future internet

MDPI

Article

Blockchain-Based Implementation of National Census as a Supplementary Instrument for Enhanced Transparency, Accountability, Privacy, and Security

Sana Rasheed * and Soulla Louca *

Digital Innovation Department, University of Nicosia, Nicosia 2417, Cyprus
* Correspondence: rasheed.s@live.unic.ac.cy (S.R.); louca.s@unic.ac.cy (S.L.)

Abstract: A national population census is instrumental in offering a holistic view of a country's progress, directly influencing policy formulation and strategic planning. Potential flaws in the census system can have detrimental impacts on national development. Our prior research has pinpointed various deficiencies in current census methodologies, including inadequate population coverage, racial and ethnic discrimination, and challenges related to data privacy, security, and distribution. This study aims to address the "missing persons" challenge in the national census population and housing system. The integration of blockchain technology emerges as a promising solution for addressing these identified issues, enhancing the integrity and efficacy of census processes. Building upon our earlier research which examined the national census system of Pakistan, we propose an architecture design incorporating Hyperledger Fabric, performing system sizing for the entire nation count. The Blockchain-Based Implementation of National Census as a Supplementary Instrument for Enhanced Transparency, Accountability, Privacy, and Security (BINC-TAPS) seeks to provide a robust, transparent, scalable, immutable, and tamper-proof solution for conducting national population and housing censuses, while also fostering socio-economic advancements. This paper presents a comprehensive overview of our research, with a primary focus on the implementation of the blockchain-based proposed solution, including prototype testing and the resulting outcomes.

Keywords: national housing and population census; blockchain; blockchain-based national housing and population census; architecture design; Hyperledger Fabric; national census prototype

Citation: Rasheed, S.; Louca, S. Blockchain-Based Implementation of National Census as a Supplementary Instrument for Enhanced Transparency, Accountability, Privacy, and Security. *Future Internet* **2024**, *16*, 24. https://doi.org/10.3390/fi16010024

Academic Editor: Paolo Bellavista

Received: 18 December 2023
Revised: 4 January 2024
Accepted: 8 January 2024
Published: 11 January 2024

1. Introduction

A national population and housing census is meant to collect comprehensive data on the demographics, individual characteristics, and living conditions within a country. These data are crucial for policymakers to make informed decisions regarding the economy, finance, healthcare, social benefits, import/export, education, and other sectors, ultimately driving a nation's development. The United Nations stipulates the necessity of a decennial census, a procedure conducted every ten years to accumulate data about the inhabitants of a jurisdiction [1]. Yet, shortcomings in the census system can substantially hinder a country's strategic planning and overall interests.

The existing census systems are flawed, with data updates occurring only once every ten years, leading to a laborious and redundant process. This system often fails to protect the rights of minorities and marginalized communities against corrupt governments and influential individuals [2,3]. Current census data collection methods encounter difficulties in enumerating elusive populations, often referred to as "missing persons". These individuals live in challenging situations, including unregistered buildings, non-compliant house extensions, shared residences, and concealed or mobile locations [2,4,5]. Deliberate efforts by certain minorities and undocumented immigrants to evade detection further complicate their inclusion in the census. The involvement of humans in data management introduces errors and potential leaks, prompting investigations into formal solutions which reduce human interaction to ensure more precise and secure outcomes. There have been instances of statistical ethnic cleansing and misrepresentation of minorities worldwide. The current system also struggles with transparency, data quality, and the protection of personally identifiable information (PII) [6]. These problems are exacerbated by external immigrants from war zones and internally displaced populations due to climate change, as the urgency for care and relief often takes precedence over data privacy and protection requirements.

The scope of this paper is to investigate the use of a blockchain-based national population and housing census and address the issue of missing persons in census data by utilizing community-level datasets from non-governmental organizations (NGOs) to verify individuals' living conditions and addresses. This method provides a granular and accurate representation of population demographics, combining diverse datasets from various stakeholders to create a reliable source of truth. This contrasts with current government-based systems, which are often susceptible to corruption and political manipulation. Implementing a national population census on a blockchain is expected to provide inherent security, transparency, and immutability advantages, while adding societal value to record, monitor, forecast, and extrapolate a country's demographic dynamics and socio-economic activities. It bridges the gap between computer science, surveying, government data collection instruments, distributed ledgers, demographics, statistical models, and population censuses, fostering a multidisciplinary approach which combines ideas, business applications, technology, and governance frameworks. The result is a new supplementary implementation model for national censuses that enhances security, accountability, and transparency and addresses the challenge of missing people unaccounted for by traditional census methods.

In this paper, Section 2 covers the methodologies followed in the enumeration process. Section 3 discusses the literature understanding of this topic based on our prior publication of a systematic literature review. Sections 4 and 5 cover the characteristics of blockchain, its implementation in the government sector, and related works in our research area. Section 6 highlights the core findings related to Pakistan's national census system (a case study). Section 7 covers a census architecture design based on Hyperledger Fabric. Section 8 mainly demonstrates a prototype of the proposed solution, followed by its testing and results in Section 9. A discussion has been initiated on the results in Section 10, with recommendations to improve the census process. The solution's implication, research contribution, and future directions are covered in the "Conclusions" section. The list of abbreviations is provided at the end of the paper.

2. Existing Enumeration and Housing Population Census Methodology

Conducting a national population and housing census demands substantial financial and human resources. The gathered data, crucial for policy-making and resource management globally, must be collected accurately by posing relevant questions and enumerating all individuals. Census methodology, therefore, plays a vital role and necessitates careful consideration to comprehensively represent a country's population.

Historically, censuses like those in the United States have exhibited bias, neglecting people of color, slaves, and Native Americans, leading to the significant underrepresentation of these groups. This highlights the need for a standardized global methodology. The United Nations addressed this by developing the Principles and Recommendations for Population and Housing Censuses, serving as a guide for countries to achieve reliable population counts at specific points in time [1].

Traditionally, a census involves full-field enumeration, employing various methods to reach households and individuals. Data are condensed into questionnaires designed to capture comprehensive information without marginalization. Conducted through electronic media, mail, or interviews, a census offers an extensive view of a country's population and socio-economic domains. Despite its advantages, a traditional full-field census is complex, time-consuming, and resource-intensive and may lack optimal technological efficiency, leading to data redundancy.

An alternative methodology, the register-based census, focuses on specific sub-groups rather than the entire population. It utilizes information from existing records to identify individuals and households in selected areas of employment, education, and tax brackets. Administrative registers, derived from crucial variables, are enumerated and compiled through the census process. Examples of administered variables include taxation, employment, education, social welfare, and businesses.

The register-based census methodology is cost-effective, utilizing existing data sources and requiring less funding than the traditional full-field methods [7]. However, it faces challenges in defining variables, potentially leading to discrepancies and the overlooking of certain subgroups, like transient populations. Information may be outdated due to restricted access to registries, limiting this method's overall accuracy [1].

An integrated approach combines register-based and traditional census methods, improving data accuracy and population coverage. This method leverages administrative registers to define subgroups and incorporates field research for recent information. This integrated method enhances data accuracy and accessibility, whilst being cost-effective and not as resource-intensive as traditional methods [8].

Identifiers are essential in both register-based and field research-based methods to prevent overlaps and handle duplicates. Cross-referencing and probabilistic methods address missing entries. A rolling census, conducted annually, complements full-fledged censuses, providing up-to-date information and allowing flexibility for future surveys [1].

Blockchain technology can address the issue of outdated data by enabling regular and real-time updates through electronic sharing. Decentralized data collection from smart devices and IoT sensors directly contributes to the blockchain network, while smart contracts automate survey processes, reducing manual administration. Blockchain's interoperability and integration capabilities enable the consolidation of data from diverse sources, creating a comprehensive view of the population. By leveraging these blockchain features, national census authorities can establish a dynamic and responsive system that continuously updates and enriches census data. Artificial intelligence can enhance data profiling and analysis. Post-census activities, including data processing, storage, and maintenance, are crucial for preventing leaks and improving future planning. Control information documents practices and errors for continuous improvement [1].

Setting bounds on variables during data processing minimizes errors. Imputation methods handle missing or error-prone entries, ensuring smoother data analysis [1]. Archives and restricted access maintain public trust and engagement, with countries adopting different models for data accessibility to researchers [9].

3. Literature Review

The United Nations, in collaboration with individual countries, has made significant efforts to enhance the standards of census methodologies. These endeavors aim to minimize errors and improve the accuracy of collected insights. However, despite these dedicated efforts, there remain challenges that hinder the optimal execution of census operations, particularly in terms of enumerating and accurately covering all individuals.

One of the fundamental reasons for these challenges is the limited scalability of existing solutions. Many census approaches struggle to provide comprehensive coverage or fail to be universally applicable across diverse regions and countries.

Addressing these limitations requires a concerted effort to develop more scalable and adaptable census methodologies. By leveraging advanced technologies, embracing data-driven approaches, and fostering cross-country collaboration, we can work towards overcoming these pitfalls and achieving more precise and inclusive census outcomes. Moreover, continuous innovation and stakeholder engagement will be vital in creating a robust and resilient census system that can cater to different populations' unique needs and characteristics worldwide.

In the preliminary stage of our study, we carried out a systematic review of the existing literature, identifying six major deficiencies in the national population census process, which we will be discussing in this section. These flaws encompassed issues such as population coverage, ethnicity and race-based discrimination, apprehensions regarding data privacy and security, obstacles in the dissemination of data, substantial financial implications of executing a census, and hurdles in public engagement [10].

3.1. Population Coverage

Census data collection has long been a concern due to its limitations in accurately covering all forms of living arrangements for individuals within a region or country. The guidelines set forth by the United Nations for enumeration recommend canvassing households, but this approach overlooks individuals who lack proper living arrangements, such as homeless communities, frequent travellers, and nomadic groups who do not associate themselves with a fixed location [11].

Unfortunately, these gaps in census data have real-world consequences. For example, during Pakistan's 2017 census, approximately 1.5 million people from Karachi were reported missing from the count [12]. Similarly, in the United States, the 2010 census missed 1.5 million children from minority communities, particularly 2.1% of Black Americans and 1.5% of Hispanics [13]. These discrepancies highlight the need for a more comprehensive and inclusive data collection approach.

The transparency, quality, and accuracy of data are also pressing challenges in the current census system, as is the need to safeguard personally identifiable information (PII) [6]. The sensitivity of these data makes it crucial to adopt robust data protection measures to ensure privacy and prevent misuse.

Efforts have been made to address the issue of "difficult-to-enumerate" population groups. The UN principles suggest incorporating the place of count for the census and whether it represents the individuals' usual residence, aiming to reduce miscounts of children and frequent travelers [1]. While this is a step in the right direction, it still falls short of accounting for dynamic variables and requires a more comprehensive and adaptable approach that encompasses homeless, stateless, and refugee groups.

Another significant concern is over-coverage, which occurs when there is duplication in the data, leading to individuals being counted more than once. This could happen with individuals having multiple jobs or addresses [12]. Although over-coverage can be managed, it may not always be handled in the most efficient manner.

3.2. Ethnic and Racial Discrimination

Throughout history, data collection by ruling authorities has been marred by discriminatory practices that either underestimate or overestimate certain groups based on factors like religion, race, or ethnicity. These inaccuracies have significant implications, as census figures play a crucial role in policymaking and funding decisions. When certain communities are underrepresented, it leads to an unfair allocation of funds, depriving them of much-needed financial assistance.

For instance, the 2010 US census underrepresented Native American populations of Navaho origins and Alaskan natives by 4.9 percent. The 2020 census is expected to exacerbate underrepresentation, especially in regions with inadequate broadband access due to the shift to online data collection [13]. In India, the implementation of the CAA law, which bans Muslim immigrants from settling in the country, has sparked country-wide uproars and riots, raising concerns about the safety and underrepresentation of the Muslim population in the upcoming decennial census [14].

Efforts have been made in census data collection to be more inclusive, such as counting Sikhs as an individual ethnic community in the 2020 US census [15]. Similarly, India plans to account for households run by transgender individuals separately, rectifying the earlier mislabeling of such households as male-headed households [16].

However, there are inherent drawbacks in the current system. The data update process for each individual is laborious and repetitive, and there is a need to protect the rights of minorities and marginalized communities from corrupt governments and selfish potentates [2,3]. For instance, Nigeria's census has witnessed instances of statistical ethnic cleansing and misrepresentation of minorities [1].

3.3. Privacy

Privacy concerns related to census data have become increasingly significant with the advancement of technology. In the past, the primary worry was the possibility of data leakage to enemy states, which could provide them with valuable information for planning attacks. For example, a cyber-attack on the Census Bureau exposed personnel data for approximately 4200 individuals, highlighting the need for stronger data protection measures [4].

In more recent times, the concerns are similar but pertain to both the use and storage of census data. In the Australian census, there was a public outcry when the option to opt-out of retaining the original names of users was removed without sufficient consultation or prior knowledge [17]. These privacy concerns can lead to reluctance or even boycotts of the census.

Centralized data storage also raises privacy concerns, as it makes data vulnerable to cyber-attacks. The rise in cybercrimes necessitates a more technologically advanced and secure approach to data maintenance, access, and storage to address privacy concerns among the population.

In the past, privacy concerns have led to legal actions. In Germany, citizens sued the government over a 160-question long census questionnaire, raising concerns about the potential identification of individuals through the collected information. The ruling established a right to "self-determination" of the information shareable by citizens, in contrast to the surveillance behavior seen in Nazi-Germany [18].

Data leakages can occur in four ways, as identified by Dunn and Austin [19]. (i) Accidental leakage may result from human errors, oversight, or ignorance in handling data during the traditional census process. The number of individuals involved in the census process increases the risk of such leakages. (ii) The second form of leakage is driven by malicious intent, where involved parties seek unauthorized access to data for criminal purposes. Strict precautions, training, and penalties are necessary to prevent this type of leakage. (iii) The third form of leakage arises from legal obligations, such as court orders which require the disclosure of information publicly. Efficient means of validating personal information without disclosing actual data can help handle such circumstances. (iv) The last form of leakage is related to statistical necessity, where data are disclosed to researchers or the public for analysis and interpretation. While insights derived from data analysis do not violate privacy, disclosing actual identifiable data can raise concerns about data misuse and potential human errors leading to leakage.

