-project-and-ITs-role-in-its-success

Date, C. J., Darwen, H., & Lorentzos, N. A. (2003). *Temporal Data and the Relational Model: A Detailed Investigation Into the Application of Interval and Relation Theory to the Problem of Temporal Database Management*. Burlington, MA: Morgan Kaufmann.

DeMuro, J. (2018, January 16). Here are the 10 sectors that blockchain will disrupt forever. Retrieved March 17, 2020, from https://www.techradar.com/news/here-are-the-10-sectors-that-blockchain-will-disrupt-forever

DeMuro, J. (2018, January 16). Here are the 10 sectors that blockchain will disrupt forever. Retrieved July 3, 2019, from https://www.techradar.com/news/here-are-the-10-sectors-that-blockchain-will-disrupt-forever

Digital Identification Document (ID) & Citizenship Consultative Meeting. (2019). Retrieved from https://www.khrc.or.ke/publications/198-report-of-digital-identification-citizenship-workshop-naivasha/file.html

Distributed Ledgers Technology and Artificial Intelligence Task Force. (2019). *Emerging Technologies for Kenya: Exploration and Analysis*. Retrieved from Ministry of Information, Communications and Technology website: https://www.ict.go.ke/blockchain.pdf

E-estonia. (2017, May 31). X-Road. Retrieved October 20, 2019, from https://e-estonia.com/solutions/interoperability-services/x-road/

Ethereum. (2020). Retrieved March 23, 2020, from https://ethereum.org/

Falazi, G., Hahn, M., Breitenbucher, U., Leymann, F., & Yussupov, V. (2019). Process-based composition of Permissioned and Permissionless blockchain smart contracts. *2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC).* doi:10.1109/edoc.2019.00019

Fzaninotto. (2020). Fzaninotto/Faker. Retrieved April 27, 2020, from https://github.com/fzaninotto/Faker

Gates, B. (2019, January 29). Making the world's invisible people, visible. Retrieved August 25, 2019, from https://www.gatesnotes.com/Development/Heroes-in-the-Field-Nandan-Nilekani

Gelb, A., & Metz, A. D. (2018). *Identification Revolution: Can Digital ID be Harnessed for Development?* Washington, DC: Brookings Institution Press.

Health Knowledge. (2018). Methods of sampling from a Population. Retrieved April 27, 2020, from https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/methods-of-sampling-population

Hirtan, L., Krawiec, P., Dobre, C., & Batalla, J. M. (2019). Blockchain-based approach for e-Health data access management with privacy protection. *2019 IEEE 24th International*

Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). doi:10.1109/camad.2019.8858469

IBM. (2015). IBM Sterling Supply Chain | Build Intelligent Supply Chains. Retrieved from https://www.ibm.com/supply-chain

Jaeger, L. G. (2018, October 4). Public versus private: What to know before getting started with blockchain. Retrieved December 9, 2019, from https://www.ibm.com/blogs/blockchain/2018/10/public-versus-private-what-to-know-before-getting-started-with-blockchain/

Jain, M. (2019, July 11). The Aadhaar Card: Cybersecurity Issues with India's Biometric Experiment. Retrieved October 20, 2019, from https://jsis.washington.edu/news/the-aadhaar-card-cybersecurity-issues-with-indias-biometric-experiment/

Kakavand, H., Kost De Sevres, N., & Chilton, B. (2017). The Blockchain Revolution: An Analysis of Regulation and Technology Related to Distributed Ledger Technologies. *SSRN Electronic Journal*. doi:10.2139/ssrn.2849251

Karafiloski, E. (2017, November 12). Blockchain: Everything you need to know about the revolutionary technology. Retrieved March 1, 2020, from https://blog.netcetera.com/blockchain-everything-you-need-to-know-about-the-revolutionary-technology-7579f82b05d

Keil, J. (2019, September 5). Self-Sovereign Identity Systems: How Businesses Win From Letting Go of Customer's Data. Retrieved December 9, 2019, from https://hackernoon.com/self-sovereign-identity-systems-how-businesses-win-from-letting-go-of-customers-data-1v3fz31n0

