Enumera and the Open Census Protocol

(Litepaper v1.1)

Abstract. A census that offers individuals the opportunity to decide what information they want to provide without compromising their personal privacy can encourage more individuals to submit accurate data. A concise questionnaire intended to provide accuracy and privacy that is designed and maintained by a central party limits genuine participation, as its questions are not relatable to a large number of people, and its privacy is unverifiable by participants outside of the central authority. We propose an Open Census Protocol that allows extensibility of the census questionnaire and transparency into the data collection process by using smart contracts developed on top of the Ethereum blockchain. The decentralized nature of our population data collection and storage methods increases the longevity and security of the data because tamper-evident, encrypted copies of the data are held by thousands of nodes participating in a global network. This technology allows individuals to enter their personal information safely and confidently. Furthermore, the data accuracy that this method promotes will lead to more informed allocation of federal funds, investment in social services, and entrepreneurial decision making.

1. Introduction

Every 10 years, each household in the United States receives a 10 question census survey. Formally named the Decennial Census, this survey provides a count of the nation's population and housing as mandated by the U.S. Constitution. It is conducted by the U.S. Census Bureau and is intended to be an accurate count of household residents, their ages, races, and genders. The Census Bureau also conducts the annual American Community Survey which asks for more detailed information. The American Community Survey is sent to a small percentage of the population on a rotating basis [1].

Conducting a census is no small task--it takes hundreds of thousands of people and millions of dollars in order to deliver the census to over 126 million homes [2]. It also has significant consequences for people, businesses, and legislation. Census data are used to draw and redraw congressional districts, so they affect the number of state representatives in Congress, and determine how more than $675 billion dollars in federal, state and local funds will be used to support schools, city infrastructure, health care facilities, child-care centers and senior centers [3]. In addition, the census forms the basis of countless government and academic studies that drive public policy decisions and legislation from Washington, D.C., to statehouses and city halls.

The 10 questions on the census form can be a point of contention, primarily because there are differing opinions on what the questions should or should not be. In 2020, the census will ask the citizenship status of household residents for the first time in 70 years. The decision to include this

question was made by the Justice Department. Many public opinions argue that this will deter non-citizens who are predominantly in racial minority groups from submitting the census. The American Civil Liberties Union, New York Civil Liberties Union, and Arnold & Porter filed a federal lawsuit on June 6, 2018 on behalf of immigrants' rights groups who charge that the current administration's order to include a citizenship question intentionally discriminates against immigrants and thwarts the constitutional mandate to accurately count the U.S. population [4]. It is important to note that the goal of the census is to count the total residents of the United States, not just the total citizens. The census has historically faced challenges of undercount. The addition of a citizenship status question may make this worse.

In 2020, the census will not ask the sexual orientation or gender identity of household residents. The census has never asked this question before and the Census Bureau has stated that there is no need to ask though it has been requested by the Department of Housing and Urban Development, the Centers for Medicare and Medicaid Services, and the Environmental Protection Agency, in addition to LGBTIQ+ rights activists [5]. This may change after 2030 if the Census Equality Act bill introduced on July 31, 2018 by Sens. Kamala Harris, D-Calif., and Tom Carper, D-Del. is signed into law [6]. To date, the census has included a checkbox for "unmarried partners", used as a proxy for same-sex relationships. Many opposite-sex partners accidentally marked this box with the intention to indicate they were unmarried to their partner, as the name implies. Currently, there remains "no reliable national data about how many LGBTIQ+ people live in the U.S. that can inform public policy" [7].

2. Solution

The aforementioned examples of how the census can fail to collect accurate information were part of the impetus for the creation of the Open Census Protocol. The Open Census Protocol has the following founding principles:

1. *Extensibility.* Individuals and collective communities can decide what questions they want to be asked and what questions they want to answer using a network consensus model.
2. *Anonymity.* An individual will not be identifiable by their data input.
3. *Verifiability.* The security and anonymity of the data can be viewed and verified by anyone.
4. *Decentralization.* A network of nodes collectively maintains the security and immutability of the data.