Addressing privacy concerns in census data collection requires robust security measures, data protection protocols, and strict compliance with privacy regulations. Furthermore, involving the public in discussions regarding data collection and ensuring transparency can build trust and confidence in the census process.

3.4. Census Data Distribution

Another vital aspect to explore concerns the mechanisms employed by census bureaus for sharing data with relevant or requested authorities. Juran [6] delves into the reports from the 2010 world program on population and housing censuses to comprehend the primary methods utilized for disseminating census data. It was found that approximately 63 countries predominantly employ paper publications for data distribution. In contrast, 34 countries rely on static web pages, and a mere 17 countries have adopted interactive online databases. Predominantly, developing nations utilize paper-based methods, alongside CD-ROMs and DVDs, for data distribution. Notably, almost all countries, including the United States, lack a distributed and decentralized system capable of meeting the requirements of all stakeholders in a timely manner.

3.5. Cost of Census

Conducting a census is an extensive and resource-intensive task that involves significant costs, including human capital, financial expenses, and materials required for data collection and enumeration. The scale of census activities contributes to their steep costs, which tend to increase over time. The type of census also influences the costs, with a full-field research census being more expensive compared to administrative censuses or small-scale enumerations in specific areas [20].

The cost of conducting a decennial census in the United States serves as an illustrative example. The 2010 census cost approximately thirteen billion dollars, which was twice as much as the cost of the census in the year 2000. Similarly, the cost of the 2000 census was double that of the 1990 census. These escalating costs are due to the growing complexity of census processes as the population increases [8]. In fact, the estimated cost for the latest decennial census in 2010 increased from an earlier projection of 11 billion dollars to 14 billion dollars. This increase was a result of abandoning the handheld devices called NRFUs and reverting to the traditional paper approach, which required more trained field staff to conduct in-person data collection [21].

3.6. Cooperation and Participation

Public cooperation and participation are crucial aspects of the census process. Historically, there has been public reluctance regarding census data collection, often stemming from concerns about potential increases in taxation or other adverse consequences. However, in recent years, this reluctance has decreased, although government still need to exert efforts to improve coverage.

To encourage participation, governments invest in media coverage and advertisements to raise public awareness about the census. For example, in the United States, in-post mail is sent to addresses of households to encourage responses. In 2010, about 83 percent of households responded to the in-post mail on time. The remaining non-responsive households required on-field workers to conduct in-person data collection. In developing countries like Bangladesh, where literacy rates may be lower, schoolteachers were engaged to spread census awareness and later perform in-person enumeration duties. This approach can be costly for the government, especially in developing states, as additional resources are needed to ensure widespread census knowledge and coverage [22].

In conclusion, census data collection faces challenges in counting hard-to-reach populations, commonly known as “missing persons”. These individuals reside in difficult-to-enumerate situations, such as unregistered buildings, non-compliant house extensions, shared residences, and mobile or hidden locations. Certain minorities and undocumented immigrants intentionally avoid detection, complicating their inclusion in the census. Human involvement in data management introduces errors and leakages, prompting the exploration of a formal solution minimizing human interaction for more accurate and secure results.

Addressing these challenges necessitates a multidimensional approach involving technological advancements, community engagement, and adaptive enumeration methods. Utilizing modern technologies, like blockchain-based solutions, could enhance accuracy, efficiency, and inclusivity in census data collection while preserving data privacy and security. Collaborating with local communities and incorporating their knowledge ensures a more comprehensive enumeration, especially for historically underrepresented groups. These steps contribute to building a more reliable and equitable census system that accurately reflects the diverse population of a region or country.

4. Blockchain

A blockchain is a distributed ledger that allows for the recording of transactions and data in a verifiable and permanent way across a decentralized network. It uses cryptographic techniques like hashing and digital signatures to create immutable records [23]. Multiple participants, called nodes, maintain a copy of the ledger each and must validate any updates. Consensus mechanisms like proof-of-work ensure that transactions are verified before being added to the blockchain [24].

A key feature of the blockchain is decentralization: i.e., it does not rely on a central authority. The ledger is maintained by a peer-to-peer network collectively adhering to protocols [25]. All participants have access to an identical copy of the ledger. Any changes have to be initiated by participants and verified by others. This eliminates intermediaries and centralized control [24].

While blockchain ledgers are public, users interact pseudonymously through generated addresses. This provides transparency and, yet, ensures a level of anonymity [25]. Data, once added to the blockchain, cannot be modified retroactively without alteration of all subsequent blocks. This makes blockchain networks immutable and tamper-proof [23].

Smart contracts allow complex applications to be built atop blockchains. These are programmable contracts that self-execute when certain conditions are fulfilled [26]. Blockchain technology enables trustless interactions and automated processes between disparate parties on a distributed network.

4.1. Characteristics of Blockchain

A blockchain can be characterized as a distributed ledger of any type of transactions, where a transaction is the exchange of data (medical data, consumer details, product data, etc.). The blockchain offers many benefits and advantages over traditional centuries-old centralized systems. The advantages range from security to efficiency and from immutability to non-alterability [27]. Its main characteristics are as follows:

- P2P: the blockchain network is peer-to-peer, which means that all members of the network interact with each other and resolve network-related issues in a collective manner.
- Cryptographically Secure: blockchains rely on public–private key cryptography for security and secrecy. Each user has a public key through which they receive cryptocurrencies and information, whereas the private key is for sending information. Each user has an address, which is derived through the hash of a public key of the user, to receive funds through. Each blockchain may use a different version of a hash function, which essentially encrypts information to add security to information transfer processes. For the Bitcoin blockchain, the hash function is SHA256, whereas for Ethereum it is ETH HASH.
- Append-Only: blockchains are append-only record-keeping decentralized technologies. This means that data can only be added to the blockchain and cannot be changed or modified once added to it. Hence, transactions and information in the blockchain remain forever.
- Time-Stamped: transactions are stored in chronological order.
- Decentralized: the blockchain is decentralized, meaning no single authority controls, influences, or manipulates the data records that exists in a blockchain.
- Consensus Mechanism: The validation mechanisms, that is, how transactions are added to the blockchain, are defined through a consensus mechanism that differs from blockchain to blockchain. For instance, for the Bitcoin network, "Proof of Work" is used, which corresponds to the computational power each user/node has. The higher the computational power, the greater the likelihood of producing a block to add to the blockchain [28].

4.2. Blockchain Implementations in Government

Blockchain-based solutions vary from an industry to another in domain and their applications. We can find various implementations, ranging from the energy sector [29,30] to finance [31], healthcare, supply chains [32], academia [33,34], and many more. Nonetheless, blockchain technology helps the government sector to improve services by offering solutions in the following areas:

- Land registry—Blockchains can create immutable land records, making title transfers transparent and reducing fraud. The Swedish government is testing a blockchain land registry system [35].

- Voting system—Blockchain-based e-voting can enable transparent, tamper-proof, and verifiable elections. Brazil has implemented a blockchain voting system [36], as well as Estonia, who is setting an example in the region [37].
- Identity management—Blockchain digital IDs can prevent identity theft and enable seamless Know Your Customers (KYC) processes. Dubai plans to put all government documents on a blockchain by 2020 [38].
- Taxation—Blockchain technology can improve tax filing, collection, and audits through trackable records and the automation of processes. The UK's tax authority is exploring using a blockchain for VAT collection [39].
- Benefits disbursement—Blockchain wallet systems can enable the direct fraud-proof transfer of government benefits and welfare schemes. The World Food Programme uses a blockchain for aid distribution [40].
- Regulatory compliance—Blockchain 'regtech' solutions can make regulatory reporting and compliance more efficient through shared ledgers between regulators and entities [41]. The US Security and Exchange Commission (SEC) is implementing a blockchain for regulatory oversight [42].
- Public procurement—Blockchain technology enables transparency and immutability in government procurement through tamper-proof contract data and spending records on a public ledger. Dubai uses a blockchain for government procurement [43].

Despite the credentials of blockchain and its various applications in the government field, it is yet to be explored and tested for national population censuses. Due to the characteristics of blockchain technology, it is expected that blockchain-based systems for population and housing censuses will also inherit these characteristics. Therefore, such systems can potentially be used for the following:

- Recording census data—Census data like name, age, address, etc., can be recorded on a blockchain in a secure and immutable way. These data would be transparent, while also protecting privacy through encryption.
- Data integrity—Blockchain's distributed nature makes census data highly secure and immutable. Once recorded, the data cannot be altered without network consensus. This ensures accuracy and trust in census figures.
- Authentication of census enumerators—Blockchain technology can enable the secure authentication of census enumerators using digital IDs. Their activities, like surveys completed, can also be immutably recorded on the chain.
- Access control for data—Census data on the blockchain can consist of granularly permissions to ensure that only authorized parties can access specific data. This maintains transparency while also prioritizing privacy.
- Automating parts of the census—Smart contracts enabled via a blockchain allow for the automation of certain census operations like notifications, tracking fieldwork, and data transfer between parties.
- Decentralized identification system—Blockchain technology can be used to create tamper-proof digital IDs for citizens that are secured cryptographically. This allows for the unique identification and authentication of each person.

Our research hypothesis is that the implementation of blockchain technology holds the promise to alleviate, if not entirely eradicate, the identified challenges and potential solutions listed in Table 1.

Table 1. Challenges of national population censuses and potential solutions [10].

Sr.	Challenges of Centralized System	Blockchain-Based Solution	Potential Solution
1	Population Coverage	improve	Decentralized NGOs, part of the blockchain, carrying out the enumeration of smart contracts

Table 1. *Cont.*

Sr.	Challenges of Centralized System	Blockchain-Based Solution	Potential Solution
2	Ethnic and Racial Discrimination	improve	Immutability of records, timestamped records
3	Privacy Concerns	resolve	Decentralization and cryptographically secured, immutable data
4	Census Data Distribution	resolve	Smart contracts
5	Cost of Census	improve	Decentralization
6	Cooperation and Participation	improve	Transparency and accountability

5. Related Blockchain-Based National Population Census Work

The exploration of blockchain technology for the housing and population census enumeration process is a recent development in the field of research. In the existing body of literature, ref. [44] introduces another relevant work. The researchers in [44] have proposed an infrastructure that integrates blockchain technology for census enumeration and internet voting applications. While these applications have distinct considerations and requirements, the researchers suggest using voting stations for census enumeration processing, indicating potential synergies between the two processes. The focus of the research has been on the cost and security aspects of the census enumeration process. Comparing blockchain technology with traditional models like the client–server model, cloud-based technology, and mobile-based services, the researchers assert that the blockchain offers superior security features [44].

The proposed implementation involves using a private blockchain platform (multichain blockchain) that provides more control and privacy. They suggest building an isolated peer-to-peer infrastructure, which will be governed by the Indian government. This infrastructure will consist of 4689 fixed stations distributed across constituencies. The stations will be connected via fiber optic cables and will function as multichain nodes of the blockchain network [44].

Given the explanation, there are a few considerable limitations of this blockchain-based proposal for the national population census, which are as follows:

- The reliance on unique IDs, like Aadhaar, risks excluding portions of the population with lost or without proper identification, especially those in remote areas or with complex living arrangements.
- Allowing multiple family members to log in could compromise data quality and accuracy.
- The proposed peer-to-peer infrastructure with thousands of nodes connected via fiber optic cables seems impractical and extremely costly.
- Voter information has different requirements than typical census data collection. Combining this to the two-on-one blockchain system may not be optimal.
- Blockchain technology brings advantages like security, transparency, and decentralization. But factors like population coverage, efficient data collection, verification, and access control for different entities are equally critical for an effective census system.

Nonetheless, India's overall literacy rate is 72% as of the year 2015 [45], as a person who can write their name in the Hindi language is also considered literate. Over 55% of Indians have access to broadband internet, and only 20% of Indians have the ability to use the internet [46]. The proposed online system might expand the gap in population coverage in the country.

A pragmatic approach would likely incorporate blockchain technology selectively for certain aspects, like identity management or data integrity, while also investing in training enumerators, using hybrid data capture modes, and building inclusive identification systems. Table 2 presents a comparison between related work and our proposed solution.

Table 2. A comparison table between related systems and the BINC-TAPS framework.

Attributes	Census Enumeration and Internet Voting App	BINC-TAPS Framework
Application Type	Online, available to citizens	Online, available to surveyors and NGOs
Platform Type	Combined	Solo
Blockchain	Multichain	Hyperledger Fabric
Population Coverage Issue	None	Improves
Racial and Ethnic Discrimination	No	Resolve
Privacy Concerns	Resolves	Resolves
Census Data Distribution	None	Resolves
Cost of Census	Reduces	Reduces
Cooperation and Participation	No	Improves

Countries Exploring the Integration of Blockchain in the Census System

There are currently no countries that have fully integrated blockchain into their national census systems. However, there are some examples of pilot projects and early-stage experimentation with blockchain for census-related applications:

- Dubai–launched the Dubai Blockchain Strategy in 2016, with census data collection mentioned as a future application, but a concrete implementation is still awaited [43].
- India–proposed using a permissioned blockchain for future census data collection by the government. Academic researchers have published design frameworks, but no system has been implemented yet [44].

Several startups like Census Chain and Procivis [47] have worked on blockchain-based identity and census data solutions, but wide-scale adoption by governments remains elusive currently.

While promising proofs-of-concept exist, most governments are still assessing feasibility or running small trials. Large-scale adoption faces challenges like population coverage, technology readiness across census machinery, and legal/regulatory hurdles. However, opportunities remain for selective blockchain integration to enhance data integrity and efficiency in the future.

