Catalyzing Innovation
Annual World Bank Conference on Land and Poverty
Washington DC, March 25-29, 2019

Self-Sovereign Identity and Disaster Resilience in Puerto Rico

Christopher Mellon
Future of Property Rights Program, New America, Washington, D.C., United States
fpr@newamerica.org

Paper prepared for presentation at the
"2019 WORLD BANK CONFERENCE ON LAND AND POVERTY"
The World Bank - Washington, D.C., March 25-29, 2019

Copyright 2019 by author(s). All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Table of Contents

1. Introduction

"People now have the freedom to have crosscutting identities in different domains. At church, I'm one thing. At work, I'm something else. I'm something else at home, or with my friends. The ability not to have an identity that one carries from sphere to sphere but, rather, to be able to slip in and adopt whatever values and norms are appropriate while retaining one's identities in other domains? That is what it is to be free."

- Elizabeth Anderson, "The Philosopher Redefining Equality," Interview by N. Heller, The New Yorker, Jan 7 2019

The above quotation is not about digital identity, but professor Anderson's definition of freedom is key to understanding the fundamentally flawed way in which identity now functions on the internet. The domains of activity she invokes -- religious, professional, personal, social -- now exist to a great and increasing degree in digital spaces.

Until recently, it was common for the average internet user to assume that his or her activities on the internet *were* free in the way Anderson describes, with discrete identities for different purposes. Thanks to a series of scandals involving hacks, privacy breaches, and above all growing public awareness of the inner workings of the data economy, more people are realizing that their sense of online privacy is largely illusory. Though it takes some effort to track and correlate the various identities which a given person uses online, that kind of surveillance is an essential part of the ad-driven business model that has allowed companies like Facebook to grow to an enormous size.

This is not to say that online anonymity is a complete fiction. There are obvious examples of digital platforms providing their users with pseudonymous online identities, enabling activities that would be much more difficult to perform in the "real" world, from political dissidence to asking embarrassing or taboo medical questions. However as more and more of our activities move online --not only our communications but also our movements, health data from wearables, etc.-- we become increasingly exposed to new privacy and security risks.

But this paper is equally concerned with a different, complementary part of the online identity problem. While it is true that in order to be free we must have the ability to use different identities for different contexts, we must also have a way to assert the core, physical, identity that underlies our various digital personas. It is not enough to have many different identities on the internet, like Facebook accounts, Gmail addresses, and Twitter handles. There are purposes for which we must use our fundamental identities, what David Birch calls our "legal identities."[1] That is a tricky thing to do remotely through an exchange of data.

The technology with which this paper is chiefly concerned, self-sovereign identity (SSI), was developed in order to give users a digital identity that enables true online privacy where appropriate, and allows them to assert their legal identities when necessary. SSI is created by the identity holder and cannot be revoked by any external authority. Moreover, it gives the user a greater degree of control over how their

[1]

personal information is shared. Users keep credentials issued by third parties in a digital wallet and can choose when and with whom to share them. These credentials can be validated without contacting their issuer. In simplest terms, SSI is an identity that functions much as identities do offline, but with a greater degree of privacy and security. By allowing users to choose which aspects of their identity they "carry from one sphere to another," SSI is an essential tool for creating the kind of freedom Anderson describes. And in addition, an SSI linked to the user's biometrics allows their fundamental, "legal identity" to be asserted remotely.

But how does the question of digital identity fit into disaster preparedness and risk reduction efforts? As we will see, SSI is an infrastructure-level enhancement of disaster management capability that allows for the exchange of trusted, verifiable data with an auditable provenance. Without the ability to assert our unique, legal identities online we cannot perform actions that require high degrees of trust. That forces those transactions offline, negating the most revolutionary characteristics of the internet, the virtual elimination of distance. Collapsing the space between people has made all sorts of important interactions faster, easier, and cheaper. This has many applications in the housing and real estate sector, where high risk has given rise to administrative practices that are cumbersome and expensive.

