# The Five Truths about Biometric Data

Contributors:  
John Callahan

Darrell Duane

Shannon Appelcline

Asem Othman

**ABSTRACT**

Biometrics are already here, and they *will* be used on the internet, whether we like it or not. They are already are; a lot of the current practices are bad and the centralized hosts know it. We know that centralized repositories of sensitive information are highly vulnerable and present irresistible targets, but better solutions are not available to current practitioners; they’re not even being considered. Fortunately, there are other fields that are already solving these problems, such as self-sovereign identity. We need to adapt their best practices to make biometrics safe and secure.

### Introduction

Biometric data is a third rail for discussions of identity. It’s broadly understood to be both permanent and sensitive, so there’s a value fear is that if the data is put online, it will inevitably be stolen and abused.

There is the potential for disaster.

Despite this fear, biometric data already exists, and it’s already being recorded digitally. We need to discuss it, or other people will determine the future of biometric data, how it will be recorded, stored, and used.

And the likelihood is, they will do so *badly*.

Good usage of biometric data should focus on privacy, security, and scalability. The owner of the data should be able to determine how it’s used, and they should good confidence that it will be protected, and they should be able to use it as part of a system that could support data from over 7 billion people.

Trying to meet all of these requirements might seem daunting, but there’s already a philosophy for identity that advocates these same needs: self-sovereign identity[[1]](#footnote-0). Christopher Allen’s “The Path to Self-Sovereign Identity” proposes ten principles for self-sovereign identity, and many of them are directly translatable to the needs for private, secure, and scalable biometrics. Biometrics, like self-sovereign identity, require: *control*, *access*, *persistence*, *portability*, *interoperability*, *consent*, *minimalization*, and *protection*.

To meet these needs, biometric data can be split in two, divided between an *identifier*, which recognizes, remembers, and relates to an identity, and *data*, which contains the actual biometric information.

Adopting self-sovereign principles for biometric data requires a new framework. This paper lays them out as five *truths* that are needed for the safe storage and usage of biometric data.

### Biometrics Should Never Be Stored in Central Databases [Darrell]

When Biometrics for Authentication were first developed, the only option to store the biometric data for comparison with the login data was in a centralized silo connected to the internet. As time has gone by, centralized silos have become more and more vulnerable, presenting irresistible targets for data thieves.

There are a number of cases where significant collections of very personal data have been hacked. In the United States, Equifax lost the financial information of 143 million people[[2]](#footnote-1), while the Office of Personnel Management lost private information for 21.5 million people[[3]](#footnote-2). And biometric data has already been endangered: AADHAR has suffered various breaches[[4]](#footnote-3).

As new data technologies and architectures have emerged, many have thought that there might be some role for the Blockchain for Biometrics.

### Biometrics Should Never Be Stored on the Blockchain [Shannon]

Blockchains are one of the most popular and well-tested decentralized tools for storing data on the internet. They have already become a prime target for biometric storage. However, they should not used for this purpose.

Storing biometric data on the blockchain is dangerous for two reasons. First, blockchains are immutable: any biometric data placed on a blockchain will be there forever. Second, blockchains are public: any data placed on a blockchain is visible to everyone on the blockchain.

Blockchain data can be made less public using a few different methods. A blockchain can be private, accessible only to a particularly group or organization. In addition, data on a blockchain can be encrypted, making it available only to people holding a specific secret. Unfortunately, each of these protections is likely to eventually be defeated. A disgruntled community member could make a private blockchain public; and encryption is likely to eventually be broken, as technology is always evolving. In other words, any data put on a blockchain is likely to become fully public at some point in the future, and that is an unacceptable result for data that is as sensitive as biometric data.

Though data should not be placed on the blockchain, that doesn’t mean that blockchains can’t be used. They’re a great decentralized tool, and there are ways to use them than don’t store biometric data on the blockchain itself.