6. A Case Study of Pakistan's National Population Census Bureau

Considering the challenges discussed in the literature review section, we picked a case study of Pakistan. The selection of Pakistan as the primary focus stems from one of the researcher's origins in the country, facilitating easier access to information for nationals compared to attempting the same for other countries. Additionally, the researcher in question expressed a desire to contribute to the well-being of her fellow citizens. Utilizing a qualitative research approach, face-to-face structured interviews were conducted at the National Census Bureau to identify challenges and propose a blockchain-based solution. This solution aims to address concerns related to census data sensitivity, privacy, and national security. This method offers the benefits of formulating relevant questions, ensuring the consistency and comparability of interview responses, and minimizing errors and biases. The Pakistan Bureau of Statistics stands as the country's principal official agency tasked with collecting, compiling, and disseminating statistical information pertaining to the national population census.

During the second phase of our research study, a list of 34 questions for structured interviews was prepared and split into five sections: participant backgrounds, existing system, missed populations, blockchain awareness, and new system requirements (see Appendix A). It was then translated into Pakistan's national language—Urdu—so that participants could be comfortable during the session and completely understand what had been asked. The target participants were selected based on the following criteria: should have 10+ years of experience and be part of a planning, data collection, consolidation, and compilation or IT support team. We set the minimum target bar for 10–15 participants because the national census bureau is a single federal body in the country, and the participants' teams will mainly use the proposed system [48].

The interviews took place during October–December 2022. Ten participants participated with consensus. A total of 450 data points were collected via 34 main questions and 11 derived questions. KoboToolbox was used for data collection, and Python and Jupyter-Notebook were used for data transformation, cleaning, and analysis. Pandas, NumPy, Plotly, and Sweetviz libraries were considered during the analysis [48].

The key findings from analyzing the structured interview responses are as follows [48]:

- Supervisors can manually alter census data without immutable audit trails, raising concerns about integrity. Only supervisors can change data when inputting missing values, lacking transparency.
- The census bureau relies solely on authorized surveyors for enumeration and does not cross-verify with external data sources, risking undercounting marginalized populations like the homeless. Some participants identified potentially missed groups, including those in remote areas, transgender individuals, and unregistered orphanages.
- The proposed blockchain system could allow accredited entities to participate in enumeration for universal coverage. A self-sovereign identity (SII) model would help count populations without official IDs. A suggested consortium blockchain aims to balance openness with security.
- The participants outlined needs such as digitization, reliability, accuracy, and upgraded infrastructure and desired features including immutability, accountability, encrypted storage, etc., as listed in Table 3.

Table 3. Priorities of the requirement and functional decomposition [49].

Requirement	Function Decomposition	Priority
Data digitalization	Website; login portal; digital questionnaire; authorized logins; data sharing	High
Data consistency, immutability, reliability, security, and encryption	Blockchain	High
Eliminate political relevance	Blockchain	High
Elimination of data redundancy	Blockchain	High
Summarized data accessibility for everyone	Easy access to data downloads' data in the required format	High
Thumbprints collection	Biometric thumbprint collection	Medium
Progress Dashboard/Reports	Visibility of ongoing census count	Low
Confirmation notification to the interviewee	Sending of SMS notifications	Low
NGOs/NPOs integration		Register themselves on the website or bureau of census directly, get them registered, and provide essential access.

In summary, the interviews revealed gaps in the current processes around traceability, transparency, and population coverage. A permissioned blockchain solution could potentially address these through cryptographic security, smart identity management, and controlled participation, enhancing integrity while retaining privacy.

Requirements and Function Decomposition

Based on the findings from the structured interviews and the challenges, requirements, and new features for the proposed system, the logical functional units for the solution are identified and presented in Table 3 below. Our primary research objective is to address the issue of missing persons in the population census count. We propose the integration of NGOs and non-profit organizations (NPOs) in the census process, as the census bureau typically holds exclusive data and does not incorporate reliable third-party or open-source information. This integration would distribute responsibility and costs, ensuring comprehensive nation-wide coverage [50].

It is worth noting that comprehending the cultural boundaries of the population and housing census bureau is crucial. Due to their limited familiarity with emerging technologies and a cautious approach to sharing trust and responsibilities, the bureau will involve NGOs selectively rather than implementing an open registration module for all. Initiating from a standpoint where they gradually share responsibility is essential to cultivating a trustworthy culture. The registration process can be streamlined as necessary after successfully piloting this project.

7. Architecture Design of a Blockchain-Based Solution

To finalize the architecture design, the following tools were selected and used in a proof-of-concept solution development:

- Hyperledger Fabric;
- InterPlanetary File System (IPFS);
- MongoDB Atlas;
- NodeJS/ReactJS;
- Ubuntu 16.04 OS.

7.1. Hyperledger Fabric

Hyperledger is a permissioned blockchain infrastructure that emerged in early 2016 as a robust solution tailored to business and industry needs. Originally named Open-Blockchain and later rebranded, Hyperledger, initiated by the Linux Foundation, functions as a framework, meaning that it is neither a blockchain nor a company. It is devised to offer deployment solutions for business infrastructure, utilizing blockchain technology to enhance operational efficiency. With the collaboration of approximately 100 companies, Hyperledger is an ongoing project fostering consistency and developing industry-driven blockchain applications to better meet business demands.

Within the Hyperledger umbrella, various projects and frameworks exist, with Hyperledger Fabric being one of them. Serving as a permissioned blockchain network, Hyperledger Fabric is established by organizations forming a consortium. Each member organization in this network configures their peers for network participation, implementing an access control mechanism to regulate information visibility among peers as needed [51].

Hyperledger Fabric, a cryptocurrency-free blockchain framework, operates on a permissioned network with restricted access. Its modular design simplifies the setup of business-specific blockchain infrastructure, emphasizing security through the Byzantine Fault Tolerance (BFT) algorithm. Using chain code for smart contracts, the framework employs Docker for containerization, ensuring modularity and portability. The voting-based consensus mechanism in Hyperledger Fabric, comprising endorsement, ordering, and validation phases, facilitates efficient and secure transaction processing. Notably, the mechanism considers users' hierarchical roles, incorporating access control lists and identity verification at each channel point [52].

It is crucial to acknowledge that platform selection involves trade-offs, and no framework is flawless. We have compared five leading permissioned blockchain technologies—Hyperledger Fabric, Ethereum, Quorum, MultiChain, and R3 Corda—considering consensus type, smart contracts, open source, support and governance, and cryptocurrency enablement [50]. Hyperledger Fabric offers cross-industry solutions instead of finance-only solutions. For the assessment of these permissioned blockchain frameworks, five metrics have been considered: latency, throughput, scalability, privacy, and adoption. Fabric, Ethereum, Quorum, Multichain, and Corda have, respectively, an average grade of 4.2, 3.3, 2.2, 2, and 2.2 [53]. Hyperledger Fabric stands out as first in terms of privacy, latency, and throughput and second in scalability and adoption.

The growing number of Hyperledger Fabric's use cases as well as government-related applications, as shown in Table 4, are visible proof of the tool's maturity and the strength of the developer community around it [54], making it a go-to framework when investigating blockchain adoption in enterprises. Its plug-and-play design provides unlimited network configuration options.

Table 4. Permissioned blockchain framework use cases and government-related applications.

Blockchain Framework	Use Cases	Government-Related Applications
Hyperledger Fabric	Manufacturing Asset Depository Direct Communication B2B Contract Assets Interoperability Trade Finance Pharmaceuticals Financial Services Banking Healthcare Internet of Things (IoT)	Secure Electronic Health Records (EHRs), drug supply chain integrity, copyright protection, royalty distribution, voting systems, land title registration, identity management, public procurement, digital identity systems, birth certificates, marriage licenses, and other civic documents
Ethereum	Decentralized Autonomous Organizations (DAOs) Token Launches (ICOs) Enterprise Ethereum Non-Fungible Tokens (NFTs) Stablecoins Decentralized Finance (DeFi) Royalties in Music Industry	Digital currency/payments, land registration, identity management, record keeping, education, supply chain traceability, healthcare, voting, and legal entities management
Quorum	Financial Services Supply Chain Healthcare Real Estate Energy	Not explicit details found
Multichain	Asset Management Data Management Financial Transactions Compliance and Regulation	Online census of India
R3 Corda	Trade Finance Insurance Healthcare Capital Markets Supply Chain	Land registry Regulatory compliance Identity verification Government bonds and securities Public sector data management

Moreover, various researchers have covered Hyperledger Fabric comparisons with other frameworks such as Hyperledger Sawtooth, ConsenSys Quorum, GoQuorum, Hyperledger Besu, Endeavor Ethereum, Openchain, Graphene, Exonum, etc. [54–56].

7.2. Architecture Design and Process Flow

Figure 1 presents our proposed blockchain-based solution for a national population census, given the selected technology stack. The process of conducting a census using this architecture involves census surveyors collecting data from residents and entering it into the client application. The data are then processed using smart contracts on the blockchain network, with only anonymized or aggregated data being stored on-chain to protect privacy. The thumbprint data are encrypted first using Advanced Encryption Standard (AES), also known as Rijndael [57], and then stored off-chain in IPFS, while the generated hash and census data will be stored in MongoDB via the Hyperledger Fabric network. The identities of census takers and administrators are verified through the identity management system, and the data are securely stored in the database and blockchain network.

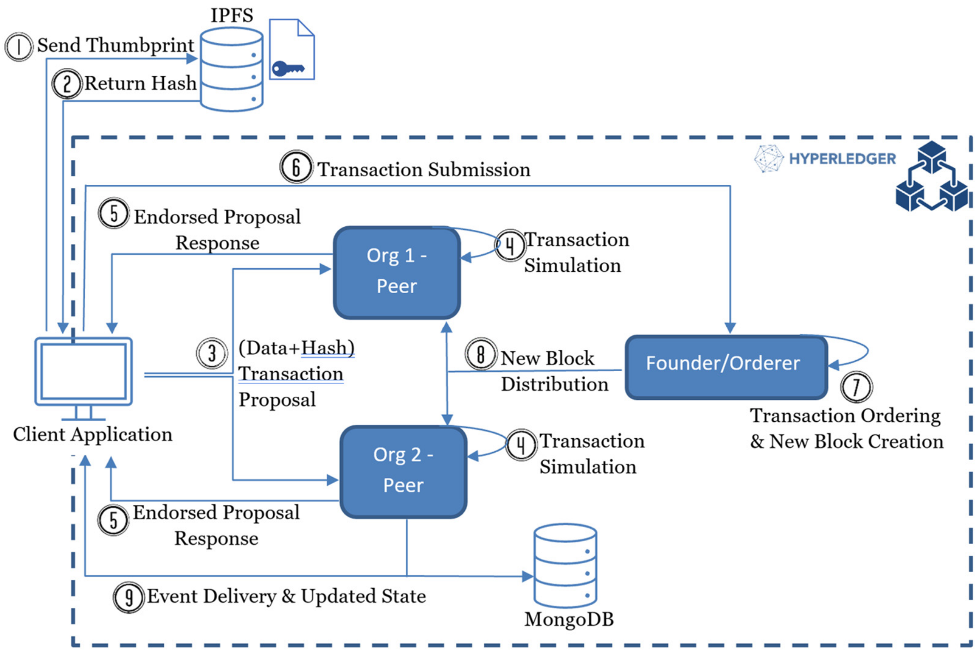


Figure 1. Architecture design.

The proposed blockchain-based national population census architecture consists of the following steps:

1. Once the census survey form is completed, the collected thumbprint image is sent to the IPFS system for storage.
2. AES encryption is applied to the file content, and IPFS generates a hash code representing the content's location and returns it.
3. The surveyors, with their identities, initiate a transaction proposal to a specific channel via applications.
4. The peer on the channel verifies the identity of the surveyors.
5. The peer then responds back to the application.
6. Subsequently, the application forwards the transaction to the peer who ordered this information.
7. The peer who ordered the information then processes the transaction, creating a new block in the process.
8. This new block is then distributed to all the peers on the network.
9. Finally, an event delivery notification is sent to the application, which then updates the state database.

8. Blockchain-Based Census Enumerator System—A Prototype

The central component of the system is the Hyperledger Fabric network, which assumes the responsibilities of overseeing access control, executing intelligent contracts, and retaining transaction data. An array of nodes constitutes this network, all managed by distinct organizations, i.e., the Bureau of Census and NGOs.

8.1. Data Structure and Privacy

The surveyors or representatives from NGOs duly complete the survey form and upload the corresponding information onto the Hyperledger Fabric network.

- A reference pointer to the census information is diligently preserved within the Hyperledger Fabric network (on-chain), and the actual data undergoes AES encryption prior to being stored within the IPFS repository (off-chain).

- The reference associated with the data conserved in IPFS serves the purpose of accessing census-related information or verifying the identities of homeless individuals through the utilization of their thumbprint data.

To safeguard privacy, access to the IPFS and Hyperledger Fabric networks is restricted solely to authorized users. The encryption and distributed storage of individual and housing information within the IPFS ensure that exclusively authorized verifiers possess the capability to retrieve said information. Within the Hyperledger Fabric network, the application of smart contracts ensures that data access and distribution adhere strictly to established regulations and permissions.

8.2. Solution Prototype

The enumerator system is accessible by administrative users, authorized surveyors, and registered NGOs and their representatives assigned to data collection. They can access the system by logging into their accounts.

8.2.1. Dashboard

The dashboard is the main landing page after successful login by all types of authorized users.

For the admin users, the main page shows the total number of active NGOs and surveyors on the system, as shown in Figure 2. The admin user has access to the following pages:

1. Dashboard—it shows the total number of active NGOs and surveyors.
2. Add Users—the admin user has access to add registered NGOs and surveyors to the network.
3. Census Records—the admin user can see records from all the users.
4. NGOs—it shows the active/inactive list of NGO representatives.
5. Surveyors—it shows the active/inactive list of authorized surveyors.
6. Settings—they include the “Profile Settings” and “Change Password” functions.

