



HiLCoE

School of Computer Science and Technology

The use of Blockchain for Government Data Management

Tsega Debebe Worku

A thesis submitted to the Graduate Programme Research Office
of HiLCoE in partial fulfillment of the requirements for the
Degree of Master of Science in Software Engineering.

May 2023

HiLCoE

School of Computer Science and Technology

The use of Blockchain for Government Data Management

Written by: Tsega Debebe Worku

Advisor: Mesfin Belachew (PhD)

Name and Signature of Members of the Examining Committee

No.	Role	Name	Signature
1.	Advisor	Mesfin Belachew (PhD)	
2.	External Examiner		
3.	Internal Examiner		
4.	Examining Committee Chairperson		

Acknowledgment

I would like to express my sincere gratitude to everyone who supported me throughout this research project. First and foremost, I thank my family for their unconditional love, encouragement and understanding. They have always been there for me in times of joy and hardship. I am truly blessed to have them in my life.

Secondly, I thank my advisor, Mesfin Belachew (PhD), for his invaluable guidance, feedback and mentorship. I acknowledge the support from HiLCoE school's instructors and staff. I am grateful to the reviewers and examiners of the thesis for their constructive comments and suggestions. They have helped me improve the quality and clarity of my paper. I also thank all my colleagues for their insights, friendship and camaraderie. I could not have completed this paper without the help of these wonderful people.

Last but not least, I thank all the known and unknown open source contributors who have dedicated their lives in creating, developing and maintaining blockchain technologies that enabled this research. Their work is an inspiration and a source of innovation for the scientific community.

Table of Contents

Acknowledgment	2
Table of Contents	3
List of Tables	4
List of Figures	5
Abstract	6
Chapter 1: Introduction	7
1.1 Background	7
1.2 Motivation	9
1.3 Statement of the Problem	12
1.4 Objectives	13
1.5 Methods	14
1.6 Scope and Limitation	16
1.7 Thesis Organization	17
Chapter 2: Literature Review and Related Work	18
2.1 Literature Review	18
2.2 Related Work	22
Chapter 3: Methodology	24
3.1 Sampling Methodology	24
3.2 Data Gathering	30
Chapter 4: Data Analysis	35
Chapter 5: Solution of the Research	37
5.1 Building the System	38
5.2 Document Verification	46
5.3 Connecting Data Sectors	49
Chapter 6: Conclusion and Recommendation	51
6.1 Conclusion	51
6.2 Recommendation	52
Chapter 7: Future Work	54
Reference	56
Declaration Sheet	59

List of Tables

Table 2.1: Tabular representation of a blockchain

Table 3.1: Reduced version of stratified sampling on a google spreadsheet

Table 3.2: Reduced version of convenience sampling on a google spreadsheet

Table 3.3: Reduced version of the data identity gathered on a google spreadsheet

List of Figures

Figure 3.1: A pictorial representation of the selected countries using stratified sampling

Figure 3.2: A pictorial representation of the selected countries using convenience sampling

Figure 3.3: Count of data type from Identity Sector

Figure 3.4: Count of data type from Automobile Sector

Figure 3.5: Count of data type from Property Sector

Figure 4.1: Count of Data types in Identity Automobile and property sectors

Figure 5.1: Adding an identity token

Figure 5.2: Using the Blockchain Explorer to verify a token

Figure 5.3: Verification Passed on an Identity token

Figure 5.4: Verification Failed on an Identity token

Figure 5.5: A visualization of connected tokens from different data sectors

Abstract

Government databases store sensitive and valuable information that affects the public interest and trust, such as citizen records, land registries, health records, tax records, etc. However, these databases are often vulnerable to tampering, corruption, or unauthorized access by malicious actors or insiders.

Blockchain technology has emerged as a promising solution for ensuring data integrity and transparency in various domains, including government databases. And it can address these challenges by leveraging its data integrity property, which ensures that once data is recorded on a blockchain, it cannot be modified or deleted without leaving a trace.

Governments can benefit from a blockchain based database system because it can ensure that the data stored in the government database is accurate, consistent and immutable, which can enhance the trust and transparency of the public sector. It can also prevent unauthorized access, tampering or deletion of the data, which can protect the privacy and security of the citizens and the government. And blockchain can also enable faster and cheaper verification and validation of the data, which can reduce the administrative costs and errors caused by government employees. This enhances the transparency and accountability of the government data management and reduces the risks of fraud, manipulation, or human error.

Keywords: Government, Blockchain, Hashing, Document Verification

Chapter 1: Introduction

1.1 Background

This research studies the practicability and benefits of using blockchain by nations around the world for the purpose of data management. In order to properly serve its citizens, every country needs to collect and store data of individuals' identity, their history, their assets, their businesses and their associations. Most developed countries use an electronic database system (for example open source relational database management systems like MySQL [16]) while other countries use the paper-based method. The electronic database system is more advantageous than the paper based for handling data entry, data amendment and data retrieval but it is vulnerable to unauthorized data usage, data loss and data corruption. This research first verified if it is possible to use blockchain for such purposes, then it discussed the benefits of using such a system for government data management systems.

The general purpose of a blockchain is allowing digital information to be recorded and distributed but never to be altered or deleted. This made blockchain an ideal method for money transaction databases. It is a common knowledge that blockchains are used for monetary purposes as a decentralized currency ledger. But this research does not include the monetary purposes of blockchains. This research only focuses on using blockchain for governmental database management purposes only.

Some of the challenges that many governments face are fabrication of identities and document forgery and they can overcome these challenges by implementing blockchain technology to their data management. *Blockchain technology can secure integrity of files stored in the database* [1], meaning once data is written on a blockchain it is mathematically impossible to alter or remove it. If we can harness this power of blockchain for governmental uses then we can build a nearly perfect system for any government in the world to store and manage their data.

A government is simply a system that manages a community or a state. In order to properly serve its citizens, any country needs to collect and store data. Most of the services given by a government are divided into different sectors. For the context of this study government data sectors are divided into six parts. Identity sectors (includes identity cards, passports, voting card registration, birth certificate, death certificate, relationship status ...), Financial sectors (includes business licenses, tax history, employee records...), Automobile sectors (includes driving licenses and vehicle licence/title registration), Property sectors (includes land deed registration), Educational sectors (include the current educational status and previous educational certificates) and Legal sectors (include police records, court rulings, government policies, constitution, legislations, regulations, contractual agreement, document authentication and registration services).

All these different government data sectors can be connected to each other. Meaning for example when an individual wishes to start a business, that individual needs services from; identity sectors to verify himself/herself, financial sectors to certify the legality of the businesses, automobile sectors to purchase and/or register vehicles and hire drivers, property sectors to register lands, buildings, office spaces, contractual agreement sectors to authenticate and register contractual documents and maybe educational sectors if the business needs such requirements. A government can prevent fraud and increase efficiency if all the databases are built on a single blockchain system linked together, meaning for example after registering individual identities of the citizens on the blockchain, the government will register their assets (for instance land deeds and vehicle certificates) and link the assets data to the individual data together. Then whenever data of an individual is requested all their assets will be retrieved and whenever data of an asset is requested all the proprietors of that asset will be retrieved. This research will study if a government can be even more effective if it developed one large centralized blockchain to gather and connect the data from all these sectors together.

1.2 Motivation

Few of the many scenarios that inspired this research's idea are listed below. The first motivation is from Mexico, indicating how criminals take advantage of the bureaucracy of verifying the authenticity of someone's identity documents that lead to tax frauds and other criminal actions. The second motivation is based on a research output done in Malaysia, indicating most frauds on land registration happen for the simple reason of neglecting to verify validity of documents. The final motivation for this research is from Australia, describing the importance of proofing one's identity during vehicle registration not only results in providing valid registration documents but in time it also discourages vehicle theft. These motivations are the inspiration for the research.

Motivation One and Proposed Solution

In Mexico, scammers create “fake” official identifications for the purposes of opening businesses, opening bank accounts, creating contracts and printing invoices. They disappear shortly after without having paid any taxes and, in the worst cases, perform other criminal activities. When the authorities detect the criminal activities, the investigation begins with the original owner of the ID [6].

In a blockchain based government system, verification of a document's authenticity is simple and easy. And if the government is willing, this verification system can be readily available for private businesses without compromising an individual's privacy. Meaning whenever someone provides their proof of self-identification or business identification, it can easily be verified for authentication. Since it is detectable if lawbreakers manipulated a blockchain based database system, the “fake” official identification could be apprehended before they incur any damages.

Motivation Two and Proposed Solution

In studies done by Low [20], Maidin and Khadouf [21], and Ismail [22] they have identified the specific patterns of fraud land registration in Malaysia ... fraud by forgery –

registering dealings using the Power of Attorney, forging of Transfer Form or Charge Form, misusing court order to register dealings illegally without verification of validity ... fraud by misrepresentation – appointing or using unauthorized persons to execute the activities of ownership ... fraud by alteration – issuing the replacement of title document under the pretext that the original title has lost ... fraud by cyber terrorist or programmer who co-operated with the system ... fraud by payment [19]

These criminal actions became successful because of the tedious process of verifying a document's authenticity by both the public and government offices. Since it is impossible for the lawbreaker to forge a verifiable document staged on a blockchain based database system, these criminal actions could have been mitigated.