This paper -- based on both primary and desk research-- examines the post-hurricane housing crisis in Puerto Rico, the ways in which informal property ownership has contributed to that crisis, and how new technologies for the secure collection and sharing of data can help the island prepare for future disasters. The first section describes briefly the nature and extent of the housing crisis in Puerto Rico and the government's recovery plans. The second section introduces three emerging technologies that can in combination be used to help implement those plans, with a special focus on blockchain-based self-sovereign identity (SSI). The third section describes how a trusted data sharing platform built on these technologies can be used throughout the disaster management process, from preparation, to response, to recovery.

2. Hurricanes and the Housing Crisis in Puerto Rico

The disastrous 2017 hurricane season exposed the degree of informal property ownership in Puerto Rico and highlighted its contribution to the ensuing humanitarian and economic disaster. Though Puerto Rico has a modern, digital registry, an estimated 55% of properties are informally owned, and 30% of documented properties are not recorded in the registry. This has greatly increased the difficulty of many aspects of the disaster assessment, response, and recovery.

Given that the construction quality of informal homes is typically low, these properties are more vulnerable to storm damage than structures that have been built to code. Moreover, the occupants of informal homes typically lack insurance and the documentation needed to access recovery aid, which makes post-disaster rebuilding very difficult. Approximately 60% of FEMA assistance claims were rejected in the aftermath of Hurricane Maria, including many cases in which applicants who needed help rebuilding could not provide titles to prove ownership of their homes. The problem was so pervasive that FEMA began accepting affidavits as evidence of property ownership in place of formal documents.

The Puerto Rican government's draft recovery plan outlines 20 measures to "repair and rebuild resilient residential housing," with a total estimated cost of more than $50 billion dollars. The bulk of this cost comes from reconstructing damaged homes and relocating vulnerable communities. The plan also calls for the formalization and registration of the island's informal properties. One of the most formidable obstacles to these initiatives is the lack of information about land and property ownership, use, and condition.

Both the current recovery efforts and planning for future disasters require data garnered from a comprehensive asset mapping and needs assessment program. Although discovering and recording the truth on the ground will be an expensive and time consuming endeavor, it also presents an enormous opportunity to simultaneously address the lack of property and occupancy data--one of the most fundamental obstacles to disaster resilience and to economic development on the island. But this data cannot remain static. Extreme weather events like Hurricanes Irma and Maria are likely to become even more frequent in the future. Effective disaster response will require access to up-to-date and, in some cases, near-real-time information on the progress of the storm, the damage caused, how to reach people in need, and the kind of aid required. Rebuilding in the aftermath of disaster requires not only information about damages, utilities, emergency incidents, and business licenses, but also demographic, financial and health data to ensure that aid is delivered equitably.

3. Crowdsourcing and Trusted Data

There has been a powerful movement in recent years to harness crowdsourced data for disaster management. There are numerous recent examples from Nepal, Haiti, the Philippines, and Syria. Adoption, however, has been slower than might be expected given the potential power of these techniques, even in highly developed and tech-saturated environments. In the response to Hurricane Harvey, for example:

"...the United States Coast Guard urged people not to tweet for help, and instead, use official channels to seek recourse. Their rationale was that social media posts were too difficult to verify and could easily be missed... They also have doubts regarding the reliability and quality of crowdsourced data… The sheer amount of crowdsourced information available during disasters can be too overwhelming for aid organisations to handle, to the extent that verifiability becomes an issue. During the Haiti earthquake, 90 per cent of aid requests sent via text messages were either inaccurate or repetitive."[2]

So the most valuable qualities of crowdsourced, data, its volume, availability, and timeliness, make it difficult to use for emergency management. Crowdsourcing typically means dealing with large amounts of information collected or reported through social media platforms. The combination of high volume and low quality data requires a lot of of verification work in order to be used productively. Indeed the processing of such data is its own category of crowdsourcing activity. In the case of real-time data, the picture is further complicated by diminished connectivity due to disruption of the power grid, cellular

[2] https://www.nst.com.my/opinion/columnists/2017/12/319041/disaster-relief-through-crowdsourced-data

networks, and transportation infrastructure. Another, related, difficulty is the inability to "crowdsource" highly sensitive information like financial and medical records.