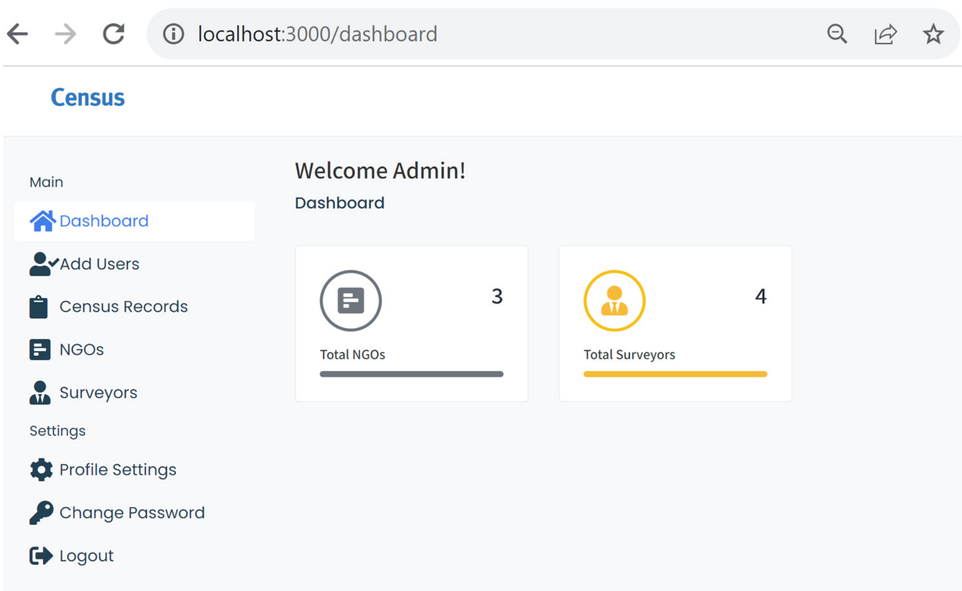
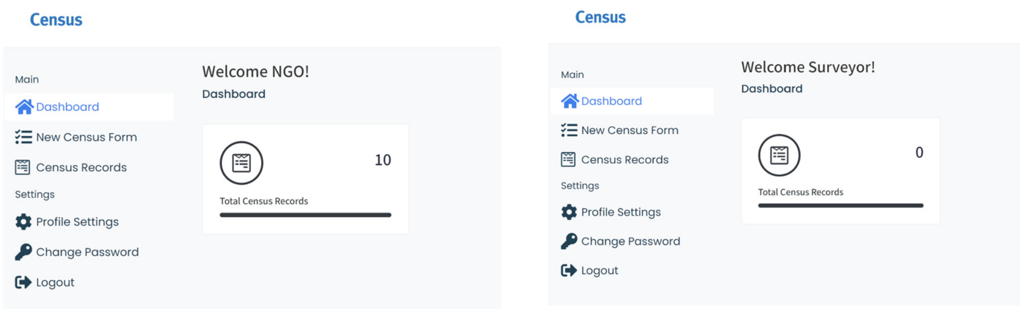


Figure 2. Dashboard of admin users.

For the NGOs/surveyors, the main page shows the user’s total collected census records, as demonstrated in Figure 3. The surveyors and NGO enumerators have access to the following pages:

1. Dashboard—it shows the user’s total collected census records.
2. Census Form—population and housing census data collection form.

3. Census Record—the NGOs and surveyors can only see records they have collected.
4. Settings—they include the "Profile Settings" and "Change Password" functions.

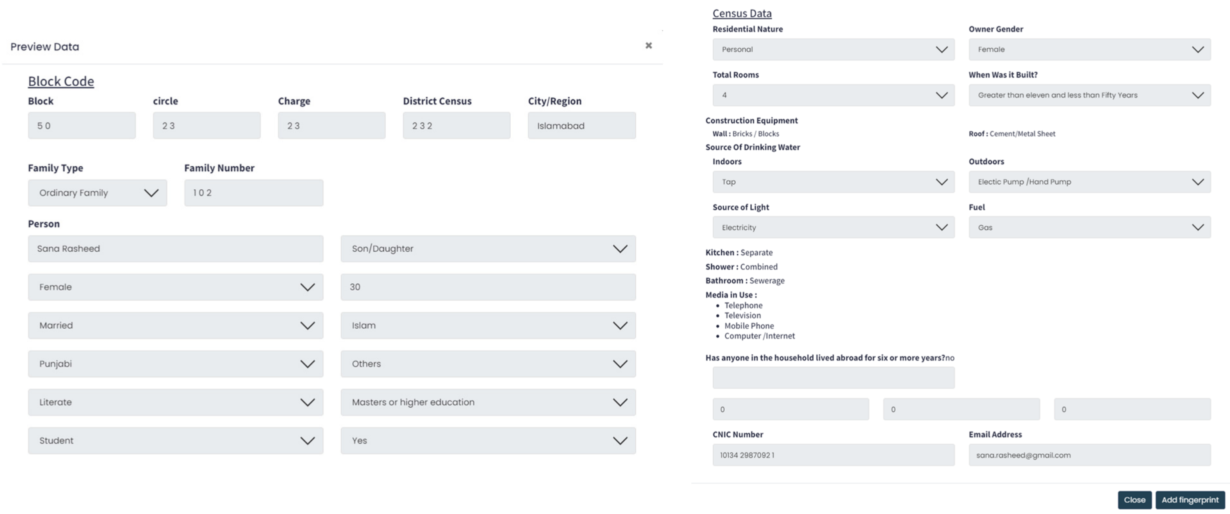


(**a**) (**b**)

Figure 3. Dashboard of (**a**) NGO representatives and (**b**) surveyor users.

8.2.2. Census Form

The census survey form has been structured based on the information provided by the bureau [58]. After collecting the data, a filled form will be reviewed by the interviewee, followed by the requesting of a thumbprint, as shown in Figures 4 and 5, respectively.



(**a**) (**b**)

Figure 4. Complete digital census form in preview view: (**a**) initial section of the census form to collect personal information; and (**b**) continuation of the census form, collecting information about the individual's house.

Individuals are then asked if they want to monetize their data and are given the authority to allow or restrain the sharing of their information with business entities, thus gaining some financial support.

We used a Futronic fingerprint device—model number FS88H—for collecting thumbprint data [59]. This scanner is suitable for high-traffic applications such as border control, identity card issuance, driver's license processing, elections, and various types of civilian Automated Fingerprint Identification Systems (AFIS). Its exceptional capability to capture high-quality fingerprint images with minimal distortions, along with its robust sensor, makes it well-suited for these purposes. Most importantly, the FS88 OEM module is available to integrate scanners into the existing systems [59].

The Census Records option presents the *IPFS hash*, *monetization* (if customer allowed), *created by*, *created at*, and *region* information collected by the enumerator. If any change occurs in the record, it will keep showing one record, but with two or more IPFS hashes (depending on the number of times a record has been altered)—for the previous hashes as intended contents see Figure 6. Admin users can see records from all users; NGOs and surveyors can only see records they have collected.

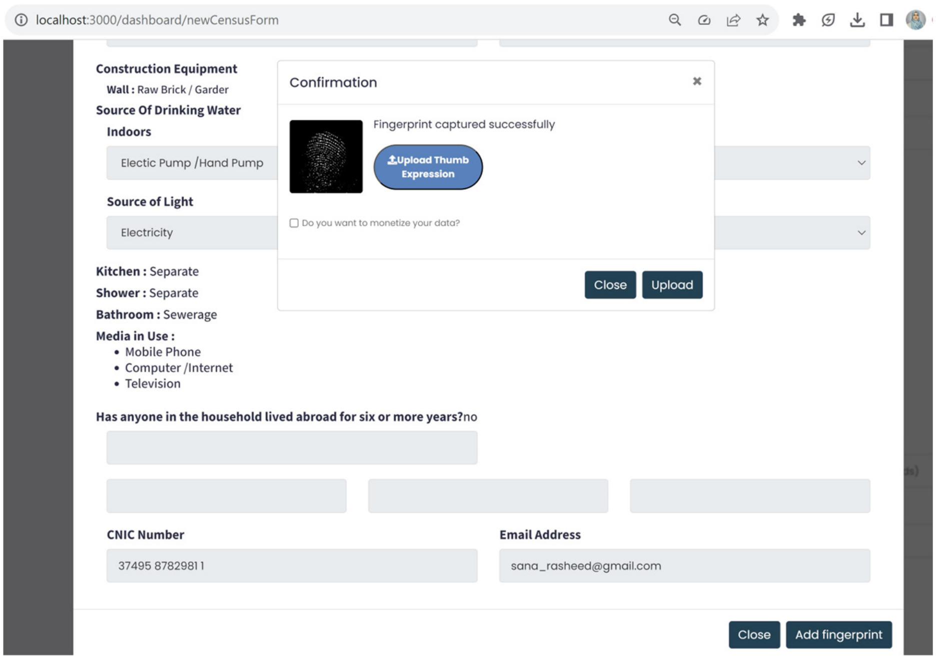


Figure 5. Thumbprint collection and upload.

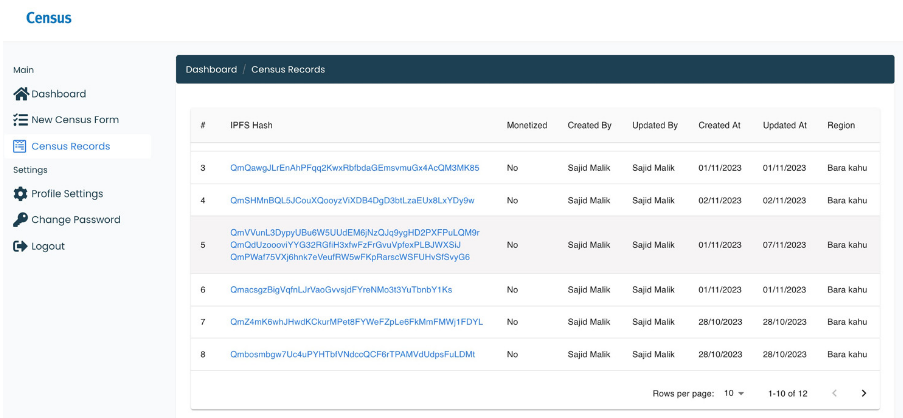


Figure 6. The Census Records option presents all transactions—record 5 is proof of an amended record.

8.2.3. Analytics

On the main page for logging in, the Analytics option is provided to see the aggregated results of the national and population census. The aggregated results are provided in the form of tables and are also available to download in a CSV (Comma-separated values) format to enhance their reusability, as shown in Figure 7.

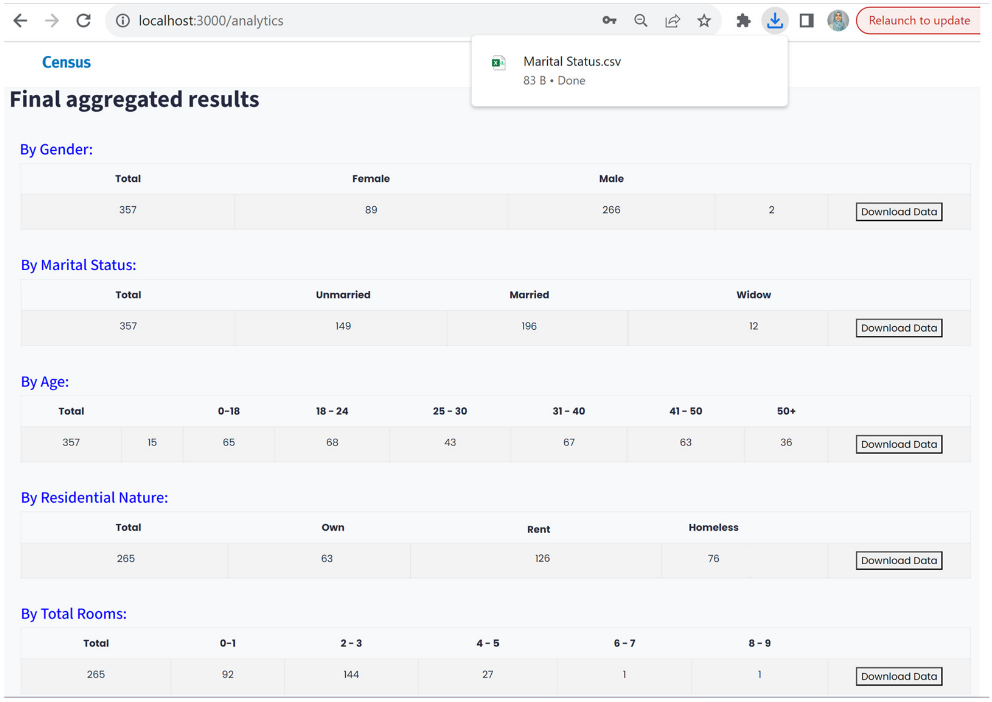


Figure 7. Analytics.

9. Prototype Testing and Results

One of the geospatial locations' boundaries was selected, defined by the National Census Bureau, to collect census data. Three local NGOs—Saylani [60], Pakistan Bait-ul-Mal [61], and Customs Healthcare Society [62]—were onboarded, with consent to support this research work.

Two out of three NGOs helped by providing one volunteer each to conduct door-to-door data collection, leveraging their familiarity with the community. They also shared pre-existing data from their records. The third NGO, with a healthcare facility in the area, gathered information from patients during checkups.

The volunteers were equipped with tablets installed with our solution for information collection. Since the volunteers were not tech-savvy, a training session lasting three hours was conducted before fieldwork. The training covered application usage and addressed any queries that the volunteers had.

We surveyed 265 units, encompassing both residential houses and homeless individuals in one selected geospatial location. Within this group, we identified 53 homeless individuals possessing national ID cards and 23 without any form of identification. For the latter group, we collected thumbprints, totaling 76 homeless individuals (refer to Figure 7—By Residential Nature section).

In the context of this proof-of-concept, it is crucial to emphasize the living conditions of the homeless community. The chosen urban location attracts individuals migrating from rural areas in search of employment opportunities. The labor community often migrates individually and resides in shared accommodations to minimize rental costs, support their families, or sustain themselves. Our interactions with them typically occurred after official working hours. Their living arrangements varied, with some utilizing workshops, placing basic beds in corners, residing in under-construction buildings, beneath bridges, on footpaths, or in open-space tents. Unlike housing owners/renters, the homeless community was cooperative, providing maximum information and granting permission for us to capture photos of their living arrangements, as depicted in Figure 8 (verbal permission obtained).