Kenya Law Reform Commission. (2019, April 3). The Bliss of NIIMS Paradise: The Legal Context for the Huduma Namba. Retrieved July 1, 2019, from http://www.klrc.go.ke/index.php/klrc-blog/645-the-bliss-of-niims-paradise-the-legal-context-for-the-huduma-namba

KLRC. (2019, April 3). The Bliss of NIIMS Paradise: The Legal Context for the Huduma Namba. Retrieved August 2, 2019, from http://www.klrc.go.ke/index.php/klrc-blog/645-the-bliss-of-niims-paradise-the-legal-context-for-the-huduma-namba

KNBS. (2019). *2019 Kenya Population and Housing Census*. Retrieved from Kenya National Bureau of Statistics website: https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county

Kshetri, N., & Voas, J. (2018). Blockchain-Enabled E-Voting. *IEEE Software*, *35*(4), 95-99. doi:10.1109/ms.2018.2801546

Kulhari, S. (2018). II. The Midas touch of Blockchain: Leveraging it for Data Protection. *Building-Blocks of a Data Protection Revolution*, 15-22. doi:10.5771/9783845294025-15

Laws of Kenya. (1928, June 9). The Births and Deaths Registration Act. Retrieved August 2, 2019, from http://kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=CAP.%20149

Laws of Kenya. (1949, May 16). The Registration of Persons Act. Retrieved August 2, 2019, from http://kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=CAP.%20107

Laws of Kenya. (2011, October 4). The Kenya Citizens and Foreign Nationals Management Service Act. Retrieved August 3, 2019, from http://kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=No.%2031%20of%202011

Laws of Kenya. (2011, August 30). The Kenya Citizenship and Immigration Act. Retrieved August 2, 2019, from http://kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=No.%2012%20of%202011

The Linux Foundation. (2009). Hyperledger Fabric – Hyperledger. Retrieved August 26, 2019, from https://www.hyperledger.org/projects/fabric

LucidChart. (2018, August 10). 4 Phases of Rapid Application Development Methodology Blog. Retrieved August 26, 2019, from https://www.lucidchart.com/blog/rapid-application-development-methodology

Middleton, F. (2020, January 13). Reliability vs Validity in Research. Retrieved April 27, 2020, from https://www.scribbr.com/methodology/reliability-vs-validity/

Mulligan, C., Scott, J. Z., Warren, S., & Rangaswami, J. P. (2018). *Blockchain Beyond the Hype A Practical Framework for Business Leaders* (120418). World Economic Forum.

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2020). Identity Verification Standards in Welfare Programs: Experimental Evidence from India. doi:10.3386/w26744

National Audit Office. (2019). *Cross-government: Challenges in using data across government*. Retrieved from National Audit Office website: https://www.nao.org.uk/wp-content/uploads/2019/06/Challenges-in-using-data-across-government-Summary.pdf

National Office for Identity Data. (2015). *Identity Management in 2030*. Ministry of Interior and Kingdom relations The Netherlands.

Nchinda, N., Cameron, A., Retzepi, K., & Lippman, A. (2019). MedRec: A network for personal information distribution. *2019 International Conference on Computing, Networking and Communications (ICNC)*. doi:10.1109/iccnc.2019.8685631

Omoka, R. S. (2020, May 19). romoka/public-data-chain. Retrieved May 29, 2020, from https://github.com/romoka/public-data-chain

Omusolo, M. (2014, October 2). Why NDRS for Kenya? Retrieved August 2, 2019, from http://c4dlab.uonbi.ac.ke/2014/10/why-ndrs-for-kenya/

OSCE /ODIHR. (2009). *Guidelines on population registration*. Warsaw: OSCE's Office for Democratic Institutions and Human Rights (ODIHR).