Motivation Three and Proposed Solution

Analysis of vehicle registration documentation showed that all eight states/territories in Australia require the person registering a vehicle to show proof of identity (typically a driver's license) to the vehicle registration authority. There has long been a requirement to prove identity when registering a vehicle in Australia, but this became more stringent when photographs were introduced on to drivers licenses in the late 1980's. Proof of identity at the point of registration helps to prevent simple vehicle cloning (either intra- or inter-state) involving the copying of an existing legitimate vehicle's identity, because the legitimate vehicle will be linked to another person's identity. [23]

Proof of identity can also prevent a change of address fraud, a variant of cloning in which a vehicle registration document is obtained by registering a vehicle owned by another person at an address of the vehicle thief's choice. Replacement registration documentation is then obtained from the vehicle registration authority, with the identity of the stolen vehicle being changed to match those on the documentation. However, in Australia, the proof of identity check means that an address cannot be changed unless the registered keeper's identity is first verified. [23]

This motivation describes the importance of verifying proof of identity using a driver's license and verifying authenticity of vehicle registration certificates to prevent crimes of vehicle theft. Since it is impossible for criminals to forge a verifiable document on a blockchain based database system, these criminal actions can be discouraged.

Motivations of the Researcher

Ever since the inception of blockchain technology many individuals and companies have been adapting it for purposes other than money management. And in most of these implementations blockchain brought more enhanced data security than the legacy database security system. This motivated the idea of using blockchain for government data management purposes, and studying the practicability and benefits of using such a system. Blockchain brings the advantages of immutable record keeping and verifiable documentation systems and this research is motivated to use these advantages of blockchain in combating identity fraud and document fabrication issues.

1.3 Statement of the Problem

Government organizations and citizens around the world face similar challenges. They are constantly fighting against fabrication of identities (IDs, Passports, Driving Licenses [23], Voting cards, TAX fraud in relation to fake business licenses [6]) and document forgery (Land Deed theft [2] [4] [19], vehicle registration certificates [23], contract agreement alteration [3], education certificates tempering [5]).

Most governments are susceptible to data manipulation. Lawbreakers could bribe government officials, hack into the government database or exploit institutional loopholes to falsely manipulate a government's data banks. Since the main advantage of using blockchain is to prevent data manipulation, this study is motivated to use blockchain technology for government data management.

A Blockchain operates by hashing its data and retaining the hash value for later verification. This makes it an ideal choice for data authentication. Even a one character alteration of the data will result in a different hash value, rendering the entire data erroneous. Which means verification of a document's authenticity will be easier as long as it is stored on the blockchain. All the above motivations (in chapter 1.2 Motivation) arose because document verification is a complex system in most countries. Let's say someone wants to open a bank account or a business license, they are required to present their self-identification, a passport, a driving license or any document issued by the government. In a blockchain based government, the government can build some type of a web system, a simple website or an api based system, that accepts the data from the identification document and verifies its authenticity, it can stop fabrication of identities. And the government can also build this system without compromising individual privacy.

Applying blockchain technologies to the governmental sectors, separately, can be helpful. But, this study will also navigate the application of blockchain to all the governmental sectors together. Meaning if a government developed one large centralized blockchain to store and link all these data together it would be able to prevent fraud and increase efficiency in giving public services.

1.4 Objectives

1.4.1 General Objectives

The general objective of this research is to explore the benefits and practicability in implementing blockchain technology for the purposes of government data management and propose a blueprint in order to build a blockchain based governmental system.

1.4.2 Specific Objectives

The specific objectives of this research are:

1. Study the feasibility of using blockchain for government database management
2. Study the benefits of using blockchain for government database management
3. Identify the best way to implement this system
4. Propose a blueprint for using blockchain in some or all parts of the governmental system.

1.5 Methods

This section describes the methodology techniques used to prepare this research. Before researching the effect of blockchain technology for government data management, it is necessary to ask which government and which data sectors should the research be focused on.

Before making a definitive claim that blockchain can be used as a governmental database system by any country, one needs to examine each country. But since it is impractical to research all 195 countries in the world, this research sampled very few of them and oversaw the practicability of implementing blockchain technology for the purposes of government data management. The sampling was performed first by eliminating countries with similar record keeping systems, then the sample size is further reduced based on their impact on the world. The sampling methods used to sample the countries are stratified sampling and convenience sampling. Stratified sampling is a type of random sampling where researchers first divide a population into smaller subgroups, or strata, based on shared characteristics of the members and then randomly select an equal number of samples from each division to form the final sample. Convenience sampling is a non-probability sampling method where units are selected for inclusion in the sample because they are the easiest for the researcher to access. This can be due to geographical proximity, availability at a given time, or willingness to participate in the research. These two sampling methods were considered after a close examination of all other sampling methods and choosing the most appropriate ones for the study.

Out of all the sectors inside a government that could use a data management system for, this research sampled the most important ones that deals with identity fabrication and document forgery. The sampling method used to choose between the data sectors is purposive sampling. Purposive sampling is a non-probability sampling technique used in research to select individuals or groups of individuals that meet specific criteria relevant to the research question or objective.

After the countries and data sectors were sampled, the next step was data gathering. The proof of documentation for all of the countries and for all the data sectors were gathered for analysis. The data gathering method used for this research is document analysis method.

Document analysis is a qualitative research method that involves evaluating electronic and physical documents to interpret them, gain an understanding of their meaning, and develop upon the information they provide. This data gathering method is employed because it is the only way one can gather the documents of different nations in different government sectors.

Surveys, interviews, case study and focus group methodologies are not applicable for this research because this research is not focused on building a system for only one nation. This research is about building a generalized system that can be applied to any country.

After data gathering, the data analysis was performed using a deductive approach. The deductive approach is a research method that involves developing a hypothesis based on existing theory and then designing a research strategy to test the hypothesis. This approach is concerned with deducting conclusions from premises or propositions.

1.6 Scope and Limitation

The research limited the focus on three government data sectors: Identity sectors, Automobile sectors, Property sectors. These three government sectors motivated the research. Other government data sectors (such as Financial sectors, Education sectors, Legal sectors, healthcare sectors, infrastructure sectors, security sectors, defense sectors...), while may be eligible for blockchain representation, are not covered in this research.

Even though this research proposes a blueprint that can be applied globally, it does not study the current system of all nations in the world. Instead, it studied a few selected countries that were chosen using sampling techniques.

One of the advantages of placing data on a blockchain is how remarkably easy it is to verify the validity of the data. But this creates privacy issues where some individuals may oppose the idea of placing their identity data accessible to the public. These types of scenarios could compel the government to encrypt the data before placing it on the blockchain. But privacy issues are not covered in this research.

Although, rightnow, blockchain technology is widely used for monetary purposes, as a decentralized currency ledger, this research does not study the monetary use case of a blockchain.

1.7 Thesis Organization

In this section the structure of this thesis is discussed, it explains how this thesis is organized and what each chapter covers.

The introduction chapter is the first chapter of this research. It includes the motivations, the problem statements, the objectives, the methods and the scope as subchapters. The second chapter discusses literature review and related works. The definitions and usage of some of the complex terms used in this research are also discussed in this chapter. The third chapter discusses the different forms of sampling and data gathering methodologies employed in this research. The fourth chapter discusses analysis of the gathered data. The fifth chapter discusses the perceived solutions to the statement of the problem and shows a proof of concept. The sixth chapter concludes this research by discussing the benefits of using such technologies and gives recommendations on how to implement the system. The seventh chapter discusses how a future researcher can extend the current work. Lastly, the reference section cites all the references used in making this research.

Chapter 2: Literature Review and Related Work

2.1 Literature Review

This thesis employs some terms that may have ambiguous or wide-ranging meanings, such as Governmental Data, other terms they may have a complex meaning, such as Blockchain and hashing. These terms are defined and clarified in the following section to avoid confusion and ensure consistency throughout the document.

2.1.1 Blockchain and Hashing

Before the creation of Bitcoin, in 2008 by an individual or a group named Satoshi Nakamoto [8], the concept of Blockchain was introduced by Stuart Haber and W. Scott Stornetta, in 1991, while they were trying to build a system in which a digital document timestamps could not be modified [9]. Nowadays, a ‘timestamp’ is referred to as a hash signature.

Hashing was first invented by an IBM engineer named Hans Peter Luhn in 1958. He created an algorithm called KWIC, for Key Word in Context, that inputs a series of words and effectively calculates its index value or in the latest term, hash value [10].

Hashing algorithms work because their output is unique to the input data. Let’s say we want to hash data, x , with a hashing algorithm $h(x)$, in which $h(x)$ will return a string value s . If a single byte was changed from the data, x' , then the hashed string value will also change; $h(x') = s'$. Which means the values of $h(x)$ and $h(x')$ will never be mathematically and cryptographically equal.

A hash signature is a fixed-length string of letters and numbers generated by a hashing algorithm that is unique to a specific document.

Blockchain is defined as a growing list of records called blocks, which are linked and stored using cryptography [11]. Every block inside a blockchain is composed of 3 things:

Previous hash: the hash signature of the previous block, Data: any digital document, and Hash: is the current block hash signature (or the hash of the current data with the previous hash).

The first block, also known as Block 0 or Genesis Block [12], does not have a previous hash value so it can be any string depending on the developer's choice. In Table 2.1, the hash value of the first block is zero. Then the second block, Block 1, will have a hash signature of combination of the first data and the previous hash which is zero. The system can continue linking blocks forever which is why it is called a blockchain.