(a) (b)

Figure 8. Various living arrangements: (**a**) in an open space; and (**b**) on a footpath and dumpster.

The system evaluation was accomplished in four stages: (i) architecture design, (ii) prototype solution, (iii) software usability, and (iv) blockchain performance.

9.1. Evaluation of the Architecture Design

A cost–benefit analysis offers essential insights into the potential expenses and advantages of the proposed system, assisting in informed decision-making regarding its development and deployment. Although the cost–benefit analysis (Table 5) is not inclusive to all, as some restrictions apply to the organization sharing confidential information, it covers the maximum knowledge and estimations based on the information provided by the National Bureau of Census of Pakistan. It is essential to highlight that Pakistan completed its inaugural digital national population census in March 2023, incurring an estimated cost of USD 3.5 million [63]. The digital census still stands on central (non-distributed) infrastructure, and news media broadcasted its same limitations and drawbacks, as identified in the literature review and case study conducted.

Table 5. Costs vs. benefits.

Costs	Benefits
The monthly hosting cost of Hyperledger Fabric could be between $4200 and $8600, considering cloud-based services with moderate traffic and data needs	Data records are resistant to tampering and can be verified
Development and implementation expenses	Enhanced security and confidentiality of the data
One-time resources training cost on blockchain infrastructure/mechanism	Smart contracts facilitate the automation of data verification and aggregation
Costs for migrating from current systems	An unchangeable audit trail could minimize fraud
Integration of fingerprint devices FS88 OEM	A synchronized system could boost the efficiency and precision of census operations
SMS notifications	Offers a prospective framework for various governmental record-keeping systems
Additional time/effort for consensus and governance	An open ecosystem could permit external applications to harness the data

Estimating an approximate cost is challenging due to several dependencies, including development, infrastructure, operations, monitoring, and training services. Typically, in response to government calls for project bids, software and information technology companies submit proposals. These proposals include the costs associated with development, operations, infrastructure, features/functionality, and various services offered, making it difficult to determine an exact cost. Nevertheless, we have attempted to outline the cost dimensions within the confines of IT systems and infrastructure.

9.2. Evaluation of the Prototype Solution

The initial plan was to verify the prototype results using data from the last national census, but this approach proved unsuccessful. In 2021, the Bureau of Census used to release population counts at the block level. However, due to recent policy amendments, this information has been withheld, including the counted figures of attributes crucial for research result verification.

To overcome this obstacle, two additional questions were introduced during the prototype execution, apart from the primary census data collection. Residents were queried about the following:

(i). Have you ever counted for the national population and housing census in 2023 or 2017? (Yes/No)
(ii). Have you faced racial or ethnic discrimination during census 2023/2017 conduct? (Yes/No)

We found that 47% of inhabitants migrated from rural areas approximately 6–9 years ago, seeking an improved lifestyle, enhanced living standards, or better financial prospects to manage family responsibilities adequately. Approximately 15% indicated uncertainty about their previous participation in a census. Only 14% of residential properties (comprising twenty-seven renters and homeowners) confirmed that the national census team had not visited their residence, totaling eighty-one residents based on the average household size of three in this area.

As mentioned earlier, approximately 23 homeless individuals lacking national identity documentation would have remained uncounted had traditional methods been used. The total population in the area is 357; when considering the 81 residents and 23 homeless individuals, the total comes to 104 people. This group represents 29% of the missing population in the selected block.

The Bureau of Census has around 185,512 defined boundary blocks—118,718 rural blocks and 66,794 urban blocks. Given the conduction of prototype testing in a complex living arrangement, we have a sample size of an urban location and, so, can perform some approximate calculations based on different scenarios. These scenarios are inclusive of urban areas only for missed populations and undercounted communities. Various remote areas exist in rural territories where NGOs are quite active in providing basic goods for human needs; this scenario was not included in our study because of a lack of sample size and approximate count of remote areas. Instead, we followed the reports of undercounted and discriminated communities and made some calculations. The following scenarios were identified during our deep study of the literature review, as well as during prototype execution.

Scenario 1: If we make a rough estimate based on the missed population figure we exposed, on average, the census team missed 29% of the population per urban block—66,794 × 104 = 6,946,576 (~7 million).

Scenario 2: Multiple families living in a residential house where all direct owners share the same equity in the house.

Scenario 3: Residential plazas with secure entrances make these locations unreachable to the surveyors.

Scenario 4: Remote areas that are primarily reachable by the NGOs. The census team probably missed some blocks of these under the rural block count.

Scenario 5: Bengalis are the most discriminated ethnic community in Pakistan. About 132 Bengali colonies in and around Karachi struggle for recognition and official identity documents. There are three million (3 million) Bengalis in Karachi alone [64]. Many individuals have resided in the country since well before the 1971 civil war that resulted in the formation of present-day Bangladesh, formerly known as East Pakistan. Despite being born in Pakistan, they face the deprivation of official recognition and citizenship and are unable to vote, attend government schools, or have access to public health. They have been fighting legal battles for their rights since 1993 [65]. Similarly, the transgender community, with reported counts of 400,000–500,000 as of 2017 [66], has not yet been approached by National Bureau enumerators, and it is said that 35% (~140,000) of the community has vanished [67]. Religious communities, i.e., Hindu and Christian, represent 2.7% [68] of the total population (869,011), and they are undercounted too. If we sum up all these communities, they approximately add up to a total of 4 million people (3,000,000 Bengalis + 140,000 Transgender + 869,011 Hindu and Christian).

Given the above-mentioned scenarios, only two of them (scenarios 1 and 5) are computable, as their sample size is available and presents an estimated 11 million people belonging to the missed population. The missed population figure must be higher considering other valid scenarios and the most populated cities. It may seem like 10 million is not a prominent figure against Pakistan's 207 million population. Still, it is of utmost importance that, by missing these people, we are not getting an accurate picture of the country's dynamics for next year's planning or policy making. The missed population mainly belongs to the blue-collar community, which may affect and skew health, economic, and overall country development projections.

As we have just identified, most of the missed population belongs to the blue-collar community. The Bureau of Census needs to reconsider its modus operandi, considering these people's work hours and status as non-ID card holders. In our study, we addressed both issues by approaching them after official working hours (after 4 p.m., when they usually call it a day) and collecting thumbprints from non-ID card holders.

The new system requirement analysis and function breakdown were performed based on structured interviews conducted at the National Census Bureau of Pakistan. As shown in Table 6, all function breakdowns (listed in Table 3) were accomplished, except for sending a *confirmation notification to the interviewee*, as it had a low priority; another exception was prototyping, as it was not suitable. Also, with respect to Table 1, where we hypothesized that the implementation of blockchain technology holds the promise to alleviate if not entirely eradicate, the identified challenges—*population coverage* improved because we combined data from NGOs that were providing support to underprivileged communities. Well-tested during prototype execution and discussed in scenarios 1–5, the *cost of census* execution can be considered to be reduced by involving NGOs. NGOs are non-profit organizations; therefore, by involving them, the cost of designating schoolteachers and providing lounging, daily allowances, and logistic costs would be avoided. Moreover, a blockchain-based system omits the need of audit expenses. As blockchain technology brings the aspect of *privacy* and easy *data dissemination* using smart contracts, we consider these issues to be resolved altogether. *Ethnic and racial discrimination* would surely be improved if we were to include the communities from scenario number 5 in the count. Lastly, the *cooperation and participation* of the public would surely improve if we were to give them a sense of trust in their data privacy and security by educating them on secure system infrastructure through new media channels. The same phenomenon was observed during the pilot when we educated the interviewees about the system being encrypted and secure.

Table 6. Challenges improved with blockchain.

Challenges of Centralized System	Applicable Scenarios/Blockchain Characteristics	Improved/Resolved
Population Coverage	Scenario 1–4	Improved
Ethnic and Racial Discrimination	Scenario 5	Improved
Privacy Concerns	Blockchain characteristic	Resolved
Census Data Distribution	Blockchain characteristic	Resolved
Cost of Census	Blockchain characteristic	Improved (in terms of reduction)
Cooperation and Participation	Blockchain characteristic	Improved

9.3. Evaluation of Software Usability

Although there are various ways to evaluate software usability, given the project scope of simulation execution, we are bound to one metric only—post-survey assessment. As mentioned before, as is the norm of the Bureau of the Census to train their surveyors before sending them for fieldwork, we conducted training for the volunteers on the use of the platform and educated them on the census survey form. At the end of the fieldwork, the volunteers were asked to attempt a questionnaire on the usability of the software used during the conduction of the census. Four questions were asked, and the overall software feasibility condition was satisfied, as shown in Figure 9. No one had faced any difficulties or technical errors during the conduction of the census.

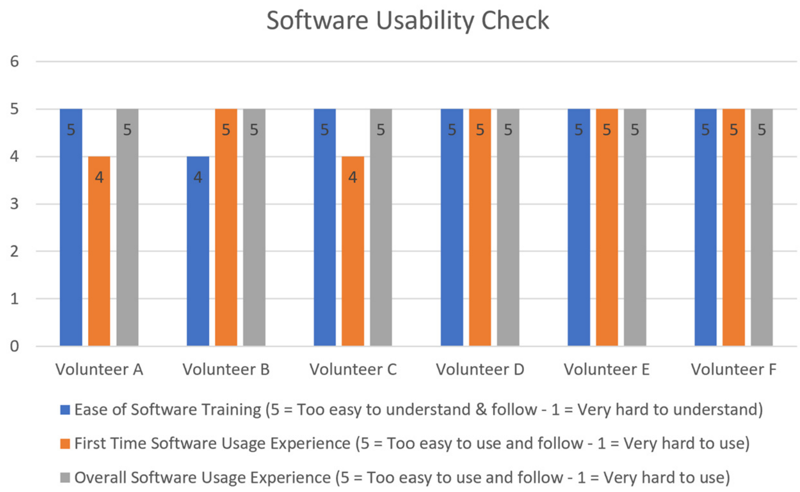


Figure 9. Evaluation of software usability.

9.4. Evaluation of Blockchain Performance

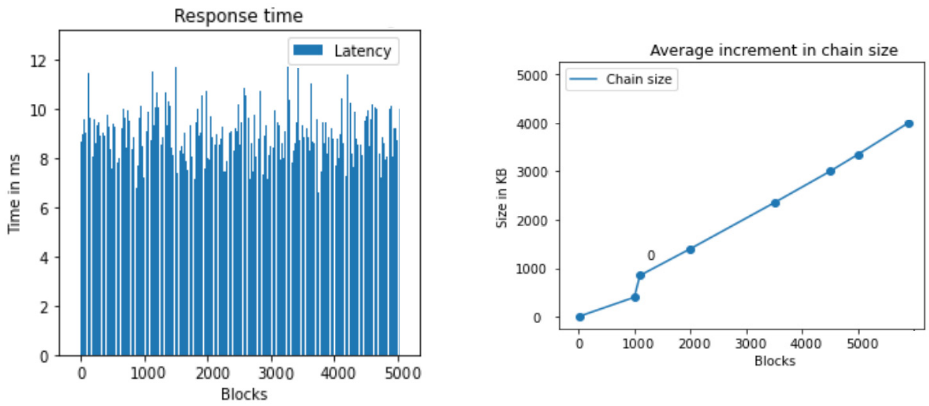
For prototype testing, we configured two peer nodes of the Membership Service Provider (MSP)—the Census Bureau node and the NGO node. The Hyperledger Fabric chain code deployed Amazon Web Services (AWS) and utilized Amazon Lightsail [69] with the specification of 4 GB RAM, 2 CPUs, and 80 GB SSD. The IPFS was configured on infura.io [70], and the total amount of data stored was 128 MB for 265 sets of household information and 30 thumbprint files. The front-end application was deployed and run directly on the tablets with fingerprint SDK installed on it. Given the above, it is also important to mention that this was non-funded research work.

The BINC-TAPS framework performance was evaluated using the latency and block size metrics, outlined as follows:

- Latency: It is characterized as the delay occurring when one component of a system awaits a response from another system component. Within the blockchain network context, latency refers to the duration between the submission of a transaction to the network and the initial acceptance confirmation by the network.
- Block size evaluation: The block size denotes the volume of data stored in a single block, with each block containing transaction data in the chain. The examination of block size involves analyzing the average increase in the size of files or the blockchain.

The performance was assessed for three main functions: (i) creating a record, (ii) obtaining a record, and (iii) updating a record. These are the primary functions that occur in a national census system.

We examined the record creation time to assess latency, as illustrated in Figure 10a. The random delays in the request execution times stemmed from the decentralized nature of the blockchain. With multiple nodes involved and no central system controlling the data, the variations in execution times occurred due to factors like individual node capacities, time-consuming and lengthy census forms, delays in transaction submission, local machine response times, machine speed, and internet bandwidth across the blockchain network. These elements contributed to latency variations across different servers.



(**a**) (**b**)

Figure 10. Blockchain: (**a**) latency and (**b**) average chain size.

The function of creating a record was reiterated 5000 times, propelling the blockchain's expansion from 0.69 KB to 409 KB over 5000 blocks. Demonstrating an average size increase of 79 KB, Figure 10b highlights a consistent correlation between the quantity of blocks and the overall size of the blockchain. Notably, for recently mined blocks containing a substantial volume of transactions, the figure reveals a linear increase in file size, particularly evident between blocks 1012 and 1125. Throughout these processes, the development strategy abstained from specifying the number of transactions per block, resulting in nuanced size differences contingent on the nature of each operation.