Enumera employs the Open Census Protocol to collect and store population data, as an alternative to the traditional census (figure 1). Enumera stores data on the Ethereum public blockchain, which can be viewed at https://etherscan.io/. Using digital identities from third party services such as uPort, private keys can be used to form a signature that verifies that an individual has submitted the census form. The encrypted form data is then saved to the blockchain by way of smart contracts.
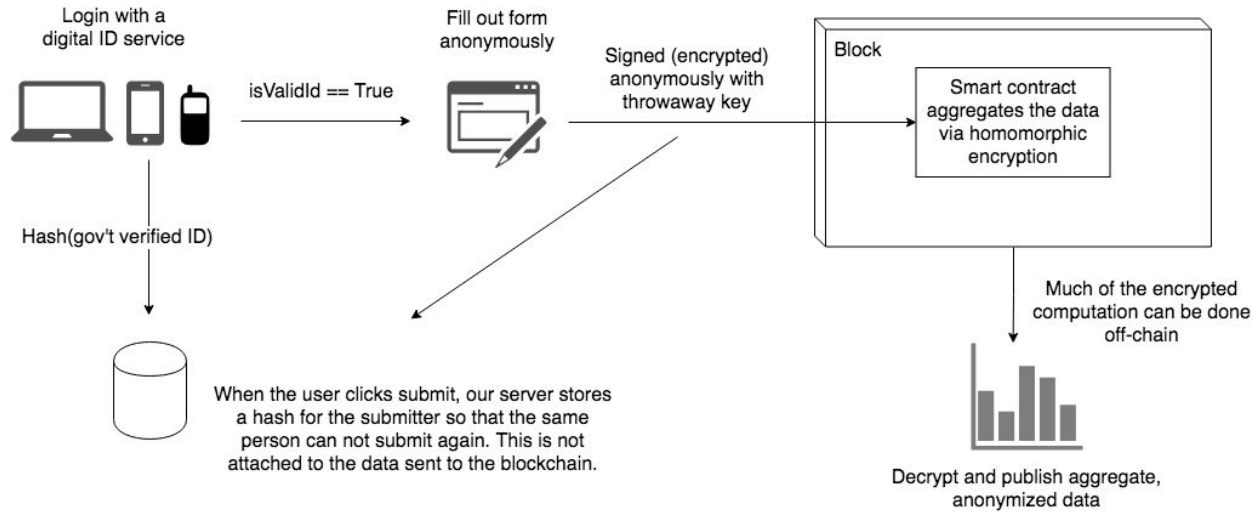
Figure 1. General technical overview of Enumera census data collection.

The Census Bureau is legally required to protect the private information of all individuals submitting the survey. It has an obligation deny all requests to access the data by anyone, including government officials and government agencies. There is currently no way for an outside party to verify that an individual's privacy is not being compromised.

The Ethereum blockchain will show that the data was securely collected; however, the raw data (for example, the answers on an individual's form submission) will not be exposed. When we decrypt the data, we will not be able to identify the submitter. Ultimately, the blockchain provides proof of data privacy without compromising that privacy.

Because the blockchain is a decentralized data store, the risk of losing the data is trivial. The thousands of nodes maintaining the blockchain will all have encrypted copies of the data. If one node loses or even alters the data, the other nodes will still have the correct copies. For this reason, bad actors attempting to tamper with the data are not a large concern. It will be easy to detect even the smallest alteration of the data. In fact, the more data stored on the chain, the more tamper-evident and tamper-resistant it becomes. Blockchain ensures data security.

In years seven, eight, and nine of a given decade, the Census Bureau's funding traditionally increases to accommodate testing and purchasing technologies, conducting a nationwide inventory of residential addresses, ordering forms, letters and advertising, and beginning to lease local offices and train temporary workers. In 2014, the U.S. Congress decreed that the 2020 census should cost no more than the 2010 census, effectively flat-lining funding [7]. This resulted in the bureau replacing workers by adopting software solutions. Simultaneously, the bureau aborted plans to test Spanish-language surveys, outreach strategies to encourage census participation, and new methods to more accurately count people in remote and rural areas. These were all plans to minimize undercounts, particularly in minority and marginalized communities. The budget cut stifled innovation.

Collecting data digitally and storing it on the blockchain requires far less manpower than the current census does, which makes our solution much less expensive than the Decennial Census. The cost of the census also climbs with population increase, while our costs will increase marginally. An additional advantage is that the security of the data stored on the blockchain increases as the amount of data

increases. This means both population and participation increases have a positive effect on the integrity of the Enumera system.

The costs benefits allow Enumera more room for innovation. In particular we place a premium on providing accessible technology that accounts for all people, including those in remote, rural areas, or less connected areas.

The programs and businesses that rely on census data are innumerable. The Census Bureau has published reports stating concrete examples of organizations and government departments using their Decennial Census data, and making alterations to their services and budgeting because of this data. To provide one concrete example, during the 2015 fiscal year, data from the Census Bureau was used by 132 programs to distribute more than $675 billion dollars in funds [8]. Enumera's statistical services under the Open Census Protocol collects more data and more accurate data. We expect our blockchain based solution will be a valuable alternative for these 132 programs and others.

In order for the Open Census Protocol to work, collaboration will be needed with local and federal governments such that individuals can be appropriately identified by some standard form of government identification. Furthermore, we acknowledge that the Enumera census will not replace the existing census, but it can provide benefits that outperform the Decennial Census. We do believe that there is potential for Enumera to collaborate with the Census Bureau by providing our services to them.

References

[1] https://unstats.un.org/unsd/demographic/sources/census/docs/P&R_Rev2.pdf

[2] https://www.brookings.edu/blog/fixgov/2017/08/31/the-2020-census-may-be-wildly-inaccurate-and-it-matters-more-than-you-think/

[3] www.census.gov

[4] https://www.aclu.org/news/aclu-immigrants-rights-groups-sue-trump-administration-over-census-citizenship-question

[5] https://www.npr.org/2017/07/18/536484467/census-bureau-found-no-need-for-lgbt-data-despite-4-agencies-requesting-it

[6] https://www.congress.gov/bill/115th-congress/senate-bill/3314/text

[7] freakonomics.com/2010/02/02/can-you-trust-census-data/

[8] https://www.census.gov/library/working-papers/2017/decennial/census-data-federal-funds.html