Table 2.1: Tabular representation of a blockchain

	Previous hash	Data	Hash
Block 0	-	-	0
Block 1	0	D_1	$h_1 = \text{hash}(D_1, 0)$
Block 2	h_1	D_2	$h_2 = \text{hash}(D_2, h_1)$
Block 3	h_2	D_3	$h_3 = \text{hash}(D_3, h_2)$
:	:	:	:
Block n	h_{n-1}	D_n	$h_n = \text{hash}(D_n, h_{n-1})$

If someone wants to change the content of D_2 , for instance, it will trigger a miss match with the original value of h_2 , so h_2 must be recalculated to accommodate the changes of D_2 . But the alteration of h_2 will create another miss match on the next block. Let's say the number of blocks at that moment is a million. This means every time someone wants to change a single byte on a single data, they must consequently change millions of hash signatures which is usually very expensive.

2.1.2 Governmental Data

From the context of this research the governmental data is the data that is gathered and stored by the government. Any and all data that is gathered and retained by the government can be considered as governmental data. Which includes but is not limited to: individual identification, passports, driving licenses, vehicle ownership certificates, business licenses, land deeds, contract agreements, educational certificates and voting cards.

All data on the blockchain needs the following attributes: Timestamp: the date and time that the data is written on the blockchain, Id: identification for that particular data, Type: the type of object the data represents in the real world (passport document, land deed certificate, etc) Government representative: the individual who is responsible to add that particular data.

There are lots of sectors where governments can use blockchain as a database system for example: for individual identity registration (including identification cards, passports, driving licenses...), for asset registration (including land deeds, vehicle titles...), for contract agreements, for voting cards, for educational certificates, for business license. All these databases are necessary for an effective operation of government services. And this study researches the benefits that could be gained by retaining these records with a blockchain based database system. Furthermore, this study will also explore whether a government can be even more effective if it developed one large centralized blockchain to link and gather data from all these sources together. Since the main advantage of using blockchain is to prevent data manipulation, this research will focus on government sectors in which data manipulation is a continuous issue.

2.1.3 Tokenization

In a blockchain based government, some original documents issued by the government can be tokenized. Meaning for those original documents, a digital copy can be generated and stored on the blockchain. This research will also investigate the benefits of using blockchain tokens for such purposes. Tokenization can be defined as the process of replacing

real world values (such as money, assets and other records) with tokens that reflect these values, making it easier and safer to use and trade them [17]. This research investigated the advantages of issuing tokens for keeping track of original documents, for the purpose of identity verification and for the purpose of asset ownership validation.

2.2 Related Work

Since the topic of blockchain is novel, the number of researches, books and journals written are fairly limited. There are even fewer works done that couple both blockchain technologies and government data management. There is no work done on the idea of a governmental use on a linked blockchain system that encompasses all data sectors (meaning a data management system that links identification database, passport database, driving licenses database, land deed database and other databases all together).

One of the very related works is titled “Blockchain for Digital Government” published in 2019 it was commissioned by the European Union. It began by defining a “Digital Government” as “a state-of-art paradigm in public administration science” “with a focus on the provision of user-centric, agile and innovative public services”. It studied seven blockchain projects by governments of European countries. These projects include land title registration, academic certification, digital identity, pension system and voucher distribution for low-income residents. It describes different blockchain architectures and relates to their cost and benefits. After analyzing the seven blockchain projects, their architectures, their cost and benefits, it concluded that “Contrary to how it is often portrayed, blockchain, so far, is neither transformative nor even disruptive for the public sector. We have not observed the creation of new business models, the emergence of a new generation of services nor direct disintermediation of any of the public institutions involved in the provision of governmental functions. Truly transformative services which enable decentralized voting or civic governance without direct involvement of governments are missing from the current landscape.” [13]

The other related work is titled “Blockchain-Based Land Registration System: A Conceptual Framework” published in February 2022. It focuses on the on land registration and management with the blockchain technology. It is an exploratory research that concluded “a land registry combined with blockchain technology has the potential to truly revolutionize governance.” It also concluded that such systems have the potential to become highly secure and private. [14]

Another related work is titled “Blockchains for Government: Use Cases and Challenges” published in November 2020. It focuses on blockchain implementation of governments around the world regarding use cases and challenges in medical, financial, infrastructural, educational and asset management sectors. It concludes by stating despite the technical challenges, “such as being fully privacy preserving, ensuring compliance when necessary, and being scalable, have yet to be fully solved, and more work is needed to address them” blockchain is the “best technology to deploy when a need to distribute data through a system that needs to guarantee data integrity and service availability exists, but the ability to make it happen is limited” which makes it ideal for government work.”[15]

These related research works discussed the benefits and challenges of using blockchain for government related works. And one must weigh the advantages and disadvantages of using such a system before applying them.

The related researches are mostly use cases, meaning they studied blockchain implementation of a specific government sector of a specific country, that employed case study and interview approaches to examine the main scenarios of application. They differ from this research in a way that this research is studying on building a generalized system that can work as a blueprint for any country to follow when building a blockchain based database system.

Chapter 3: Methodology

3.1 Sampling Methodology

Sampling was performed on two areas of this research: on government data sectors and on countries in the world.

3.1.1 Sampling of Government Data Sectors

A government may need a data management system for a variety of sectors. These sectors may include but are not limited to: identity sectors, citizen census, financial sectors (includes business licenses, tax history, employee records...), automobile sectors, property sectors, educational sectors (include the current educational status and previous educational certificates), legal sectors (include police records, court rulings, government policies, constitution, legislations, regulations, contractual agreement, document authentication and registration services), health care, infrastructure, security (criminal record, CCTV video surveillance, traffic management...), defense (military and intelligence data for own or allied nations), environmental, economics, utilities (electric, water, telecommunication...)... When starting this research the first work was to select the data sectors that can benefit from a blockchain based database system.

The sampling method used to choose between the data sectors is purposive sampling. Purposive sampling method is applied when the researcher is using their expertise to select a sample that is most useful to the purpose of this research. And, since the purpose of this research is to use blockchain to deal with identity fabrication and document forgery, only related data sectors are chosen.

This research will consider the following three government data sectors: Identity sectors (includes identity cards, passports, voting card registration, birth certificate, death certificate, relationship status...), Automobile sectors (includes driving licenses and vehicle ownership registration), Property sectors (includes land deed registration).

3.1.2 Sampling of Countries in the World.

One of the objectives of this research is to propose a blueprint for a government database system and the first question that comes to mind is which government. There are close to 195 countries in the world and the goal of this research is to provide a generalized system that any of these countries can adapt to their system. In order to build a blockchain based government database system the first work is to analyze what the data is and since it is impractical to analyze every country's system sampling of countries must be performed.

The first sampling method used that satisfies this area is stratified sampling. Stratified sampling works by first dividing the population into subgroups who all share similar characteristics and then picking equal sample size from each subgroup for further analysis. In the context of this research the population are the countries and after they are divided into a subgroup one country will be picked from each subgroup because it is redundant to investigate more than one country if they share similar characteristics.

On a closer examination of all the countries in the world, it is understood that most countries have similar data keeping techniques. This is because most countries choose to adapt their system from other countries. Throughout history most countries started building their system on their own when they gained independence from colonization but they adapted it from their colonizers [18]. Out of 195 countries in the world 153 countries were colonies (and protectorate in some cases), and 11 countries were colonizers. Assuming most colonies adapted their system from their colonizers, it will significantly reduce the sample pool by 153 countries.

Some other countries started building their system when their union was dissolved (for example Dissolution of the Soviet Union in 1991) and adapted it from their host countries. 18 countries were found that fulfill this criteria.

The close examination of each of the countries can be found on a google spreadsheet with sheet label 'Stratified Sampling', saved with a viewer only access link:<https://docs.google.com/spreadsheets/d/1x1xDqmufhJs2R7TzST09t2gGGSZOxoq-aNx>

[13nx8J_Q/edit#gid=1636470947](#)[25] and a reduced version of the spreadsheet is tabulated in Table 3.1.

Table 3.1: Reduced version of stratified sampling on a google spreadsheet

No.	Country	Reason for exclusion	Included/Not	
1	Afghanistan	Former British Colony	Excluded	0
2	Albania	Former Italy Colony	Excluded	0
3	Algeria	Former French Colony	Excluded	0
.
.
.
16	Belarus	Former Soviet Union	Excluded	0
17	Belgium		Included	1
.
.
.

Using the stratified sampling methodology the 195 countries are reduced to twenty four: Belgium, China, Denmark, Ethiopia, Finland, France, Germany, Greece, Greenland, Hungary, Italy, Liechtenstein, Monaco, Mongolia, Netherlands, Norway, Palau, Portugal, Russia, San Marino, Spain, Sweden, Switzerland and the United Kingdom. A pictorial representation of the selected countries using stratified sampling is shown in Figure 3.1. The countries expressed in green are the selected countries and the ones expressed in red are not. The picture is generated from the previously stated table (Table 3.1) using the ‘Geo chart’ feature of google spreadsheet.