Regardless of blockchain performance, we tested this method with limited resources and within the neighborhood of one of the researchers (neighbor advocates around their neighborhood), primarily to validate the idea of improving the population coverage issue of national housing and population censuses. However, the entire system sizing had been performed well in another publication [71], as well as the assessment of the network's traffic (workload) and solution deployment architecture. To determine the peak workload value for the system, an estimation was made based on the Pakistan Census facts from 2017 and 2023. With 121,000 enumerators working 8 h a day for 30 days to enumerate 32,185,605 households, each enumerator collecting data on an average of 6.39 individuals per household, the total number of data points to be recorded would reach 237,752,338. Additionally, considering one thumbprint collected and stored on the IPFS for each household representative, the system's workload would involve handling approximately 16,511 transactions per minute or 275 transactions per second. This calculation ensured that the system would be capable of efficiently processing the substantial volume of data generated during the 30-day census period. To accommodate this workload, the system was designed for horizontal scalability by adding peer nodes as needed, with load-balancing mechanisms to distribute traffic evenly and ensure optimal performance based on hardware, software, and data complexities [71].

10. Discussion

Blockchain technology has been explored in the government sector for various use cases, but the enumeration process of population census has never been in the spotlight. In this study, a systematic literature review was performed on the existing systems for population censuses and identified quite a lot of limitations and drawbacks. There remains a noticeable lag in adopting technologies for the efficiency and security of extensive operations like census and enumeration processes. An opportunity exists for substantial improvements that can enhance the results, diminish political influences, and lower the overall financial resource burden of said processes. Countries around the world have the potential to implement more effective measures for data collection, data management, and enhancing the coverage of population census efforts. Public blockchain networks, like Ethereum, are often unsuitable for enterprise applications due to several factors. These include their limited throughput, transaction fees, inefficient consensus protocols, and an inability to store sensitive data securely [72]. In contrast, permissioned blockchain frameworks empower enterprises to establish private or consortium networks, fostering secure and trustless collaborations. The Hyperledger Fabric framework best suits this research and has a track record of identity management use cases.

By leveraging blockchain technology in the census enumeration process, this research aimed to address missed population coverage and ethnic and racial discrimination and enhance data security, transparency, and efficiency. This research also highlighted the potential for blockchain technology to revolutionize traditional census processes, providing a more secure and robust data collection and management solution. However, implementing such a complex infrastructure requires careful planning, investment, and stakeholder cooperation to ensure its successful deployment and integration in the census enumeration process.

Although collecting anonymized census data for research is tedious, the software feasibility test and blockchain network performance results are rejuvenating and satisfactory upon completion. With the help of three NGOs and three volunteer surveyors, we were able to conduct the successful execution of the prototype, the software, and the blockchain network. The proposed system was then evaluated at four stages, from architecture design and prototyping to software usability and blockchain performance.

Validating the national population census results through a prototype faced challenges due to Bureau of Census's policy changes, notably the discontinuation of block-level data publication. Additional survey questions revealed significant undercounting, with evidence suggesting that up to 29% of the population in certain areas, including marginalized groups and non-ID card holders, were potentially missed using traditional methods. Extrapolating these findings suggests a considerable underrepresentation at a national level, possibly exceeding 10 million individuals, particularly in densely populated areas. This underscores the need for the Bureau of Census to revise its methodologies, ensuring more inclusive and accurate data collection strategies which accommodate diverse population segments and their unique circumstances.

Pakistan's National Bureau of Census conducted its first digital census with the collaboration of the National Database and Registration Authority (NADRA), which took care of the technological solutions behind the census [73,74]. In general, NADRA issues computerized national identity cards to the citizens of Pakistan, manages sensitive information (i.e., thumbprints of the index finger and a thumb from both hands) in government databases, and safeguards national identities from theft. NADRA plays a crucial role in identity protection. A national identity card is the first thing enumerators ask during the conduction of a national housing and population census. The Census Bureau is responsible for covering data collection in the houses of identity card holders; the major limitation of the Bureau, therefore, was its ignorance of homeless/minority/discriminated communities, refugees, and non-identity card holders (referred to as missing persons in the census count). We proposed thumbprint collection in the census process—only mandatory for non-ID card holders, so that they could be verified by the NADRA. In contrast, if their records did not exist, these people would provide a corner case for policymakers to analyze such cases and propose some form of identity to have them listed in the official database. In the future, the verification of thumbprint could be possible on the go, and the usage and cost of the IPFS would be minimized as well.

In terms of managing user identity, an encryption method has been applied in our model to protect identity data (national ID card and thumbprint) both at rest and in transit. This ensures that, even if data are intercepted, they remain unreadable and secure. In the future, either a national ID card or a thumbprint could potentially be used to establish an SSI model. In our study, we collected mobile contact information and email addresses so that multi-factor authentication could be applied to verify the identity of individuals participating in the census. This reduced the risk of impersonation or duplicate entries. Privacy risks such as identify theft and data breaches still exist, even when strong encryption applied. There are a few strategies that might help in mitigating these risks, such as the following:

- Zero-Knowledge Proofs (ZKP): It allows one party to prove to another that a statement is true without revealing any information beyond the validity of the statement itself. Implementing ZKP in a blockchain-based census system can enable individuals to verify their identity or other sensitive information without exposing the actual data. This approach is particularly valuable in census data collection, ensuring data privacy and security while maintaining accuracy [75,76].
- Homomorphic Encryption: This form of encryption allows computations to be carried out on ciphered text, generating an encrypted result which, when decrypted, matches the result of operations performed on the plain text. It enables the processing of encrypted census data without needing to decrypt it, thus maintaining data privacy throughout the analysis phase [77–79].
- Private Smart Contracts: These are smart contracts that maintain the privacy of the contract's data and state. Implementing this could ensure that the census data collected and processed on the blockchain remain confidential and only accessible to authorized entities [80,81].
- Self-Sovereign Identity (SSI): SSI is a digital identity that gives individuals control over the storage and management of their personal data. By incorporating SSI, individuals can manage their demographic data and consent to their use in the census. This not only empowers citizens but also enhances the security and accuracy of the census data, as individuals are more likely to provide accurate information if they retain control over them [82–84].
- Decentralized Identifiers (DIDs): DIDs are a new type of identifier that enables verifiable, self-sovereign digital identities. They can be used to create a secure and verifiable identity system for census data collection, ensuring that each individual's data are accurately linked to them without exposing personal information [85,86].

With these strategies, a blockchain-based census system can effectively manage user identities while mitigating privacy risks, thus significantly strengthening the security and privacy aspects of the system. This system will be aligned with global trends towards greater data protection and individual privacy rights, potentially increasing the acceptance and trustworthiness of the system among the public and stakeholders. The key to this is to strike a balance between accurate and efficient data collection and the protection of an individual's privacy rights.

Managing data provenance in a blockchain-based census system is crucial to ensuring the integrity, transparency, and trustworthiness of the data collected. Data provenance refers to the record of the origins, custody, and transformations of the data. In the context of blockchain technology, this means tracking and documenting every interaction with the data, from their initial entry into the system to their final state, including any updates or changes. Blockchain's inherent structure serves as a natural provenance tool. Its characteristics, such as an (i) immutable ledger where each block in the chain contains timestamp and transaction data, which provide an immutable record of all interactions with the data, and (ii) the traceability of the blockchain ledger allow for the complete traceability of all data transactions. This means that any changes, access, or updates to the census data can be traced back to their origin. The other way of handling this is with smart contracts. Smart contracts can be programmed to automatically record data interactions on the blockchain. For instance, when census data are updated, the smart contract can log the transaction details, including who made the change and when. A challenge may arise in that managing data with extensive provenances on a blockchain can lead to scalability issues.

Nonetheless, there are a few recommendations for safeguarding and data strategies that could be combined with a blockchain census system to resolve missing persons issues more completely. The key to this is taking a multifaceted approach, combining technology, data integration, operational procedures, and governance. These recommendations are the results of the discussions carried out during our qualitative research and during the design and testing of the prototype and overall blockchain experience.

- Continuous Census Data Updating: Rather than just conducting a census at fixed intervals, integrate a more continuous updating of data using civil registration systems, surveys, administrative records, and individuals' own reports. This keeps the data more dynamic and accurate.
- Cross-linking of Datasets: In addition to linking census data on-chain, link them to other official government datasets like birth/death certificates, medical records, tax records, social services, etc. This provides a more holistic view of individuals and families, making it harder for people to fall through the cracks.
- Privacy-preserving Record Linkage: Use statistical techniques to link individuals across datasets in a privacy-preserving way. This allows data to be integrated without compromising individuals' privacy with a public, transparent blockchain.
- Offline Backups and Versioning: Maintain offline backups of census data in case the blockchain is disrupted or provides an additional level of security. Store historic versions of the data to allow for the auditing and resolution of errors.
- Verification Procedures: Implement robust verification procedures to ensure that individuals are counted correctly and that the data associated with them are accurate. This could include cross-referencing other data sources, audits, and a process for individuals to report errors.
- Outreach Programs: Create proactive outreach programs to identify and count marginalized groups that may be underrepresented in a census. Partner with local communities and nonprofits to improve coverage.
- Governance and Security: Establish proper governance and security procedures over the census data and blockchain system to prevent unauthorized access or tampering. Add built-in auditing to monitor issues.
- Testing and Auditing: Conduct regular testing, simulations, and auditing of the census blockchain system to identify weaknesses and make continuous improvements to the system and procedures.
- Individual Ownership and Access: Give individuals access to their own census records and a means to report errors or changes. This helps engage individuals and keeps the data more accurate.

On the other hand, if we talk about specific marginalized groups, a few effective measures can ensure that these groups are counted during the census process.

- Targeted Outreach: Conduct targeted outreach campaigns specifically focused on marginalized groups like homeless populations, migrants, refugees, indigenous groups, etc. This could include on-the-ground outreach workers, partnerships with local nonprofits and community groups, and multilingual messaging.
- Accommodate Diverse Living Situations: The census should be designed to accommodate people with diverse living situations, including those without traditional housing, stable addresses, or documentation. Allow options to report temporary addresses, cross-streets, or describe locations, and accept reports from shelter administrators or outreach workers.
- Incentivize Participation: Provide incentives for marginalized groups to participate in the census, such as access to essential services, food assistance, healthcare, cash payments or credits, etc. This can help overcome distrust or a lack of motivation to engage with the census.
- Accommodate Disabilities and Limitations: The census procedure and forms should accommodate individuals with visual impairments, learning disabilities, limited literacy, or other issues that could deter participation. This could include braille, large-print and audio questionnaires, a simplified census form, and census workers with relevant training.
- Build Trust and Reassure Privacy: Conduct education campaigns to build trust in the census process and reassure individuals, especially marginalized groups, that their data will be kept private and secure and will not be used against them. Transparency about data usage and consent policies can help build confidence.

- Partner with Advocacy Groups: Census agencies should work closely with advocacy groups representing marginalized communities. They can provide input into census design, help promote participation, and reassure their communities that the census is necessary and safe to engage with.
- Provide Census Surveyors with Diversity Training: Require census takers and outreach workers to complete diversity and inclusion training so they have a better understanding of the barriers and issues faced by marginalized groups. They will be better equipped to effectively conduct outreach and properly count these populations.
- Follow Up on Undercounted Areas: Analyze census results to identify geographic areas or communities that appear to be undercounted. Then, conduct follow-up outreach to count missing individuals and work to determine the causes for the undercounting so the next census can be improved.
- Accept the Self-reporting of Demographics: Allow individuals to self-report demographic information, like ethnicity, tribe, disability status, income, etc., rather than having census takers assign categories. This will result in more accurate counts, especially for marginalized groups.

These are some of the critical ways in which a census can help ensure that even the most marginalized groups in society are correctly counted. A census has a better chance of achieving a complete count with targeted outreaches, accommodations, trust-building practices, partnerships with advocates, and follow-ups on undercounts.

From a controversial point of view, the following question can be raised: why use blockchain if these problems can be resolved or improved in a centralized system? Building a transparent and secure system for a census using centralized database and technologies is feasible, but it comes with some inherent challenges and trade-offs compared to a blockchain-based system. A comparison has been made regarding the advantages and disadvantages of both approaches in Table 7.

Although centralized databases can be simpler to set up, they may lack some of the security and transparency features that blockchain offers. We can say that the *population coverage (missed population)* issue can be improved in either approach through the centralized infrastructure of thumbprint collection devices, but, again, the decentralization, privacy, security, trust, and transparency of the system are today's public needs. *Ethnic and racial discrimination* issues in censuses can only be resolved with a blockchain-based system. Blockchain does introduce complexity but provides enhanced security, transparency, and decentralization. Overall, our proposed blockchain-based census system offers robustness, scalability, and improvements in the system as well as promotes transparency and trust in the general public so that they can participate in the census process confidently. The same results cannot be accomplished with any sort of centralized system.

Table 7. Advantages and disadvantages of centralized vs. blockchain-based approaches.

Approach	Advantages	Disadvantages
Traditional Centralized Approach	Simplicity: Centralized databases are typically easier to set up and manage compared to distributed systems. They require less complex infrastructure.	Single Point of Failure: A centralized system has a single point of failure. If the central database is compromised, all the data are at risk.
	Performance: In certain scenarios, centralized databases may offer better performance for reading and writing operations since all the data are stored in one location.	Data Integrity: With a single database, ensuring data integrity can be challenging. Any corruption or tampering could have widespread effects.
	Control: Centralized systems provide a single point of control and authority, making it easier for administrators to manage access and permissions.	Security Concerns: Centralized databases are more susceptible to targeted attacks. Once breached, an attacker can potentially gain access to all sensitive information.
		Lack of Trust: Centralized systems may face issues related to trust. Users may be concerned about how their data are being handled and whether they are being misused.

Table 7. *Cont.*

Approach	Advantages	Disadvantages
Blockchain-based Approach	Decentralization: Blockchain operates on a decentralized network, reducing the risk of a single point of failure. Data are distributed across multiple nodes.	Complexity: Implementing and managing a blockchain system can be more complex than a centralized database, requiring specialized knowledge.
	Security: The immutability of blockchain ensures that once data are recorded, they cannot be easily altered or tampered with, enhancing data security.	Regulatory Challenges: There may be regulatory uncertainties and challenges associated with the use of blockchain technology in certain jurisdictions.
	Transparency: The transparent nature of blockchain allows participants to trace and verify transactions, promoting trust and accountability.	
	Privacy: Advanced encryption techniques in blockchain systems can provide a high level of privacy and data protection.	