OSCE, & ODIHR. (2013). *Addressing the Link between Travel Document Security and Population Registration/Civil Registration Documents and Processes*. Retrieved from https://www.osce.org/odihr/110685?download=true

Perrigo B. (2018, September 28). India Is Collecting a Vast Database of Eye Scans and Fingerprint Records. Retrieved September 23, 2019, from https://time.com/5409604/india-aadhaar-supreme-court/

SSH Communications Security Corp. (1995). What is PKI (Public key infrastructure)? Retrieved April 27, 2020, from https://www.ssh.com/pki/

Swedish Tax Administration. (2010). Population registration in Sweden. Retrieved September 16, 2019, from https://unstats.un.org/unsd/vitalstatkb/KnowledgebaseArticle50188.aspx

Tech2 News. (2018, September 25). Aadhaar security breaches: Here are the major untoward incidents that have happened with Aadhaar and what was actually affected. Retrieved October 20, 2019, from https://www.firstpost.com/tech/news-analysis/aadhaar-security-breaches-here-are-the-major-untoward-incidents-that-have-happened-with-aadhaar-and-what-was-actually-affected-4300349.html

Types of Research - Research Methodology. (n.d.). Retrieved from https://research-methodology.net/research-methodology/research-types/

United Nations, U. N. (1969). Methodology and evaluation of population registers and similar systems. Retrieved from https://unstats.un.org/unsd/publication/SeriesF/Seriesf_15e.pdf

United Nations. Statistical Division. (2001). *Principles and Recommendations for a Vital Statistics System*. New York, NY: United Nations Publications.

UN-OICT. (2017). *#Blockchain4Humanity : Use the blockchain technology to help combat child trafficking in Moldova*. Retrieved January 2, 2020, from https://ideas.unite.un.org/blockchain4humanity/Page/Home

Veres One. (2018). The Veres One Project. Retrieved March 15, 2020, from https://veres.one/summary/

Verhoef, R., & Van de Kaa, D. J. (1987). Population Registers and Population Statistics. *Population Index*, *53*(4), 633. doi:10.2307/3643792

WFP. (2016). *Building Blocks*. Retrieved February 26, 2020, from https://innovation.wfp.org/project/building-blocks

Wilson, P. M., Petticrew, M., & Calnan, M. W. (2010). Disseminating research findings: what should researchers do? A systematic scoping review of conceptual frameworks. *Implementation Science*, *5*, 91. Retrieved from https://doi.org/10.1186/1748-5908-5-91

World Economic Forum Strategic Intelligence. (2018). Explore the latest strategic trends, research and analysis on Blockchain. Retrieved December 9, 2019, from https://intelligence.weforum.org/topics/a1Gb00000038qmPEAQ?tab=publications

World Identity Network. (2018). *Turning Invisible Children into Invincible Ones*. Retrieved from World Identity Network website: https://win.systems/wp-content/uploads/2018/09/blockchain-for-humanity.pdf

World Population Review. (2019, July 11). Kenya Population 2019. Retrieved August 2, 2019, from http://worldpopulationreview.com/countries/kenya-population/

Wust, K., & Gervais, A. (2018). Do you Need a Blockchain? *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. doi:10.1109/cvcbt.2018.00011

Xu, X., Weber, I., & Staples, M. (2019). *Architecture for blockchain applications*. Springer.

Yego, P. K. (2014). *Data Sharing Among Selected Government Institutions in Kenya* (Master's thesis, University of Nairobi, Nairobi, Kenya). Retrieved from http://erepository.uonbi.ac.ke/bitstream/handle/11295/90235/Yego_Data%20Sharing%20Among%20Selected%20Government%20Institutions%20in%20Kenya.pdf?sequence=4&isAllowed=y

Young K. (2019). The Promise of Public Interest Technology: In India and the United States. Retrieved September 23, 2019, from https://www.newamerica.org/fellows/reports/anthology-working-papers-new-americas-us-india-fellows/key-differences-between-the-us-social-security-system-and-indias-aadhaar-system-kaliya-young/

Zhang, P., White, J., Schmidt, D. C., Lenz, G., & Rosenbloom, S. T. (2018). FHIRChain: Applying blockchain to securely and Scalably share clinical data. *Computational and Structural Biotechnology Journal*, *16*, 267-278. doi:10.1016/j.csbj.2018.07.004

Zyskind, G., Nathan, O., & Pentland, A. '. (2015). Decentralizing privacy: Using blockchain to protect personal data. *2015 IEEE Security and Privacy Workshops*. doi:10.1109/spw.2015.27

Appendix A : Originality Report