### Biometrics Can Be Accessed Via the Blockchain [Jack]

If you need to store biometric data, you can store a ***pointer*** to it on a blockchain but not the data itself. Biometric data always needs to be secure and under your control. For example, a mobile device is currently a popular platform for storing biometric data due to the available hardware locks (e.g., a TPM) and the fact that you physically possess the device. Your device is assigned several public addresses (e.g., IP, phone, device id) but biometric data on the device is supposed to be accessible only by explicit consent.

Biometric data is private, not secret[[5]](#footnote-4). A password is a secret and should never be shared by definition. Your biometrics are private because you must consent to share them on each use. Like a house, your address is public but your possessions inside the house are private under lock and key. Only by explicit consent should anyone be granted permission to enter your house because that implies access to your private possessions.

In the near future, cloud storage will be private as well including sensitive information like your biometrics and medical records. This private data should be secure (both legally and technologically with locks and encryption). You may need to share this private data, but only under explicit consent. The keys to your house give you control to access even though the address is public.

[Additional Topic of Discussion: Offline First. How can we make this work when you’re offline without internet access, then sync up lately?]

### Biometrics Should Be Under A User’s Control [Shannon]

When a user decides where to place his actual biometric data, he should ensure that it stays under his control: he should be the authority who determines who can access his data, which data they can access, and which capabilities they have. He should also have the authority and ability to move his data as he sees fit. In other words, his relationship with his biometric data should adopt many of the rules of self-sovereignty.

This requirement does not enforce any specific physical implementation. Currently, many people are advocating storing biometrics on mobile devices, but there is no guarantee that personal electronic devices will remain common in the future; a model of ubiquitous, shared electronic devices is equally possible. When considering the ability to control his biometric data, a user might consider many other options, including cloud storage. Microsoft’s Identity Hubs[[6]](#footnote-5) and Blockstack’s Gaia Storage[[7]](#footnote-6) may both be viable options. Besides deciding where their data will be stored, users may also make decisions about *what* control entails, as some particular storage solutions might reduce control, but only by their ultimate decision.

Supporting total user control maximizes the chances that a user will be able to constantly access and protect his data, even over a very large span of time. It allows the user to ensure that his data storage matches the previous criteria for biometric storage: that it’s not centralized and that it’s not stored on a blockchain. It also allows them to make important privacy decisions, such as whether their biometric data remains accessible in the future.

The largest challenge with user control of biometric data is that it may not be protected legally. This is a topic that will need to be addressed legally with new data sovereignty rules: data needs to be protected by property laws, to ensure that a user’s ability to control access and capabilities for his data remains intact.

With biometric storage (at last) stored in a way that maximizes protection and authority for the user, the next question is: *how do you use it?*

### Biometrics Traits Should Be Reliable [Asem/Shannon]

Not just any biometrics will do the job. When choosing which biometrics will be stored and used, care should be taken to ensure that the biometrics are reliable.

This is a problem for two reasons. First, some biometrics are less reliable than others[[8]](#footnote-7). For example, fingerprints and iris scans are very different than face and voice scans. Second, there’s more opportunity to spoof some biometrics than others: Googling a picture of someone is easy while lifting a fingerprint is harder; and producing a fake image of someone is easy while producing a fake finger is harder.

This isn’t to say that any biometric is necessarily superior, let alone perfect. Instead, biometrics need to be selected that are reliable in relation to the specific application that they’re being used for.

And even with that, biometrics might not be quite enough ...

### Biometrics Should Be Part of a Systematic Procedure [Asem/Shannon]

Biometrics aren’t the end-all and be-all. They need to be part of a larger systematic procedure. There are a number of possible steps for a biometric system:

1. Biometrics can have passive liveness (such as an Apple face recognition), or a more extensive system can require active liveness (such as lip-reading a person’s face when they specified words), to facilitate presentation attack detection (PAD).
2. Biometrics can combine multiple factors, such as requiring a passphrase in addition to biometrics.
3. Biometrics can produce confidence ratings; rather than offering a binary result, they can suggest how likely a result is to be accurate, based on the reliability of the biometric trait for the purpose and the use of other systematic elements such as active liveness and multiple factors.
4. Biometrics can have fallbacks. If either the measurement device or the person suffers a temporary or permanent change (such as a scar, a lost finger, or a broken phone), then the biometric system should permit other means of access.