11. Conclusions

Blockchain technology offers promising potential to help address some key challenges in conducting an effective national population census. The decentralized and distributed nature of blockchain could enable diverse humanitarian entities and government to coordinate and collaborate securely on a shared platform. This could assist in improving population coverage, especially reaching marginalized communities through trusted non-profit partners. Blockchain's inherent verification mechanisms via distributed consensus could also strengthen the reliability of census data collection from multiple sources. Individual identity management on blockchain makes it feasible to build an accurate database covering all residents across jurisdictions while securing their privacy. If thoughtfully designed with appropriate permissions, such a robust digital census database could be invaluable for policymaking bodies and researchers to gain rapid insights for social planning. However, blockchain is not a silver bullet—factors like infrastructure readiness, legal frameworks, blockchain network design tailored for census needs, and community participation remain pivotal for its successful adoption. With careful implementation, blockchain's decentralization and cryptographic security properties could provide a foundational layer for the complex mechanics of a national census to thrive.

Our research has made significant contributions to the field, enhancing the body of knowledge in several key areas related to national population censuses and the innovative application of blockchain technology. Firstly, we developed a comprehensive knowledge bank addressing the global challenges encountered in conducting national population censuses. Our work also highlights the pivotal role of NGOs in achieving accurate population coverage, particularly for marginalized, deprived, and homeless communities.

Our major contribution to the research is the development of a blockchain-based system that mitigates the missing population problem that occurs in existing national housing and population censuses. We presented a case where NGOs collaborate with the National Census Bureau (cross-agency collaboration), enabling a more holistic approach to identifying and addressing missing person challenges by leveraging data from various sources. Our research highlights the fact that blockchain's decentralized and immutable nature ensures the reduction in the risk of data loss (even in the events of natural disasters or system failures) or manipulation and minimizes the chances of missing persons occurring due to accidental errors or deliberate tampering. Blockchain's transparency brings trust in the census process and encourages individuals to participate, reducing the likelihood of intentional avoidance or underreporting. With secure and private data handling, individuals may feel more confident that their personal information is being protected, encouraging their active participation and reducing their concerns about privacy breaches. Using biometric data, such as thumbprints, on the blockchain reduces the chances of duplicate records and aids governments in accurately identifying and counting individuals missing identity documents or from discriminated communities, reducing the risk of missing persons.

Our research has various sub-contributions, i.e., the development of a blockchain-based system architecture for housing and population censuses and the collection of thumbprint data. This architecture is not just theoretical; we implemented a proof-of-concept analysis and conducted pilot testing to assess the system's functionality and the performance of the underlying blockchain network. An innovative aspect of our approach involves using fingerprint devices to identify individuals who lack traditional identity cards, such as immigrants, refugees, and the homeless, ensuring their inclusion in census data.

Our study contributes significantly to understanding how blockchain technology can be utilized for the betterment of humanity. It opens new avenues for future research in blockchains, particularly focusing on Hyperledger Fabric, chain code, and Distributed Ledger Technology (DLT) in the context of social good. Additionally, our research provides insights into the optimization and monetization of census practices, offering a more efficient and economically viable approach to population data collection.

Furthermore, our work serves as a gateway for social interventions aimed at improving the quality of life of nations. By leveraging advanced technologies like blockchain, we demonstrate the potential for more accurate, inclusive, and humane approaches to census-taking, ultimately contributing to better-informed policy-making and resource allocation for societal benefit.

12. Future Directions

We will be exploring the integration of advanced privacy-preserving technologies, such as ZKP, SSI, DIDs, and homomorphic encryption. The primary researcher is collaborating with the National Census Bureau of Pakistan on people's awareness of blockchain technology, providing consultation on its seamless integration with existing government databases and systems. We need to ponder upon the incentive mechanisms for citizens, NGOs, and agencies to further encourage their active participation in order to reduce the likelihood of intentional avoidance. Another important aspect we will be working on is the development of blockchain standards specifically tailored for census applications as well as the integration of machine learning algorithms for predictive analytics and trend analysis based on census data. Models card tools will be considered for the transparent documentation of machine learning models. They can be used to document the algorithms and models used for data analysis and prediction, and they can detail the data inputs, model training processes, versioning, performance metrics, and any biases identified in the models. The exploration of the implementation of a decentralized governance model for decision-making within the census blockchain network described in this study is another dimension we will be looking into.

Author Contributions: Writing—original draft, S.R.; Writing—review & editing, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The questionnaire of the structured interviews is provided in Appendix A. The data analysis of the structured interviews is published in the IEEE conference [48].

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The list of abbreviations used in the paper is as follows: AES—Advanced Encryption Standards; AFIS—Automated Fingerprint Identification Systems; AWS—Amazon Web Services; BINC-TAPS—Blockchain-Based Implementation of National Census as a Supplementary Instrument for Enhanced; Transparency, Accountability, Privacy, and Security; BFT—Byzantine Fault Tolerance; CAA—Citizenship Amendment Act; CD-ROM—Compact Disc Read-Only Memory; CSV—Comma-separated Values; DIDs—Decentralized Identifiers; DLT—Distributed Ledger Technology; DVD—Digital Video Disc; ETH—Ethereum; ID—Identification; IPFS—InterPlanetary File System; IT—Information Technology;

KYC—Know Your Customers; MSP—Membership Service Provider; NGO—Non-government Organization; NADRA—National Database and Registration Authority; NPO—Non-Profit Organizations; NRFU—Non-response Follow-up; P2P—Peer-to-peer; PII—Personally Identifiable Information; SEC—Securities and Exchange Commission; SHA—Secure Hash Algorithm; SMS—Short Message/Messaging Service; SSI—Self-Sovereign Identity; TPS—Transaction Per Second; UK—United Kingdom; UN—United Nations; US—United States; VAT—Value-added Tax.

Appendix A

Questionnaire—Target Audience: Bureau of National Population Census

Section 1. About You:

1. What is your designation/position in the organization? (text)
2. How long have you been working with the Census Bureau? (9 years or high/7+ years/5+ years/3+ years/2 years or less)

Section 2. Existing Population Census System:

3. Is the current system immutable? (Yes/No)

* immutable means unable to change/alter once a record is entered in the system, even from the inspector/supervisor.

4. How reliable is the current population census system? (Yes/No) * reliable means totally free of technical errors; consistently performs according to its specifications.
5. Do you share the population census data with other entities like government departments or private companies? (Yes/No)
    5.1. If yes, what kind of organizations (text)
    5.2. Why do they need the data? (text)
6. Are the data easily accessible to these organizations/government departments when needed? (Yes/No)
7. How do they access the data? (direct access via portal/assigned login accounts/Dropbox-Email File Transfer/via USB drive/obtain data via CD drive)
8. Is the current system secured from data leakage and authorized access? (Yes/No) * secure means free from or not exposed to danger or harm; safe.
9. Is the current system recoverable? (Yes/No) * recoverable means to be restored to a normal or usual condition.
10. Do you take complete ownership/responsibility of your own collected data points only and do not rely on 3rd party datasets? (Yes/No) * means are you only responsible for your own collected data points?
11. Do you consider any reliable open-source network data to cover the missing data? (Yes/No)

* e.g., NGO data are available for people living in hand-to-mouth circumstances, would you consider this and ensure your data coverage?

Section 3. About Missing Population—Futuristic Approach/System

12. Research says that people who are not living in houses (homeless communities, under the bridge, on footpaths, jail, asylum, etc.), transgender, orphan houses, hospital patients, and victims of ethnic/racial discrimination are missed from the census population counting, do you agree with it? (Strongly Agree, Agrees, Neutral, Disagree, Strongly Disagree)
    12.1. If disagree: Please explain (text)
13. Do you use any alternative way (solution to integrate them in census population counting? (Yes/No)
    13.1. If yes, please state alternative way(solution): (text)
14. Do you have any plans to integrate homeless and transgender, travelers' communities, orphan houses, and hospitals into the current census population system? (Yes/No)

15. If Yes or No, please explain?
16. Would you consider reliable open-source network data to cover the missing person from NGOs who provide aid to striving communities?
17. Which measure have you taken to achieve this goal? (text)
18. Anything you like to share regarding this problem? (text)

Section 4. About Blockchain:

19. Are you familiar with distributed ledger (DL) database (blockchain technology) and what benefits it holds? (Yes/No) * The distributed ledger database is spread across several nodes (devices) on a peer-to-peer network, where each replicate and saves an identical copy of the ledger and updates itself independently. The primary advantage is the lack of central authority.
    19.1. If yes, do you think a blockchain is better than a centralized system to address the missing persons issue? (Yes/No)
    19.2. If yes, in what ways have you ever considered permissioned distributed ledger (private blockchain) technology for the enhancement of census population system? (text)
20. Do you think that DL/blockchain can help to solve the missing person problem? (text)

Section 5. Open Question:

21. Any risks/challenges/security issues/blockers/pain points you would like to share regarding existing population census system? (text)
22. What do you suggest for improving the population census in the future so it would be more inclusive for all and technologically better? (text)
23. What functionality do you think is more appropriate to have in the census system? (text)

References

1. United Nations. *Principles and Recommendations for Population and Housing Censuses*; Department of Economic and Social Affairs Statistics Division: New York, NY, USA, 2017.
2. United States Congress. Is Brooklyn Being Counted? Problems with the 2010 Census. 2010, Serial No. 111–101. Available online: https://www.congress.gov/event/111th-congress/house-event/LC6955/text (accessed on 17 December 2023).
3. Swaroop, S.R. *Truth about Muslim Population Explosion in India: Evidence from Census 2011*; Amazon Asia-Pacific Holdings Private Limited: Mumbai, India, 2022.
4. Consumer Affairs. Yet Another U.S. Government Cybersecurity Breach; This Time, It's the Census Bureau. 2017. Available online: https://www.consumeraffairs.com/news/yet-another-us-government-cybersecurity-breach-this-time-its-the-census-bureau-072415.html (accessed on 20 January 2020).
5. Karim, M. Missing People in Census. Daily DAWN, 17 October 2017. Available online: https://www.dawn.com/news/1364371 (accessed on 23 July 2018).
6. Anderson, M.J. *The American Census: A Social History*, 2nd ed.; Yale University Press: New Haven, CT, USA, 2015.
7. Juran, S.; Pistiner, A. The 2010 Round of Population and Housing Censuses (2005–2014). *Stat. J. IOAS* **2017**, *33*, 399–406. [CrossRef]
8. Nordholt, E.S. *Efficiency in Population Censuses: The Situation of European Register-Based 2011 Censuses*; Statistics: The Hague, The Netherlands, 2012.
9. Fienberg, S.; Kenneth, P. Save Your Census. *Nature* **2010**, *466*, 1043. [CrossRef] [PubMed]
10. Rasheed, S.; Louca, S. Exploring the need for Blockchain-based national population census. In *Lecture Notes in Business Information Processing*; EMCIS 2021; Springer: Berlin/Heidelberg, Germany, 2021; Volume 437. [CrossRef]
11. CBS News. Available online: https://www.cbsnews.com/news/2010-census-missed-15-million-minorities/ (accessed on 23 July 2018).
12. Hogan, H. The Accuracy and Coverage Evaluation: Theory and Design. *Surv. Methodol.* **2003**, *29*, 129–138.
13. 'We Are Still Here': Native Americans Fight to Be Counted in US Census. Guardian 2020. Available online: https://www.theguardian.com/us-news/2020/jan/15/we-are-still-here-native-americans-fight-to-be-counted-in-us-census (accessed on 10 January 2020).
14. Crossette, B. For Census-Takers Worldwide, 2020 Could Be a Rough Year. Available online: https://www.passblue.com/2019/12/21/for-census-takers-worldwide-2020-could-be-a-rough-year/ (accessed on 2 January 2020).
15. Sikhs to Be Counted as Separate Ethnic Group in US Census for First Time. The News 2020. Available online: https://www.thenews.com.pk/print/599780-sikhs-to-be-counted-as-separate-ethnic-group-in-us-census-for-first-time (accessed on 15 January 2020).