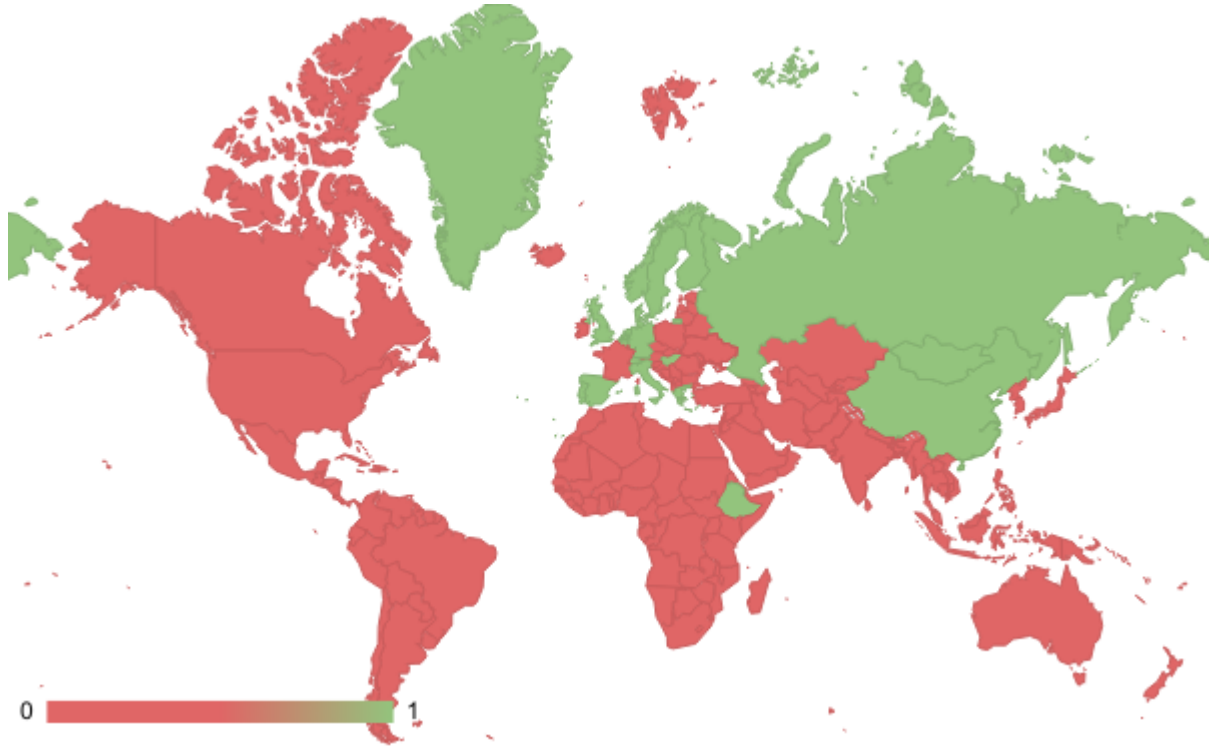


Figure 3.1: A pictorial representation of the selected countries using stratified sampling

After stratified sampling was performed, the convenience sampling method was used to further reduce the sample size because of the inconvenience of studying the selected data sectors for those few countries. Convenience sampling works by including individuals who happen to be most accessible to the researcher.

In the context of this research the convenience is based on the accessibility of a country in terms of data gathering and the impact of a country based on population. Since the countries in the European Union have similar economic, social and security structure with the exception of the United Kingdom [24], countries Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Portugal, Spain and Sweden can be represented with one country and France was chosen for convenience. And since countries with a population less than ten million people are generally assumed to have less

impact on the world, countries Greenland, Liechtenstein, Monaco, Mongolia, Norway, Palau, San Marino and Switzerland are not considered.

The close examination of each of the countries can be found on a google spreadsheet with sheet label ‘Convenience Sampling’, saved with a viewer only access link: https://docs.google.com/spreadsheets/d/1xlxDqmufhJs2R7TzST09t2gGGSZOxoq-aNx13nx8J_Q/edit#gid=941066011[25] and a reduced version of the spreadsheet is tabulated in Table 3.2.

Table 3.2: Reduced version of convenience sampling on a google spreadsheet

No.	Country	Population (2022)	Reason for exclusion	Included/Not	
1	Belgium	11,655,930	Part of the EU	Excluded	0
2	China	1,425,887,337		Included	1
3	Denmark	5,882,261	Part of the EU	Excluded	0
4	Ethiopia	123,379,924		Included	1
.
.
.

After applying the convenience sampling method on the twenty four countries listed above only five of them were chosen: China, Ethiopia, France, Russia and the United Kingdom. A pictorial representation of the selected countries using convenience sampling is shown in Figure 3.2. The countries expressed in white are countries deselected using stratified sampling method, the countries expressed in green are the selected countries and the ones expressed in red are not. The picture is generated from the previously stated table (Table 3.2) using the ‘Geo chart’ feature of google spreadsheet.

3.2 Data Gathering

The method used to collect the data for this thesis is the document analysis method. Document analysis is a method of data collection which involves analysis of content from written documents in order to make certain deductions based on the study parameters. The method is mainly used in qualitative research as a method of qualitative analysis. One of the advantages of document analysis is that it allows researchers to access a variety of documents, such as public records, personal documents, and physical evidence, that may contain relevant and rich information for their studies.

For this research document analysis method is used since all governments track individual identity by providing their citizens with original documents, the best way to understand the data they deal with is to collect these documents first. Therefore, the data gathered for the research was the documents that were prepared by different countries.

Data gathering was conducted by collecting templates of these documents from different online sources for the selected countries. Then all the data referenced on these documents was collected for further study.

The close examination of the gathered data can be found on a google spreadsheet with sheet, saved with a viewer only access link: https://docs.google.com/spreadsheets/d/16hSoNVIOIfhwNeyLVjOUwlZ10Ax_mzuURSD0mB9OWA4/edit#gid=0[26] and a reduced version of the spreadsheet is tabulated on Table 3.3. The data gathered from an online source, the website link is indicated on the google spreadsheet. The google spreadsheet also shows the translation methods used for some of the documents that are not in English.

Table 3.3: Reduced version of the data identity gathered on a google spreadsheet

Country	Document Type	Document Image (Image is attached using google sheet's image inserting function "=IMAGE(_)")	Data in Question (Translation was performed using google sheet's translate function "=GOOGLETRANSLATE(_)")	Data type
China	Passport (All passport data in two languages english and chinese)	(IMAGE)	Type Name Country Code ...	String String String ...
.
.
.

For identity sectors, these documents were; passports, national identification, birth certificate, marriage certificate and death certificate documents. A total of twenty five documents were collected from five different countries. The data of interest were extracted from these documents, resulting in five hundred five data entries. The data can be found on a google sheet with the label "Identity." While 92.1% of the data type are String, the rest 7.9% of the data type are Images like shown on Figure 3.3.

Count of Data Types from Identity Sector

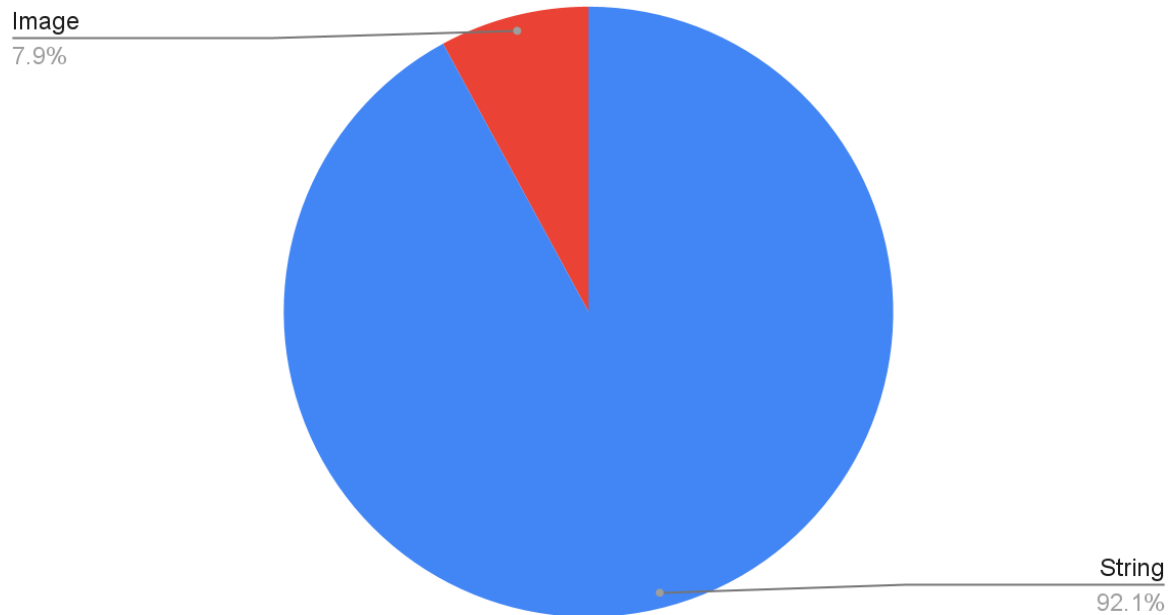


Figure 3.3: Count of Data Types from Identity Sector

For automobile sectors, these documents were; driving licenses and vehicle ownership certificates. A total of ten documents were collected from five different countries. The data of interest were extracted from these documents, resulting in two hundred thirty four data entries. The data can be found on a google sheet with the label “Automobile.” While 92.7% of the data type are String, the rest 7.3% of the data type are Images like shown on Figure 3.4.