The need for these more in-depth systems is a fact that’s already been stated: no biometric is perfect. This means that any system will suffer from False Rejection Rate (FRR) and False Acceptance Rate (FAR): the right users may be rejected and the wrong users may be accepted. NIST’s SOFA metrics provide more details on these problems and possible attacks[[9]](#footnote-8).

Systematic approaches can be actively aware of these challenges and compensate for them.

## A Biometric Use Case

1. Give the Solution
   1. Architecture of Safe Use of Biometrics
      1. Store in an Identity Hub
      2. Point to from a Blockchain
      3. [and example of DID-Auth too]
   2. And we explain how it meets our 5 or 6 criteria
   3. This solution will cause new problems…

[WE NEED TO INCLUDE THE SIX TRUTHS HERE]

BIOMETRIC ENROLLMENT

(initial setup)

STEP 1: USER stores IBV in his agent (identity hub, maybe mobile)

A Key-Pub is Generated for accessing the IBV.

[HAVE THE USER GENERATE TWO BIOMETRICS, SO THAT HE CAN HAVE An OK/EASY ONE AND A GOOD/HARD ONE]

STEP 2: USER creates or reuses DID appropriate to personality

STEP 2.1: If USER creates DID, then his agent must register it on blockchain

STEP 3: USER links to the IBV and its public key in his identity hub from the DID Document referenced by the DID

[THIS DEMONSTRATES: NOT CENTRALIZED, NOT BLOCKCHAIN, UNDER USER CONTROL]

[THIS MIGHT BE A GOOD POINT TO EXPAND ON USER CONTROL: HE CAN DELETE HIS BIOMETRIC AT A LATER POINT, HE CAN CHANGE WHAT’S IN HIS DID DOC AND MAYBE EVEN MOVE IT DEPENDING ON WHICH DID METHOD HE’S USING]

STEP 4: USER asks ISSUER for a credential, giving them the DID

[THIS DEMONSTRATES: THE WHOLE VIA SYSTEM, AS DO A FEW LATER STEPS]

STEP 5: Identity Assertion: ISSUER asks USER for additional metadot, such as drivers license, social security card, or other KYC-required information. This is IDV (Identity Verification)

STEP 6: ISSUER creates or reuses DID appropriate to personality

STEP 7: ISSUER creates a verifiable credential, binding the IBV to the USER’s identity and signs it and sends it back to the USER and put a hash of the information on the BLOCKCHAIN.

STEP 8: The USER adds the verifiable credential to his DID Document. (If He Wants)

You now have a DID Doc that proves your biometric is linked to your person.

(BIOMETRIC) IDENTITY ASSERTION

STEP 9: USER wants to link a biometric to a new or existing RELYING PARTY, such as a bank. USER sends DID to RELYING PARTY to link the biometrics to his bank account. And also metadata (like bank account number and password for Identity Assertion).

STEP 10: RELYING PARTY uses the universal resolver to look up DID. The RELYING PARTY then requests the DID Document from USER.

STEP 11: USER’S AGENT sends back DID Document, which contains the biometric public key and the signed verifiable credential linking the biometric public key to the identity.

[THE BANK REQUIRES A STRONG BIOMETRIC. WE CAN USE THIS TO DEMONSTRATE OUR TRUTH ABOUT RELIABLE BIOMETRICS.

WE CAN ALSO TALK ABOUT AN ALTERNATIVE: IF THE USER HAS A WEAK BIOMETRIC, THE BANK MIGHT ACCEPT THAT ALONGSIDE ANOTHER OPTION, SUCH AS A SOCIAL SECURITY NUMBER. THIS DEMONSTRATES SYSTEMATIC TRUTH. SO THAT”S ALL SIX.]

STEP 12: RELYING PARTY verify hash of biometric binding on the BLOCKCHAIN

[This does not require issuer to be still alive]

STEP 13: RELYING PARTY encrypts a challenge with the public key.