These problems share several common components. The verification of real-time data could be streamlined with an identity system for the users who are generating it, allowing for attribution, deduplication, and reputation management. As for the sharing of sensitive, high-value data like financial and medical records, there must be a system for issuing verifiable data and storing it under the user's control in the cloud. In addition to guaranteeing the provenance of this data, the system must allow it to be shared with the appropriate authorities in a timely fashion and without imposing an unrealistic burden on the user to manage the exchange.

The contention of this paper is that self-sovereign identity platforms provide the best foundation on which to build this kind of trusted data sharing. The proposed infrastructure for exchanging trusted data would not replace traditional crowdsourcing or centralized government databases. Rather, it creates an additional high-value data layer that allows sensitive/personal information to be used for emergency preparation, response, and recovery in a way that is private and secure.

Before a disaster strikes, people can create repositories of critical identity, financial, and medical documents that will remain accessible to them wherever they go and cannot be revoked or disputed. Moreover, SSI can reduce both the likelihood of data breaches and, depending on the implementation, reduce the harm caused by a breach. Credentials on the Sovrin Network, for example, are designed in such a way that they can only be used by the identity to which they were originally issued. Given the amount of sensitive personal and financial data that must be shared in the course of disaster response and recovery, there is a substantial privacy risk. It was recently revealed, for example, that FEMA exposed the addresses and financial information of millions of victims of Hurricanes Harvey, Irma, and Maria, as well as the 2017 California wildfires.[3]

[3] https://www.washingtonpost.com/national/health-science/fema-data-breach-hits-25-million-disaster-survivors/2019/03/22/3e2c6232-4cec-11e9-93d0-64dbcf38ba41_story.html?noredirect=on&utm_term=.ab52aa0023ec

4. Technologies for Trusted Crowdsourced Data

A. Self Sovereign Identity

The basis of the trusted data platform envisioned in this paper is the use of SSI paired with a personal data store. The leading SSI solutions leverage blockchain to provide users with a persistent and secure digital identity that cannot be revoked, altered, or accessed without their explicit permission.

The great advantage of SSI is that it can make identity in the digital world function more like identity in the physical world, in which every person has a unique and persistent identity which is represented to others by means of both physical attributes and a collection of credentials attested to by various external sources of authority. These credentials are stored by the identity holder --in a digital wallet on a smartphone or in the cloud-- and presented to different people for different reasons. Crucially, the identity holder decides what information to present based on the environment, trust level, and type of interaction. A user's fundamental identity persists even though the credentials by which it is represented may change over time.

In emerging models, a self-sovereign identity and related personal data are encrypted in a distributed storage system like IPFS. Via cryptography, the identity holder can use a credential to access many different systems and services; but there is no third-party tracking the services to which the user authenticates. Furthermore, cryptographic techniques called "zero-knowledge proofs" (ZKPs) can be used to prove possession of a credential without actually revealing the credential itself, helping to preserve the privacy of sensitive information. As we will see later the same effect can be accomplished through the use of Secure Multi-Party Computing.

Although the concepts behind SSI have existed for decades, actual implementation was technically infeasible until recently. The arrival of blockchain and the advancement of biometrics have brought SSI from concept to reality. Blockchain enables both distributed storage and peer-to-peer transactions, both of which are helpful for a model that requires users to control data instead of having it under the control of a centralized authority.