Count of Data Types from Automobile Sector

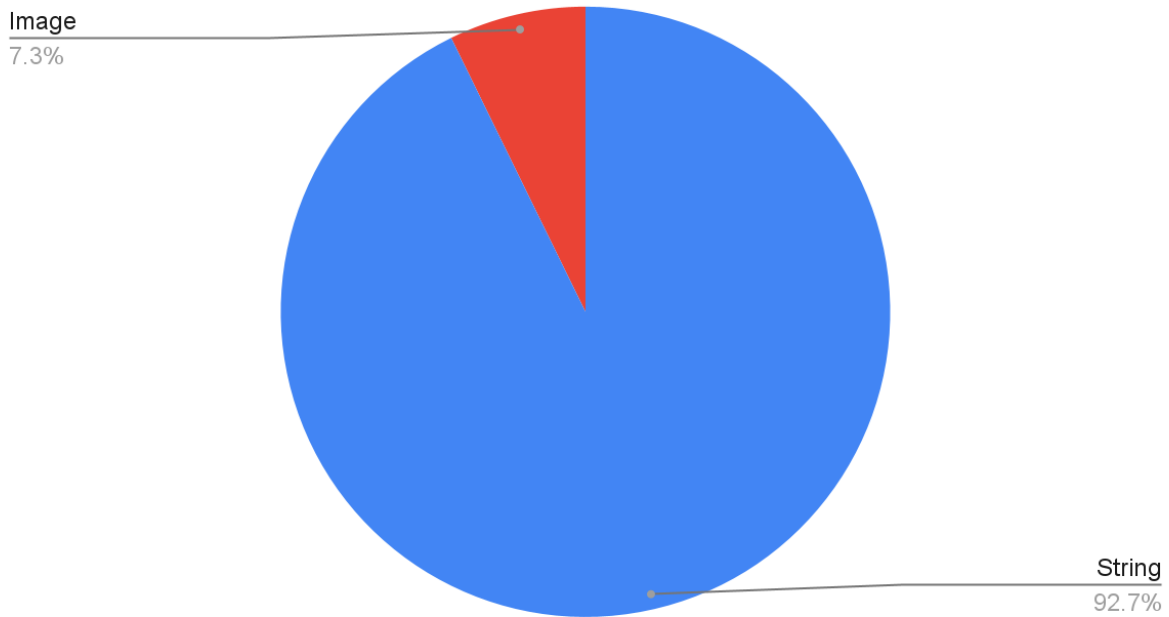


Figure 3.4: Count of Data Types from Automobile Sector

For property sectors, these documents were; land deed certificates (titles). Five documents were collected from five different countries. The data of interest were extracted from these documents, resulting in eighty-three data entries. The data can be found on a google sheet with the label “Property.” While 91.6% of the data type are String and 6.0% of the data type are Images, the rest 2.4% of the data type are Vector like shown on Figure 3.5.

Count of Data Types from Property Sector

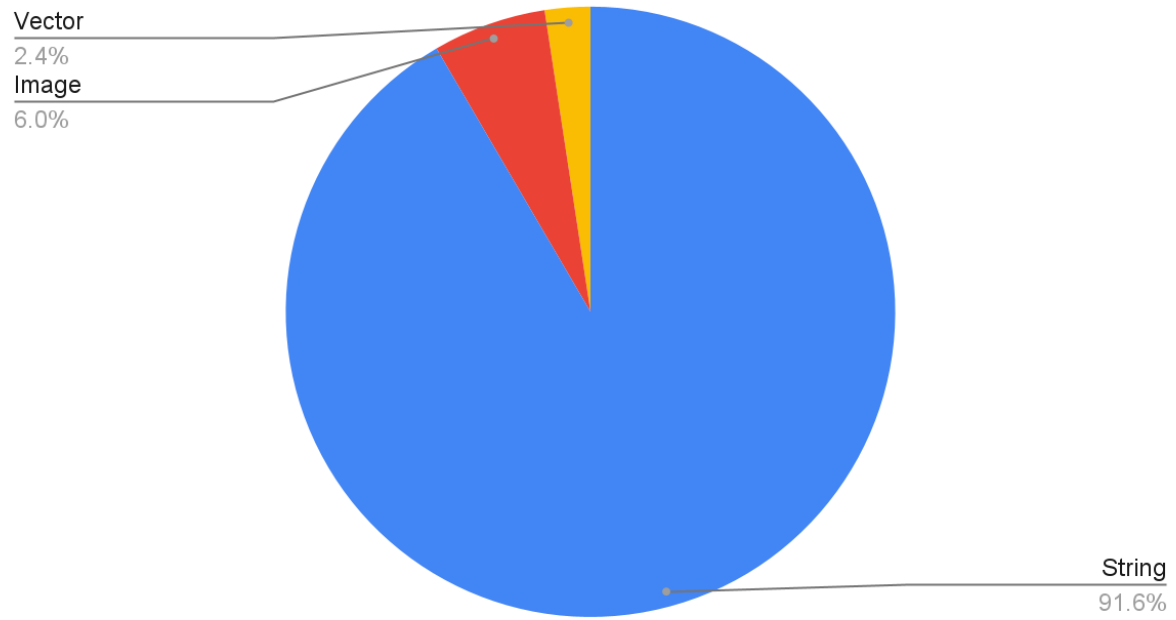


Figure 3.5: Count of Data Types from Property Sector

Chapter 4: Data Analysis

Data analysis was performed on the gathered documents of the five selected countries. The data analysis was performed on a google spreadsheet with access link: https://docs.google.com/spreadsheets/d/16hSoNVIOIfhwNeyLVjOUwlZ10Ax_mzuURSD0mB9OWA4/edit#gid=0[26] and a shortened version of the spreadsheet is tabulated on Table 3.3.

A deductive approach was used for the data analysis. Deductive analysis generally means applying theory to the data to test the theory. It's a kind of “top-down” approach to data analysis. In qualitative analysis, this often means applying predetermined codes to the data. This approach was selected to check the theory that all the data gathered from the five countries and from the three government sectors can be represented on the blockchain.

From the collected eight hundred twenty two data points, ten different types of data were found: words, numbers, dates, qr-codes, bar-codes, photos, stamps, seals, signatures and a map. The next step is to find appropriate data types to store these ten different types of data.

The words, numbers and dates can be stored as a string on the blockchain by using the Unicode Standard to represent multiple different languages. Strings are the ideal choice for most of these cases because they are designed to store human-readable texts. The qr-codes and bar-codes can also be converted and stored to a string data type.

The photos, stamps, seals, and signatures can be stored in the form of images on the blockchain using two methods. The first method is to compress and store the binary data of the images directly on the blockchain. Although this approach might somewhat slow down the blockchain network during retrieval, the images will never be lost or tampered with. The other method is to place the images on a conventional database and store both the link and hash value of the image on the blockchain. This approach might result in a faster blockchain network and unmodifiable storage but the images will be susceptible to deletion. Since the goal of this research is to utilize blockchain technology for its advantageous features, the first method is the acceptable one.

The maps can also be represented using two methods. The first method is to convert them into pictures and retain them in the form of images. The second method is to convert them into geometrical vector image formats and retain the vector data. Retaining the vector data is the preferable method because it results in faster network and precise representation.

To summarize the data findings, from the collected five hundred five data entries of the identity sector, four hundred sixty five of them were string data and the rest forty were image data. From the collected two hundred thirty four data entries of the automobile sector, two hundred seventeen of them were string data and the rest seventeen were image data. From the collected eighty three data entries of the property sector, seventy-six of them were string data, five were image data and the rest two were vector data. Figurative summarization is shown on Figure 4.1 [26].

Count of Data Types

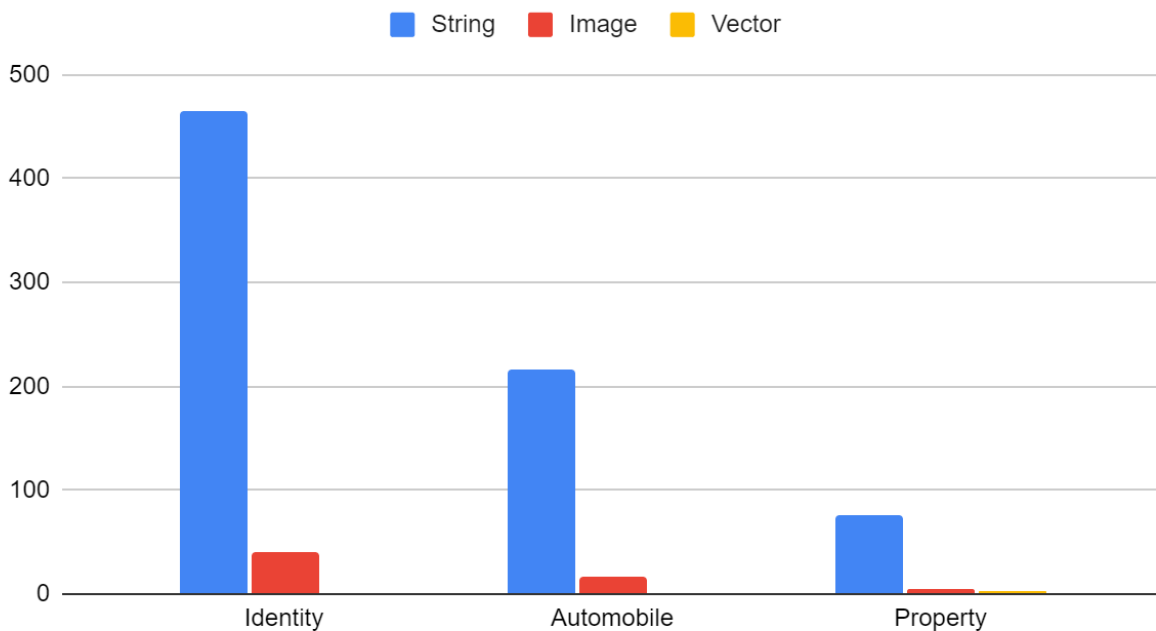


Figure 4.1: Count of Data Types in Identity Automobile and property sectors

This proves that data collected concerning data sectors can be adequately represented on the blockchain.