16. In a First, Census to Categorise Data on Households Run by Transgenders Separately. The Hindu—Business Line 2020. Available online: https://www.thehindubusinessline.com/news/national/in-a-first-census-to-categorise-data-on-households-run-by-transgenders-separately/article30592094.ece (accessed on 15 January 2020).
17. Norton, T.S. The Census Is Too Important to Boycott, Despite Serious Privacy Concerns. Available online: http://www.smh.com.au/comment/the-census-is-too-important-to-boycott-despite-privacy-concerns-20160804-gqllvs.html (accessed on 15 December 2019).
18. Why Germans Are So Private about Their Data. Available online: https://www.handelsblatt.com/today/handelsblatt-explains-why-germans-are-so-private-about-their-data/23572446.html?ticket=ST-1119959-eZWdKzPuhLf4ofbXVloa-ap4 (accessed on 2 January 2020).
19. Dunn, C.S.; Austin, E.W. *Protecting Confidentiality of Archival Data Resources*; IASSIST Quarterly, University of Alberta: Edmonton, AB, Canada, 1998; Volume 2. [CrossRef]
20. Skinner, C. *Annual Review of Statistics and Its Application*; Department of Statistics, London School of Economics and Political Sciences: London, UK, 2015.
21. Frey, W.H. *Investigating Change: Web-Based Analyses of US Census and American Community Survey Data*, 3rd ed.; Cengage Learning: Boston, MA, USA, 2011.
22. Bair, R.; Torrey, B.B. *The Challenges of Census Taking in Developing Countries*; Government Information Quarterly; ELSEVIER: Amsterdam, The Netherlands, 2000; Volume 2, pp. 433–452. [CrossRef]
23. Makridakis, S.; Polemitis, A.; Giaglis, G.; Louca, S. Blockchain: The next breakthrough in the rapid progress of AI. *Artif. Intell. Emerg. Trends Appl.* **2018**, *10*, 197–219.
24. Weinstein, S. Blockchain Neutrality. *Ga. Law Rev.* **2021**, *55*, 499–591.
25. Zafar, M.A.; Sher, F.; Janjua, M.U.; Baset, S. Sol2js: Translating Solidity Contracts into Javascript for Hyperledger Fabric. In Proceedings of the 2nd Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers, Rennes, France, 10 December 2018; ACM: New York, NY, USA, 2018; pp. 19–24. [CrossRef]
26. Dabbagh, M.; Kakavand, M.; Tahir, M. Towards Integration of Blockchain and IoT: A Bibliometric Analysis of State-of-the-Art. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 1010, pp. 27–35.
27. Drescher, D. *Blockchain Basics: A Non-Technical Introduction in 25 Steps*, 1st ed.; Apress: Berkeley, CA, USA, 2017. [CrossRef]
28. Viriyasitavat, W.; Hoonsopon, D. Blockchain characteristics and consensus in modern business processes. *J. Ind. Inf. Integr.* **2019**, *13*, 32–39. [CrossRef]
29. Scheller, F.; Reichelt, D.; Dienst, S.; Johanning, S.; Reichardt, S.; Bruckner, T. Effects of implementing decentralized business models at a neighborhood energy system level: A model based cross-sectoral analysis. In Proceedings of the 2017 14th International Conference on the European Energy Market (EEM), Dresden, Germany, 6–9 June 2017; pp. 1–6.
30. Scheller, F.; Johanning, S.; Seim, S.; Schuchardt, K.; Krone, J.; Haberland, R.; Bruckner, T. Legal Framework of Decentralized Energy Business Models in Germany: Challenges and Opportunities for Municipal Utilities. *J. Energy Ind.* **2018**, *42*, 207–223. [CrossRef]
31. Chen, Y.; Bellavitis, C. Blockchain disruption and decentralized finance: The rise of decentralized business models. *J. Bus. Ventur. Insights* **2020**, *13*, e00151. [CrossRef]
32. Yanovich, Y.; Shiyanov, I.; Myaldzin, T.; Prokhorov, I.; Korepanova, D.; Vorobyov, S. Blockchain-Based Supply Chain for Postage Stamps. *Informatics* **2018**, *5*, 42. [CrossRef]
33. Themistocleous, M.; Christodoulou, K.; Iosif, E.; Louca, S.; Tseas, D. Blockchain in academia: Where do we stand and where do we go? In Proceedings of the 53rd Hawaii International Conference on System Sciences, Maui, HI, USA, 7–10 January 2020. [CrossRef]
34. Ayub Khan, A.; Laghari, A.A.; Shaikh, A.A.; Bourouis, S.; Mamlouk, A.M.; Alshazly, H. Educational Blockchain: A Secure Degree Attestation and Verification Traceability Architecture for Higher Education Commission. *Appl. Sci.* **2021**, *11*, 10917. [CrossRef]
35. Sweden Tests Blockchain Smart Contracts for Land Registry. Cointelegraph. Available online: https://cointelegraph.com/news/sweden-tests-blockchain-smart-contracts-for-land-registry (accessed on 12 November 2023).
36. Brazilian Voting Authority Might Include Blockchain Tech in Future Elections. Available online: https://news.bitcoin.com/brazilian-voting-authority-might-include-blockchain-tech-in-future-elections/ (accessed on 14 November 2023).
37. Estonia Leads World in Making Digital Voting a Reality, Financial Time. Available online: https://www.ft.com/content/b4425338-6207-49a0-bbfb-6ae5460fc1c1 (accessed on 15 November 2023).
38. The United Arab Emirates' Government Portal, Blockchain in the UAE Government. Available online: https://u.ae/en/about-the-uae/digital-uae/digital-technology/blockchain-in-the-uae-government (accessed on 15 November 2023).
39. UK Tax Agency to Pay $100K for Blockchain Surveillance Software. Available online: https://cointelegraph.com/news/uk-tax-authority-to-pay-100k-for-blockchain-surveillance-software (accessed on 12 November 2023).
40. World Food Program—Building Blocks. Available online: https://innovation.wfp.org/project/building-blocks (accessed on 17 December 2023).
41. How Regtech Can Transform Your Regulatory Compliance. Available online: https://legal.thomsonreuters.com/en/insights/articles/how-regtech-can-transform-your-regulatory-compliance (accessed on 24 November 2023).
42. U.S. SEC Approves New U.S. Exchange with Blockchain Feed, Faster Settlement. Available online: https://www.sec.gov/page/blockchain (accessed on 24 November 2023).

43. Emirates Blockchain Strategy 2021. Available online: https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/strategies-plans-and-visions-untill-2021/emirates-blockchain-strategy-2021 (accessed on 24 November 2022).
44. Tirodkar, V.; Patil, S. Proposed Infrastructure for Census Enumeration and Internet Voting Application in Digital India with Multichain Blockchain. In *Advanced Computing Technologies and Applications—ICACTA*; Springer: Berlin/Heidelberg, Germany, 2020. [CrossRef]
45. World Population Review. Available online: https://worldpopulationreview.com/country-rankings/literacy-rate-by-country (accessed on 25 December 2023).
46. Srivastava, S. Improving Digital Literacy in India—A Review, ITU. 2022. Available online: https://www.itu.int/hub/2022/12/improving-digital-literacy-in-india-a-review/ (accessed on 25 December 2023).
47. Procivis. Available online: https://www.procivis.ch/en/home (accessed on 23 August 2023).
48. Rasheed, S.; Louca, S. Data analysis & core findings of Pakistan national population census survey. In *Lecture Notes in Network and Systems*; BLOCKCHAIN 2023; Springer: Berlin/Heidelberg, Germany, 2023; Volume 778, pp. 144–153. [CrossRef]
49. Rasheed, S.; Louca, S. A Blockchain-based Architecture for National Population Census. In Proceedings of the 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Spain, 19–21 July 2023. [CrossRef]
50. Open Source Blockchain Effort for the Enterprise Elects Leadership Positions and Gains New Investments. Hyperledger, 2016. Available online: https://www.hyperledger.org/announcements/2016/03/29/open-source-blockchain-effort-for-the-enterprise-elects-leadership-positions-and-gains-new-investments (accessed on 23 December 2019).
51. Hyperledger Fabric. Available online: https://www.hyperledger.org/use/fabric (accessed on 23 December 2019).
52. Paul, M.S. Hyperledger Frameworks and Modules. Medium 2018. Available online: https://medium.com/swlh/hyperledger-chapter-2-hyperledger-frameworks-modules-cabf50e12105 (accessed on 10 December 2019).
53. Polge, J.; Robert, J.; Traon, Y.L. *Permissioned Blockchain Frameworks in the Industry: A Comparison*; ICT Express: Amsterdam, The Netherlands, 2021; Volume 7, pp. 229–233; ISSN 2405-9595. [CrossRef]
54. Capocasale, V.; Gotta, D.; Perboli, G. Comparative analysis of permissioned blockchain frameworks for industrial applications. In *Blockchain: Research and Applications*; Elsevier: Amsterdam, The Netherlands, 2023; Volume 4; ISSN 2096-7209. [CrossRef]
55. Soelman, M. Permissioned Blockchains: A Comparative Study—A Deep Dive into Hyperledger Fabric and Hyperledger Besu. Master's Thesis, University of Groningen, Groningen, The Netherlands, 2021.
56. Daria, R.; Dmytro, H. Top 6 Blockchain Frameworks to Build Your App. Ruby Garage 2020. Available online: https://rubygarage.org/blog/best-blockchain-frameworks (accessed on 17 December 2023).
57. Mouha, N. Review of the Advanced Encryption Standard, NIST—U.S. Department of Commerce. 2021. Available online: https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8319.pdf (accessed on 26 December 2023).
58. Census Survey Form—2017. Available online: https://www.pbs.gov.pk/sites/default/files/population/short_form_2.jpg (accessed on 13 January 2023).
59. FS88H FIPS201/PIV Compliant USB2.0 Fingerprint Scanner. Available online: https://www.futronic-tech.com/prodetail.php?pro_id=1535 (accessed on 17 August 2023).
60. Saylani Welfare. Available online: https://saylaniwelfareusa.com/en (accessed on 3 August 2023).
61. Pakistan Bait-ul-Mal. Available online: https://www.pbm.gov.pk/ (accessed on 3 August 2023).
62. Customs Healthcare Society. Available online: https://www.customshealthcaresociety.org/ (accessed on 3 August 2023).
63. Abbas, G. First-Ever Digital Census Begins Tomorrow, Pakistan Today. 2023. Available online: https://www.pakistantoday.com.pk/2023/02/27/121000-field-enumerators-deployed-as-first-ever-digital-census-begins-on-march-1/ (accessed on 3 April 2023).
64. AlHasan, R. Bengalis in Pakistan: A Neglected Community Crying for Recognition, South Asia Monitor. 2021. Available online: https://www.southasiamonitor.org/spotlight/bengalis-pakistan-neglected-community-crying-recognition (accessed on 17 December 2023).
65. Maryam, H. Stateless and Helpless: The Plight of Ethnic Bengalis in Pakistan, Aljazeera. 2021. Available online: https://www.aljazeera.com/features/2021/9/29/stateless-ethnic-bengalis-pakistan (accessed on 17 December 2023).
66. Reuters, Don't We Count? Transgender Pakistanis Feel Sidelined by Census. Available online: https://www.reuters.com/article/pakistan-transgender-census-idUSL8N1ME398/ (accessed on 22 November 2023).
67. Azam, O. A Transgender-Blind Census, Geo New. Available online: https://www.geo.tv/latest/487921-a-transgender-blind-census (accessed on 22 November 2023).
68. Religion in Pakistan, Country Meters. Available online: https://countrymeters.info/en/Pakistan#google_vignette (accessed on 22 November 2023).
69. Amazon Lightsail, Amazon Web Services. Available online: https://aws.amazon.com/lightsail/ (accessed on 27 April 2023).
70. IPFS, Infura. Available online: https://docs.infura.io/networks/ipfs (accessed on 27 April 2023).
71. Rasheed, S.; Louca, S. Architecture Design & System Sizing of a Blockchain-based National Population Census. In Proceedings of the 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering ICECCME, Tenerife, Spain, 19–21 July 2023. [CrossRef]
72. Khan, S.N.; Loukil, F.; Ghedira-Guegan, C.; Benkhelifa, E.; Bani-Hani, A. Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 2901–2925. [CrossRef] [PubMed]

73. Digital Census, NADRA Pakistan. Available online: https://www.nadra.gov.pk/local-projects/social-protection/digital-census/ (accessed on 26 December 2023).
74. Pakistan's First-Ever Digital Census Proves to Be a Massive Success, ProPakistani. Available online: https://propakistani.pk/2023/02/27/pakistans-first-ever-digital-census-proves-to-be-a-massive-success/ (accessed on 26 December 2023).
75. Sun, X.; Yu, F.R.; Zhang, P.; Sun, Z.; Xie, W.; Peng, X. A Survey on Zero-Knowledge Proof in Blockchain. *IEEE Netw.* **2021**, *35*, 198–205. [CrossRef]
76. Hasan, J. Overview and Applications of Zero Knowledge Proof (ZKP). *Int. J. Comput. Sci. Netw.* **2019**, *8*, 436. Available online: https://ijcsn.org/IJCSN-2019/8-5/Overview-and-Applications-of-Zero-Knowledge-Proof-ZKP.pdf (accessed on 17 December 2023).
77. Yi, X.; Paulet, R.; Bertino, E. Homomorphic Encryption. In *Homomorphic Encryption and Applications*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 27–46. [CrossRef]
78. Acar, A.; Aksu, H.; Uluagac, A.S.; Conti, M. A Survey on Homomorphic Encryption Schemes: Theory and Implementation. *ACM Comput. Surv.* **2018**, *51*, 79. [CrossRef]
79. Naehrig, M.; Lauter, K.; Vaikuntanathan, V. Can homomorphic encryption be practical? In Proceedings of the 3rd ACM Workshop on CCSW '11, New York, NY, USA, 21 October 2011; pp. 113–124. [CrossRef]
80. Steffen, S.; Bichsel, B.; Baumgartner, R.; Vechev, M. ZeeStar: Private Smart Contracts by Homomorphic Encryption and Zero-knowledge Proofs. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–26 May 2022; pp. 179–197. [CrossRef]
81. Kerber, T.; Kiayias, A.; Kohlweiss, M. KACHINA—Foundations of Private Smart Contracts. In Proceedings of the IEEE 34th Computer Security Foundations Symposium (CSF), Haifa, Israel, 7–10 August 2021; pp. 1–16. [CrossRef]
82. Pöhn, D.; Grabatin, M.; Hommel, W. eID and Self-Sovereign Identity Usage: An Overview. *Electronics* **2021**, *10*, 2811. [CrossRef]
83. Laatikainen, G.; Kolehmainen, T.; Abrahamsson, P. *Self-Sovereign Identity Ecosystems: Benefits and Challenges*; Parmiggiani, E., Kempton, A., Mikalef, P., Eds.; SCIS: Schiller Park, IL, USA, 2021. Available online: https://aisel.aisnet.org/scis2021/10/ (accessed on 17 December 2023).
84. Stokkink, Q.; Ishmaev, G.; Epema, D.; Pouwelse, J. A Truly Self-Sovereign Identity System. In Proceedings of the IEEE 46th Conference on Local Computer Networks (LCN), Edmonton, AB, Canada, 4–7 October 2021; pp. 1–8. [CrossRef]
85. Alzahrani, B. An Information-Centric Networking Based Registry for Decentralized Identifiers and Verifiable Credentials. *IEEE Access* **2020**, *8*, 137198–137208. [CrossRef]
86. Decentralized Identifiers (DIDs) v1.0—Core Architecture, Data Model, and Representations. World Wide Web Consortium (W3C), July 2022. Available online: https://www.w3.org/TR/did-core/ (accessed on 26 December 2023).