Biometrics are also critical for enabling SSI, allowing intrinsic characteristics of the individual to be extended into the digital world. When we go in person to renew a driver's license or to have a document notarized, we are undergoing a series of implicit identity checks that may not be obvious. The first and most important of these is biometric. human beings are exceptionally sophisticated 'sensors' when it comes to recognizing other living humans and their physical features. This check is accompanied by the submission of documents and, taken together, these checks furnish proof of identity. This is proof not only that you are who you say you are, but that you have a number of other qualities and credentials required for that transaction. You might have to prove that you live in the jurisdiction issuing the driver's license, perhaps with a piece of mail addressed to you by a trusted party like a utility company. You may also have to furnish proof that you completed a driver's education course.

These different credentials cannot currently be conveyed electronically in an easily authenticated way. They often include physical security features to resist duplication, and when accommodations are made

for the sake of convenience, for example scans of ID documents, then security is reduced. Another invisible, behavioral authentication factor is the need to spend money, time, and effort travelling to a physical location to get a driver's license. This also introduces the risk of physical exposure to people and cameras for someone trying to obtain a document illegally. So the question is: how do we replicate these features digitally so that people no longer have to go into the office to get a driver's license?

The problem has a few different components, as we have already seen in this example. A core legal identity must be established And the properties of that legal identity must be gathered in a format that can be verified and shared . Biometry is essential to establish a unique core identity and to guarantee that when data is accessed or shared the core identity holder is the one doing so.

B. Image Authentication

Self-sovereign identity and verifiable credentials provide the backbone of trusted data, but they are not sufficient. User-generated content is critical for disaster management, especially photos and videos documenting conditions on the ground during a disaster and damage and aid requirements in the aftermath.

Visual media are also have the greatest impact on public perception and narratives surrounding the event. Moreover, they are susceptible to manipulation. In the context of natural disasters the danger of this kind of fraud is likely to be negligible during the event, becoming a concern only with respect to aid and insurance payments in the aftermath. But in man-made disasters caused by political violence, the risk of deliberate deception is high.

The need to verify pictures and videos has given rise to a number of image-authentication services. Mounir Ibrahim, who is now Vice President of Strategic Initiatives at Truepic, was convinced of the need for image-authentication by his experiences as a US Diplomat in Syria during the Arab Spring: "Other countries—people who wanted to deny the reality of what was going on in Syria—undermined the validity of user-generated content... It was a surprisingly effective argument.[4]"

When a picture is taken with the Truepic app it is run through a series of tests to verify that it is a unique, unmodified image of a 3D scene. The time and location of the image capture are established through a variety of sensors, including GPS, network signals, the phone's barometer, and a timestamp from a clock on a Truepic server. Once verified, the image is watermarked, embedded with a unique ID number, and hashed to the Bitcoin blockchain. The image ID is incorporated into a unique URL that people with whom the image is shared can visit to confirm the authenticity of the image and the associated time and location data.

Truepic is adding functionality to identify deliberate manipulation of images, including deepfakes -- hyper-realistic human images synthesized by AI. That may not be necessary in the case of a hurricane, but in responding to future humanitarian crises caused by human conflict it is virtually certain that the misinformation spread by social media will include photos and videos that are either outright fake or are being presented out of context for propaganda purposes. For proof of location in particular it may

[4] https://www.fastcompany.com/90299000/truepic-most-innovative-companies-2019

eventually be necessary to have the location data processed in a tamper-proof, trusted computing environment which is secured at the hardware level and can sign the data with it's own private key.

C. Secure Multi-party Computation

Granting users control of their data comes with one enormous downside: it puts them in charge of governing access to that data. That can be burdensome under the best of circumstances, with users overwhelmed by the number of data transactions they are expected to manage. The end result would likely be that valuable data remains compartmentalized in individual storage instead of being made available to responsible users. That would be even more likely in a disaster situation where both connectivity and attention are limited resources. So there has to be a data governance scheme allowing permissioned access to user data by groups such as verified aid organizations. Access must be possible without the user having to approve access requests in real time and there must be additional privacy safeguards to compensate for the fact that the individual access requests are not being vetted and approved.