Chapter 5: Solution of the Research

One of the objectives of this research is to propose a blueprint as a solution for the research problem. First the blockchain system must be developed which is the main database system that contains the data. Different government sectors require different data but in a blockchain system there are certain data that shall be included. Second, the infrastructure must be built which is concerned with the servers and backup servers. Lastly, a document verification system must be built in which users can verify the authenticity of a government supplied document.

5.1 Building the System

One of the key aspects of creating a blockchain network is to ensure that the data, or the blocks, are connected in a sequential manner. This means that each new block must include the hash value of the previous block as part of its data. The hash value is a unique identifier that depends on the content of the block. Therefore, if someone tries to alter the data in a block, they would also have to change its hash value, which would affect the next block in the chain, and so on. This makes it very difficult to tamper with the blockchain data.

A proof of concept was built that illustrates such a system using reactjs framework for the user interface and firebase for the backend. The website can be found at <https://blockchain-gov-sys.web.app>[27] and the source code can be found at <https://github.com/TsegaDEBEBE/blockchain-gov-sys> [28].

For the simplicity of the thesis it is assumed that every individual document was a token and one token was placed on each block. Which entails every document will have a token-id that is the hash value of the token or the hash value of the block. Every token has a token-id, token type, token issuer, timestamp, previous block's hash value and main data. The details inside the main data differs between different types of tokens.

The proof of concept has three features; token adding, token verifying and token showing. The token adder for an identity token is shown on the Figure 5.1 and can be found on <https://blockchain-gov-sys.web.app/add> [27]. In addition to the identity token adder; passport token, birth certificate token, marriage certificate token, driving license token, vehicle registration token and land deed token adder were built.

The use of Blockchain for Government Data Management

Identity Card Token




TOKEN TYPE	identity
FULL NAME	
NATIONALITY	
GENDER	
DATE OF BIRTH	mm/dd/yyyy 
DATE OF ISSUE	mm/dd/yyyy 
PLACE OF BIRTH	
DATE OF EXPIRY	mm/dd/yyyy 
PHOTO	<input type="button" value="Choose File"/> No file chosen
TOKEN ID OF GOVERNMENT REPRESENTATIVE	9d6200de4e0a8d1f08beba0265dad5ac
CURRENT TIME STAMP	Thu Apr 20 2023 01:24:39 GMT+0300 (East Africa Time)
<input type="button" value="Add Token"/>	
<input type="button" value="Go Back"/>	

Figure 5.1: Adding an identity token

The identity token has the following structure:

```
{
  "token id": "alkdsfjalsdfnladskfja", //hash for current token and block
  "type": "identity", //token type
  "name": "XXXXXX XXXXX XXXXX",
  "nationality": "XXXXXXXXXXXXXXXXXX",
  "gender": "X",
  "date of birth": "XX/XX/XXXX",
  "date of issue": "XX/XX/XXXX",
  "place of birth": "XX XXXX",
  "date of expiry": "XX/XX/XXXX",
  "photo": BLOB,
  "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
  that issued the token
  "timestamp": XXXXXXXX, //automatically recorded the moment the token is issued
  "previous hash": "ds35fads5f6aes4r6w8e5", //hash for previous block
}
```

The passport token has the following structure:

```
{
  "token id": "alkdsfjalsdfnladskfja", //hash for current token and block
  "type": "passport", //token type
  "birth": "a5d4f16ds54f6sdsdf54a", //token id for birth certificate
  "identity-card": "6w5e4f2d0cads56fw8ef4", //token id for national identity card
  "marriage": "2a30sd1fsa56d4fasd5f1", //token id for marriage certificate
  "type": "P",
  "country code": "XXX",
  "passport id": "XXXXXXXXXX",
  "name": "XXXXXX XXXXX XXXXX",
  "nationality": "XXXXXXXXXXXXXXXXXX",
  "gender": "X",
  "date of birth": "XX/XX/XXXX",
  "date of issue": "XX/XX/XXXX",
  "place of birth": "XX XXXX",
  "date of expiry": "XX/XX/XXXX",
  "signature": BLOB,
  "photo": BLOB,
}
```

```

    "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
    that issued the token
    "timestamp": XXXXXXXX, //automatically recorded the moment the token is issued
    "previous hash": "ds35fads5f6aes4r6w8e5", //hash for previous block
}

```

The birth certificate token has the following structure:

```

{
    "token id": "a5d4f16ds54f6sdsdf54a", //hash for current token and block
    "type": "birth",
    "father identity card": "fa2s4dfasdfadsf54a6sd", //father's token id from his identity card
    "mother identity card": "5f4sad651f4a6sd5f4a6s", //mother's token id from her identity card
    "name": "XXXXX XXXXX XXXXX",
    "nationality": "XXXXXXXXXXXXXXXXXX",
    "gender": "X",
    "date of birth": "XX/XX/XXXX",
    "date of issue": "XX/XX/XXXX",
    "place of birth": "XX XXXX",
    "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
    that issued the token
    "timestamp": XXXXXXXX, //automatically recorded the moment the token is issued
    "previous hash": "46ads5f2ds023oiewkdkp", //hash for previous block
}

```

The marriage certificate token has the following structure:

```

{
    "token id": "a5d4f16ds54f6sdsdf54a", //hash for current token and block
    "type": "marriage",
    "husband identity card": "fa2s4dfasdfadsf54a6sd", //husband's token id from his identity card
    "wife identity card": "5f4sad651f4a6sd5f4a6s", //wife's token id from her identity card
    "date of marriage": "XX/XX/XXXX",
    "date of issue": "XX/XX/XXXX",
    "husband photo": BLOB,
    "wife photo": BLOB,
    "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
    that issued the token
    "timestamp": XXXXXXXX, //automatically recorded the moment the token is issued
    "previous hash": "46ads5f2ds023oiewkdkp", //hash for previous block
}

```

The driving license token has the following structure:

```

{
  "token id": "vnq2r0358qr98wes4d5d5", //hash for current token and block
  "type": "driving",
  "name": "XXXXXX XXXXXX XXXXXX",
  "date of issue": "XX/XX/XXXX",
  "date of expiry": "XX/XX/XXXX",
  "function": "XXXX",
  "code": "XXXX",
  "signature": BLOB,
  "photo": BLOB,
  "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
  that issued the token
  "timestamp": XXXXXXXXXX, //automatically recorded the moment the token is issued
  "previous hash": "46ads5f2ds023oiewkdkp", //hash for previous block
}

```

The vehicle registration token has the following structure:

```

{
  "token id": "vnq2r0358qr98wes4d5d5", //hash for current token and block
  "type": "vehicle",
  "driver identity card": "6w5e4f2d0cads56fw8ef4", //driver's token id from his/her identity card
  "residence": "XXXXXXXXXX",
  "date of issue": "XX/XX/XXXX",
  "vehicle identification no": "XXXX",
  "brand": "XXXX",
  "model": "XXXX",
  "color": "XXXX",
  "engine number": "XXXX",
  "engine model": "XXXX",
  "engine power": "XXXX",
  "weight": "XXXX",
  "issuing authority": "XXXX",
  "issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
  that issued the token
  "timestamp": XXXXXXXXXX, //automatically recorded the moment the token is issued
  "previous hash": "46ads5f2ds023oiewkdkp", //hash for previous block
}

```

The land deed certificate token has the following structure:

```

{
  "token id": "vnq2r0358qr98wes4d5d5", //hash for current token and block
  "type": "land",
  "owner identity card": "6w5e4f2d0cads56fw8ef4", //owner's token id from his/her identity card
}

```

```

"land deed certificate id": "XXXXXXXXXX",
"location": "XXXXXX XXXXXX XXXXXX XXXXXX",
"area": "X",
"type of land": "XXXX",
"purpose of land": "XXXX",
"date of issue": "XX/XX/XXXX",
"issuer": "a3s2d1f0a3sd5fcas5dfa", //token id for identity card of the government representative
that issued the token
"timestamp": XXXXXXXXX, //automatically recorded the moment the token is issued
"previous hash": "46ads5f2ds023oiewkdkp", //hash for previous block
}

```

Tokenid

The tokenid is the hash value of the token. Every piece of data on a token will first be arranged into an ascending order before converting them into a string. Then the string will be hashed using the hashing algorithm, such as SHA-256 function, and the resulting hash string will be the unique id of the token. Every block must contain the tokenid which is later used to verify if the data is altered in any way.

Type

The type attribute describes the type of the token. Every block must contain this attribute.

Issuer

The issuer attribute describes the government representative that issued the token. Before the government employee issues a token they must first login to the system with their credentials. Then the system will automatically link their identity token to every token they

issue which will be assumed to be their signature. This discourages government employees from participating in fraudulent cases because their identity token id will be forever stored on an immutable ledger as long as the system is operational. Every block must contain this attribute.

Timestamp

A timestamp is a piece of data that indicates when a certain event or document was created or modified. In blockchain, timestamps are used to record the date and time of each block. Timestamps help to ensure the integrity and security of the blockchain, as they prevent tampering and double minting with the data.

The moment a data is stored on a blockchain a snapshot will be taken of the timestamp. And the timestamp together with the rest of the data is hashed and stored on the blockchain. If that same data is minted on another moment the hash value will be completely changed and thus prevents tampering and double minting.

Previous Hash

This attribute simply contains the hash value of the previous token or block thus ensuring the blockchain link. Every block must contain this attribute.

Main Data

The main data contains all the details about that particular token; such as name, gender and photo. While tokens of the same type contain similar data structure, different token types contain different data types of data, meaning while it is mandatory that all passport tokens contain a data entry for country-code, other tokens like identity token will not include such data.

Add Token Button

Like the name specifies, the add token button adds the current token to the blockchain after the data has been filled. This is done by first taking a snapshot of the data in question including the particular timestamp, then ordering the data in ascending order, then converting data into a string and finally hashing the data. All the data objects and hash string will finally be stored on the database. The hash string can be used as a unique token-id for the recently added token and a file name for that particular data. The hash string is necessary in later identifying if the token data is tampered with in any way.

5.2 Document Verification

Document verification is performed using smart contracts. Smart contracts are self-executing programmes that encode the rules and logic of the blockchain. They can automate processes, enforce compliance, and reduce errors and fraud.

Smart contracts must be built to automate the process of document verification. Document verification must be performed at multiple levels of data process; when a new data is added, during an audit of the system and when a document is being retrieved.

Adding new data

When a new data is added, each document is verified by the network of nodes (servers) that run the blockchain protocol, and then added to a block of data that is linked to the previous block, forming a chain.

To verify a block in a blockchain, a node must first check that the block has a valid format and structure. Then it must check that the block's timestamp is within a reasonable range of the current time and the previous block's timestamp.

If all these checks pass, then the node can accept the block as valid and append it to its local copy of the blockchain. The node can then broadcast the block to other nodes in the network, or request new blocks from other nodes if it is not up to date with the latest block.

Audit

An audit is an automated program that runs on the server nodes and it can be random, scheduled or triggered by some events. It usually starts from the latest block and verifies if the block link has not been broken and the data on every block has not been tampered with.

Retrieving a document

The key area of this research is verifying a document during retrieval. But first the system will need to have a blockchain explorer. The government will need to develop some type of a web-based system so that any one, such as a government employee, private

company or a civilian, can verify a document. Whenever a document needs to be verified, the user could simply write the token id and search for it in the blockchain explorer.

An instance of a blockchain explorer is shown on Figure 5.2 and a working proof of concept can be found on <https://blockchain-gov-sys.web.app/verify> [27].



The image shows a web application interface with a light blue background. At the top center is the HiLCoE logo, which consists of a stylized 'N' inside a circle, followed by the text 'HiLCoE' and 'School of Computer Science & Technology' below it. Below the logo, the title 'The use of Blockchain for Government Data Management' is displayed in a large, black, sans-serif font. Underneath the title is a white rectangular input field containing the text 'TOKEN-ID' and a long alphanumeric string '6f35f7a8f5e8ca84b4c113deab761c6a'. To the right of the input field is a white rectangular button with the word 'Search' in black text.

Figure 5.2: Using the Blockchain Explorer to verify a token

After the blockchain explorer retrieves the document there will be an automated program that checks if the data has been tampered with. The program will calculate the hash value of the retrieved document and check it against the token id, which is the original

hash value of that document. This document verification system will reduce frauds caused by document fabrication.

An instance of a token retrieval is shown on Figure 5.3 and Figure 5.4 and a working proof of concept can be found on <https://blockchain-gov-sys.web.app/verify/6f35f7a8f5e8ca84b4c113deab761c6a>[27]. Figure 5.3 shows the verification passed because the data was not altered with.



Figure 5.3: Verification Passed on an Identity token

Figure 5.4 shows the verification as failed because the data was altered, the year on the date of expiry was increased by one year, which changes the hash value of the original document.



Figure 5.4: Verification Failed on an Identity token

5.3 Connecting Data Sectors

Building one centralized blockchain system that contains the data from different gives a new advantage by linking different government data sectors together. Meaning, if a

government body needed to identify to whom a land certificate token belongs to or who the parents are on a birth certificate token, their corresponding identity token can be found linked to their respective tokens. This system will streamline government related processes and tiresome bureaucratic procedures for the public and civil servants.

Figure 5.5 shows a visualization of linked tokens starting from a marriage token then listing every token that relates to it [29].

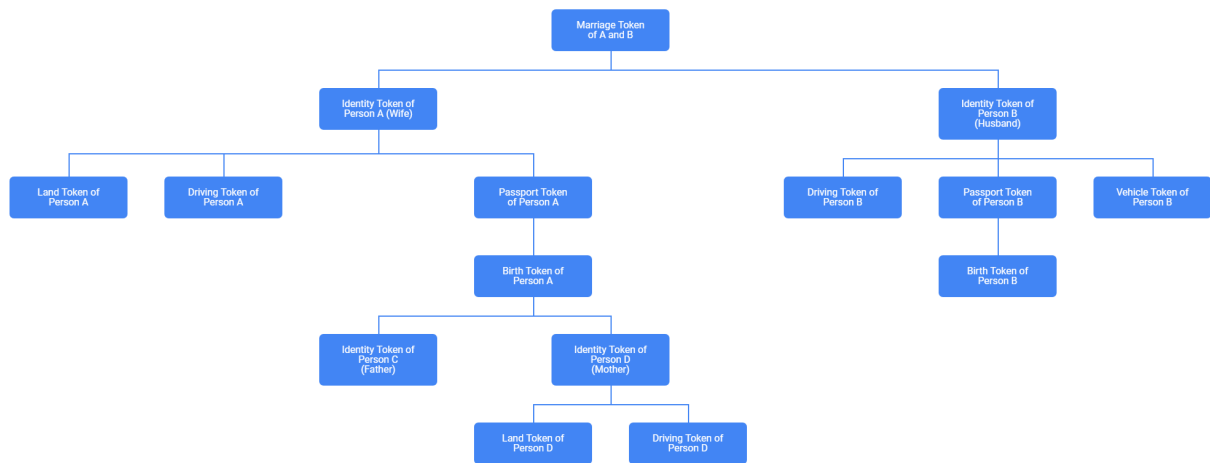


Figure 5.5: A visualization of connected tokens from different data sectors

Chapter 6: Conclusion and Recommendation

6.1 Conclusion

In conclusion, blockchain technology has the potential to revolutionize government related databases by providing a secure, transparent, and efficient way of storing and sharing data among multiple sectors.

After a thorough examination of the three data sectors of a government, identity sectors, automobile sectors and property sectors, it is evidence that a blockchain based database system can be used to store such documents.

Blockchain provides secure storage of sensitive government and citizen data. Blockchain can store data in a distributed network of computers, making it more difficult for hackers or unauthorized parties to access or manipulate the data.

Blockchain provides digitalization and reduction of labor-intensive processes. Blockchain can automate and streamline administrative processes that involve multiple agencies or stakeholders, such as identity management, vehicle registration management, land deed management or other citizen services. This can reduce costs, errors, and delays associated with paper-based or manual systems.

Blockchain can increase transparency and prevent corruption cases. Blockchain can provide an immutable audit trail of all activities that take place on the network. This can increase accountability and trust between the government and its citizens, and deter any attempts to tamper with or falsify the data.

Blockchain provides enhanced efficiency and automation of administrative processes. Blockchain can enable smart contracts, which are self-executing agreements that are triggered by predefined conditions. This can facilitate document verification.

6.2 Recommendation

When implementing blockchain for a government database the first step is to understand the use case for the particular government sector. All government sectors have their own registration, verification, retrieval policies and privacy issues. When building such systems the following questions must be answered. What data should each sector include? How to represent each data? Which programming language should one choose? What hashing algorithm should the system employ?

Then choose the appropriate blockchain platform and architecture. Depending on the use case and the requirements, different types of blockchain platforms may be suitable. The government can develop its own blockchain from scratch or use an existing blockchain and start from there. There are many public blockchain projects, such as Ethereum, Solana, TON..., each with their own property, such as speed, performance and scalability. These blockchain projects are open source so that any government or private organization can copy their source code and use it in any way they see fit.

Then design and develop the smart contracts and applications. Smart contracts are self-executing agreements that are triggered by predefined conditions. Applications are the user interfaces that allow stakeholders to interact with the blockchain network and access the data stored on the ledger. They can be web-based, mobile-based, or integrated with existing systems.

Then test and deploy the blockchain solution. Before launching the blockchain solution to the public, it is essential to test its functionality, performance, security, and usability. Testing can be done in a simulated environment or a pilot project with a limited number of participants. Once the testing is successful, the blockchain solution can be deployed to the target audience and integrated with other government systems.

Lastly, monitor and maintain the blockchain network. After deployment, the blockchain network needs to be monitored and maintained to ensure its reliability,

availability, and security. This may involve updating the software, resolving technical issues, managing access rights, auditing transactions, and evaluating user feedback.

Chapter 7: Future Work

The most important part of a database system is the infrastructure that enables it, such as the servers and the backup servers. On a blockchain based system, the servers are called nodes. They are interconnected to each other and communicate through a peer-to-peer protocol. By having a copy of the blockchain stored on multiple nodes, the network becomes more robust and resistant to malicious attacks. If one node is compromised or offline, the others can still maintain the integrity and availability of the data on the blockchain. There are different types of nodes in a blockchain network, such as full nodes and light nodes. Full nodes store a complete copy of the blockchain ledger and validate every transaction and block according to the consensus rules of the network. Light nodes only store a fraction of the data and rely on full nodes to verify transactions. Infrastructure related issues, such as server type, server size, connection type, power supply, number of full nodes and light nodes, are not discussed in this thesis and they are left for future work.