A possible answer to this dilemma is found in Secure Multi-Party Computation, which allows data to be processed without ever being decrypted or transferred from its place of storage. The Enigma project[5] out of MIT is a leading example. It is beyond the scope of this paper to describe in detail how the Enigma protocol and similar project like OpenPDS accomplish this. For our purposes these protocols are a modification of SSI and verifiable credentials[6] that resolves the tension between the privacy and security benefits of personal data storage and the need to process aggregated data. [7]

Again, access permissions to this data should not be managed directly by the end users. Policies should be instituted at a group level, through a trusted organization serving as a data fiduciary. This might be a community organization or an NGO. Users would opt in once to a set of access policies agreed amongst the members of the group. For example it might be agreed that the local emergency management organization can query the medical records of community members in order to ascertain which medicines will be needed or how many dialysis patients will need priority access to generators. Using secure multi-party computation, answers to these questions can be returned without transferring, or even exposing, the underlying records.

D. Reality Mining

Since the advent of smartphones, human activity has increasingly been reflected in a trail of data. This "breadcrumb" data can lead us to many kinds of important knowledge about the people creating it. The term "reality mining" was coined by Nathan Eagle and Sandy Pentland of the MIT Media Lab for the collection and analysis of this kind of data.[8] We need to think about how to capture and use that data for disaster prep and make sure it persists through crises. If it is collected and stored in a standard, verifiable

[5] https://enigma.co/enigma_full.pdf
[6] https://blog.enigma.co/off-chain-identity-claims-with-enigma-2d5b23c31f
[7] https://blog.enigma.co/off-chain-identity-claims-with-enigma-2d5b23c31f92
[8] http://www2.technologyreview.com/news/409598/tr10-reality-mining/

claims format, it will do much to empower people to use their data in creative ways without putting their privacy at risk.

One approach would be to turn some of the events reflected in cell phone data into credentials. A property claim might be bolstered by breadcrumb data showing that the claimant has slept at that GPS coordinates corresponding to the property for years or decades. Additional elements of a compound occupancy proof might come from Amazon delivery data or the start and end points of Uber trips.

Data from social networks and employers would be invaluable for a reputation management system to cut through.

5. A Trusted Data Platform for Resilience and Recovery

In order for the government of Puerto Rico to plan and execute the housing recovery, it must have accurate data on property ownership, condition, and occupancy. As the draft recovery report acknowledges, "the creation of a central source of data related to the housing stock, including title, permits, land use, property tax and location will be crucial in the recovery effort."

This data may form the foundation of the housing recovery, but it can become transformational when integrated with secure digital identity and a geospatial map layer. Many aspects of disaster response can be dramatically improved by bringing together critical data about people, properties, and locations with an extremely robust data supply chain. Such a system would have a wide range of applications, including:

A. Data Provenance and Reputation Management

As was discussed earlier, the usefulness of real-time crowdsourced data is limited by the resources needed to verify it. Incoming information can be weighted according to a combination of credentials and reputation scores built up over time. Input from verified first responders, veterans, scientists, or anyone with relevant experience or skills can be given greater weight.

Many other applications can be imagined, from allowing people to document property damage with their smartphones, uploading photographs linked to their ID, property ID if available, GPS coordinates, etc, to provide a quick initial damage assessment. All data (damage, repairs, permits, etc) related to a given property could be logged over time and linked via the property ID. Neighbors could cross-check or endorse each others property claims with digital signatures, which would be especially helpful in establishing eligibility for aid where formal titles do not exist. The claimant could add financial records to determine reconstruction assistance grants.

The ability to share verified health records without compromising privacy could save lives. It is critical to know where the most vulnerable people are during a prolonged power outage. Health records could be used to identify an urgent need for access to electricity or to allow first responders to plan distribution of supplies like medications.

From the government/NGO side the same platform could be used for financial transparency, tracking funding and service delivery for all the way down to the individual subcontractor who replaces a broken toilet or fixes a roof. The blockchain technology on which SSI is based could eventually be used not only to track the flow of money but to provide a value-transfer layer with which to actually deliver financial aid.