Even though transparency is one of the positive features of blockchain, when we implement it for governmental databases, privacy issues could emerge. To help mitigate this issue, encrypting the data on the blockchain will be necessary. Encrypting governmental databases is a crucial measure to protect the sensitive and confidential information of the citizens and the state from unauthorized access, theft, or tampering. The techniques of encrypting data on a blockchain are not discussed in this thesis and they are left for future work.

The other feature of a blockchain is its immutable record keeping property. Meaning once a data is placed on a blockchain it is remarkably hard to modify it. The government employees tasked to add data on the blockchain could make spelling errors and if these errors are placed on the blockchain it becomes a problem. Therefore the government must build a system to handle such issues. The techniques of amending data or invalidating a document on a blockchain are not discussed in this thesis and they are left for future work.

Because of some limitations in resources this research only studies identity, automobile and property sectors of a government but the assumption is this technology can apply to many other government sectors. If this assumption is proved to be correct, then advantages and application of connecting different government data sectors together will be immeasurable. An institute with no limitations should examine this approach to handling government data and discover its complete capabilities.

Using NFT for government issued original documents. NFTs, or non-fungible tokens, are digital assets that can prove ownership and authenticity of a unique asset on a blockchain ledger. NFTs can be used for various purposes, such as art, music, gaming, and sports. But they may also be used for government issued original documents, such as birth certificates, passports, or driver's licenses. For example, the government could create an NFT that represents an individual's identity document, such as a passport, and store it on a secure wallet. The NFT could contain encrypted personal information, such as name, date of birth, nationality, biometric data, and a digital signature. The individual could then use the NFT to prove their identity to various entities, such as banks, airlines, or government agencies, by scanning a QR code or using a smart contract. However, using such a system might have legal, technical and ethical challenges and risks which requires an extensive study. Therefore techniques of using NFTs are not discussed in this thesis and they are left for future work.

Reference

- [1] I. Zikratov, A. Kuzmin, V. Akimenko, V. Niculichev and L. Yalansky, "Ensuring data integrity using blockchain technology," 2017 20th Conference of Open Innovations Association (FRUCT), 2017, pp. 534-539, doi: 10.23919/FRUCT.2017.8071359.
- [2] Abdulai, Raymond & Ochieng, Edward. (2017). Land registration and landownership security: An examination of the underpinning principles of registration. *Property Management*. 35. 24-47. 10.1108/PM-09-2015-0051.
- [3] Gupta, Vijaya. "Contract Alteration: a detailed study of its effects." *iPleaders*, 3 August 2020, <https://blog.ipleaders.in/contract-alteration-detailed-study-effects/>. Accessed 10 August 2022.
- [4] United Nations Economic Commission for Europe. "Study on the Challenges of Fraud to Land Administration Institutions." vol. 1, no. 1, 2011, p. 51. ECE/HBP/165.
- [5] Dongre, Jayesh. (2020). Education Degree Fraud Detection and Student Certificate Verification using Blockchain. *International Journal of Engineering Research and*. V9. 10.17577/IJERTV9IS070156.
- [6] Organization for Economic Co-operation and Development. "Identity Fraud: Tax Evasion and Money Laundering Vulnerabilities." 2006, p. 17. <https://www.oecd.org>, <https://www.oecd.org/tax/exchange-of-tax-information/42223740.pdf>. Accessed 15 August 2022.
- [7] Levitt, Justin. (2007). The Truth About Voter Fraud. *SSRN Electronic Journal*. 10.2139/ssrn.1647224.
- [8] Nakamoto, Satoshi. (2008) Bitcoin: A Peer-to-Peer Electronic Cash System. <https://bitcoin.org/bitcoin.pdf>. Accessed 01 July 2022.
- [9] Haber, S., Stornetta, W.S. How to time-stamp a digital document. *J. Cryptology* 3, 99–111 (1991). <https://doi.org/10.1007/BF00196791>
- [10] H. Stevens, "Hans Peter Luhn and the birth of the hashing algorithm," in *IEEE Spectrum*, vol. 55, no. 2, pp. 44-49, February 2018, doi: 10.1109/MSPEC.2018.8278136.

- [11] H L, Gururaj & Athreya, A & Kumar, Ashwin & Holla, Abhishek & Nagarajath, S & Kumar V, Ravi. (2020). BLOCKCHAIN. 10.1002/9781119621201.ch1.
- [12] Forsström, Stefan, "Blockchain Research Report", December 2019, p. 12.
- [13] Allessie, D., Vaccari, L., & Sobolewski, M. (2019). Blockchain for Digital Government: An Assessment of Pioneering Implementations in Public Services (F. Pignatelli, Ed.). Publications Office of the European Union.
<https://joinup.ec.europa.eu/sites/default/files/document/2019-04/JRC115049%20blockchain%20for%20digital%20government.pdf>
- [14] CLAVIN, J., DUAN, S., ZHANG, H., JANEJA, V. P., JOSHI, K. P., YESHA, Y., ERICKSON, L. C., & LI, J. D. (2020). Article 22. Blockchains for Government: Use Cases and Challenges, 1(3), 21.
- [15] Khalid, M. I., Iqbal, J., Alturki, A., Hussain, S., Alabrah, A., & Ullah, S. S. (2022). Research Article. Blockchain-Based Land Registration System: A Conceptual Framework, 2022, 21.
- [16] Oracle. "A Guide to MySQL and Open Source in Government." *Government Technology Insider*, Oracle and/or its affiliates, 2013, <https://governmenttechnologyinsider.com/wp-content/uploads/2014/09/Acrobat-Document.pdf>, Accessed 1 September 2022.
- [17] Kharitonova, Anna. (2021). Capabilities of Blockchain Technology in Tokenization of Economy. 10.2991/aebmr.k.210318.006.
- [18] Ocheni, Stephen & Nwankwo, Basil. (2012). Analysis of Colonialism and Its Impact in Africa. 46-54. 10.3968/j.ccc.1923670020120803.1189.
- [19] Yusri Zakariah, Salfarina Samsudin & Norshafadila Ngadiman. "An Overview of the Fraud and Forgery Challenges in Land Registration System". *European Journal of Molecular & Clinical Medicine*, vol 7, issue 3, 2020, 274-282.
- [20] Low, R. (2008). The use of technology to automate the registration process within the Torrens system and its impact on fraud: An analysis. Queensland University of Technology,

- [21] Maidin, A. J., & Khadouf, H. A. (2009). Weaknesses in the Registration of Land Dealing System in Malaysia: Suggestion for improvement for enhancing the system. Law Review, 4.
- [22] Ismail, M. S. (2011). Measures undertaken to safeguard against fraud in land dealings. Jurnal Pentadbiran Tanah, 1(1), 85-99.
- [23] Brown, R. Crime prevention design in a vehicle registration system: a case study from Australia. Crime Sci 4, 25 (2015). <https://doi.org/10.1186/s40163-015-0038-1>
- [24] Liapis, Konstantinos & Rovolis, Antonis & Galanos, Christos & Thalassinou, Eleftherios. (2013). The Clusters of Economic Similarities between EU Countries: A View Under Recent Financial and Debt Crisis. European Research Studies Journal. 16. 41-66. 10.35808/ersj/380.
- [25] Tsega Debebe, [Sampling, https://docs.google.com/spreadsheets/d/1xIxQmufhJs2R7TzST09t2gGGSZOxoq-aNx13nx8J_Q](https://docs.google.com/spreadsheets/d/1xIxQmufhJs2R7TzST09t2gGGSZOxoq-aNx13nx8J_Q), Accessed 20th April 2023.
- [26] Tsega Debebe, [Data analysis, https://docs.google.com/spreadsheets/d/16hSoNVIOIfhwNeyLVjOUw1Z10Ax_mzuURSD0mB9OWA4/edit#gid=0](https://docs.google.com/spreadsheets/d/16hSoNVIOIfhwNeyLVjOUw1Z10Ax_mzuURSD0mB9OWA4/edit#gid=0), Accessed 20th April 2023.
- [27] Tsega Debebe, [Blockchain for Government System \(blockchain-gov-sys.web.app\), https://blockchain-gov-sys.web.app/](https://blockchain-gov-sys.web.app/), Accessed 20th April 2023.
- [28] Tsega Debebe [Blockchain for Government System \(blockchain-gov-sys.web.app\), https://github.com/TsegaDEBEBE/blockchain-gov-sys](https://github.com/TsegaDEBEBE/blockchain-gov-sys), Accessed 20th April 2023.
- [29] Tsega Debebe, [Connected Tokens, https://docs.google.com/spreadsheets/d/1sFStL2GXxhfzNzbPy0s-o7hBri17kqqakg90bv83aS4/edit#gid=0](https://docs.google.com/spreadsheets/d/1sFStL2GXxhfzNzbPy0s-o7hBri17kqqakg90bv83aS4/edit#gid=0), Accessed 25th April 2023.

Declaration Sheet

This thesis is my original work and has not been presented for a degree in any other university, and that all sources of material used for the thesis have been duly acknowledged.

Submitted by:

Tsega Debebe

May 2 2023


Student Name

Signature

Date

Approved by:

1. Mesfin Belachew (PhD)



May 1 2023

Advisor Name

Signature

Date

2. _____

Coordinator, Graduate

Signature

Date

Programme Research Office

3. _____

Director, Graduate

Signature

Date

Programme Office