B. Generating Land Occupancy Documents

When it comes to land-related issues in Puerto Rico, the most obvious intervention would be to begin addressing the lack of formal documentation which resulted in so many aid claims being rejected. First, for disaster preparedness, people who do not have titles should be issued signed digital documents that establish their occupancy.

This could be done in a number of different ways. The simplest, most immediately impactful solution would be to create a repository of land occupancy affidavits together with the existing supporting documentation, e.g. utility bills. FEMA relaxed its documentation criteria for individual assistance claims in the wake of hurricane Maria to accept signed affidavits stating that the occupant has lived there for a certain number of years, maintains the property, and has been unable to obtain a deed. A Puerto Rican legal aid organization, Ayuda Legal, has already created a digital version of this document.[9]

Having them created and stored for all occupants of informal property before the next hurricane would be enormously valuable in the aftermath. Creating and storing them digitally on an SSI platform would have several advantages. Each document would be irrefutably linked to the signer with biometry and cryptographic signatures, and in the event of another major disaster could be filed remotely. The documents would be stored in a private and decentralized way, assuring occupants that their information cannot be accessed or used in any way without their knowledge and consent.

C. Registering Informal Property

The more radical approach to the lack of land documentation would be to use a community mapping methodology to generate land titles. The process of formalization in Puerto Rico already implicitly embraces this concept. As is the case in many jurisdictions around the world, establishing ownership in the absence of prior documentation requires asking the neighbors to attest to the claimants occupancy or ownership.

In Puerto Rico, a surveyor creates a parcel map and asks the owners of the adjoining properties to verify the boundaries (they sign off by verified mail). The map is then filed in court and a hearing is held to get a judgment. If everything is in order and ownership is not contested, a deed is issued. This whole process costs costs approximately $2,500 if there are no competing claims to be adjudicated. Only $500 of that cost is spent on surveying, the rest is administrative fees.

A sufficiently robust digital ID cannot be repudiated, and is a cheaper, more effective way of attesting to a property claim than using notaries and certified mail. Moreover, in most cases the precision of a

9 https://ayudalegalpr.org/resource/entrevista-sobre-declaracin-bajo-juramento?ref=Zc2NJ#122A4ADE-1FE4-428A-AEBC-071A6DC871C7

professional surveyor is an unnecessary expense. Mobile mapping platforms like Cadasta and MAST can be used. If boundary accuracy is an issue, the output need not be a full freehold title. A provisional title would be sufficient for most purposes, especially with respect to emergency management, and there is already a comparable “use and enjoy title” under Puerto Rican law.

Conclusion

The goal of this paper was to demonstrate the potential of trusted data to improve emergency management. The main components of that capability, SSI and secure multi-party computing, are still being developed, though they are seeing increasing deployment. But the development of the technology means nothing if it is not adopted, and the efficacy of these sorts of systems depends on network effects. An SSI wallet is not worth much if trusted authorities like government agencies, hospitals, and universities are not issuing compatible credentials.

What is more, people do not want to download, set up, and learn to use apps with very limited functions. Even a fully functional emergency management app that was not useful in daily life would probably fail. Many different organizations have made, or have planned to make, apps to address different parts of emergency management, to find that people will not dedicate thought to downloading and learning them before a disaster or memory to storing them in preparation for an unlikely event.

For these reasons an app like the one described here would ideally be built into a broader sort of citizen services app, where the core credentials, like driver’s license, voter registration, etc are issued by the government and used for official services and access to financial services. Key management might be done through the telecom companies, as has sometimes proven the most efficient method with other blockchain products. Or a digital ID/wallet could be built into the telecom company app. Telecom companies can be effective ID and key management partners, as demonstrated by a number of previous projects like Chromaway’s land registry in Sweden.

As was mentioned in the section on secure multi-party computing, SSI with personal data stores introduces a serious problem of data governance. But if that gap is bridged, it will be an unprecedented opportunity to build systems that maximise the sharing of sensitive data for ethical use.