STEP 14: USER biometrically authenticates to prove that he controls DID and is the right person. He encrypts it with the private key and sends it back [DID AUTH COMES IN HERE]

[Is this a different type of challenge? PUBLIC KEY + TYPE (like BOPS or FIDO)

STEP 15: USER sends encrypted message back to RELYING PARTY, who now links their account with the DID

## Conclusion

1. Conclusion
   1. A Roadmap is Needed for this Future World
      1. Train has already left station for doing this in a bad way with ADHAAR and server-side matching
      2. Legal & Technical & Social Challenges for doing it better
   2. How do we Get There?
   3. That’s the next paper

============================================

FINAL NOTES

OPENING:

The use of self-sovereign identity offers:

1. Privacy

2. Security [group/national]

3. Scalability

[maybe can only have two? We want to say all three are achievable.]

Biometrics is one of the ways to prove who you are

* The naked man problem, refugee crisis

Biometrics include (various physical characteristics, embedded devices): can be ephemeral or permanent. We’ll use them, even though they’re ephemeral.

It’s the third element of: What you have, what you know, and what you are

ROADMAP

How to we get from legacy systems to respecting all three

PURPOSE OF PAPER

Pro/con

STRUCTURE

* ½ page to a page, but definitely not more than a pags
* 2-3 paragraphs
* 1. What are we proposing
  + Expansion and explanation of the topic statement
* 2. Why?
  + Why is this important?
  + For the Negatives include current state
* 3. Segue into Next Section

The Six Truth About Biometrics

1. Abstract
2. Discuss the problem
   1. They Exist
   2. And They’ve Very Important: the Naked Man Problem
   3. They’re Bad
   4. We need privacy, security, and scalability
   5. Self-sovereign identity teaches us things, like being able to control your identity, minimize it, protect it, make it portable [go through the ten points, and list the ones the that are applicable]
   6. Cite: Christopher’s Paper
   7. Another principle: We separate identifiers and data (identity)
      1. We separate the identifier from data immutably related to an identity
   8. So, how do these principles apply to Biometrics
   9. Like This ...
3. Lay Out the Six Truths
   1. [this is the intro paragraph]We Need to Do Biometrics Before Someone Does It Wrong
      1. THis is a sensitive topic
      2. People have strong opinions about personal data
      3. We have decided to attack this head on
      4. And to make the community sentiments obvious[these are the things that people might do wrong, and so we list them as truths]
   2. Biometrics Should Never Be Stored in Central Databases
      1. Not in Silos

From reports that AADHAAR, the system run Unique Identification Authority of India (UIDAI) having a rogue operator selling access to the database to

AADHAAR [<https://www.hindustantimes.com/tech/aadhaar-breach-everything-you-need-to-know/story-VhCKHDIL8lziw6OcnhL4wO.html>]

Joe Andrieu’s People Centered Identity article?

Equifax {https://www.nytimes.com/2017/09/07/business/equifax-cyberattack.html]

OPM [https://www.washingtonpost.com/news/federal-eye/wp/2015/07/09/hack-of-security-clearance-system-affected-21-5-million-people-federal-authorities-say/?utm\_term=.e5301c5c6f53] , [<https://en.wikipedia.org/wiki/Office_of_Personnel_Management_data_breach>]

* 1. Biometrics Should Never Be Stored on the Blockchain

*NOTES: Blockchains are immutable, everything on the blockchain is visible to everyone on the blockchain [everyone on chain for closed, but eventually someone will be a blackhat within the closed community; everyone in the world for open]*

*Someone could just walk off with the blockchain*

*Even if encrypted, it could be broken at some point in the indeterminate future, and it’s there forever so there’s no way to protect against that retroactively [it’s like a time capsule]. (HV: technology is always evolving and developing. All technology will become old technology. We could take a long term view that this is a solution to a problem at this moment in time, but in the future the solutions were are exploring may not be the best for long term solutions. Theses solutions will introduce future problems.)*

* 1. Biometrics Should Be Stored Via the Blockchain

*NOTES: [this is our first affirmative statement]*

*[so how do I do it?]*

*Put IDENTIFIER on the blockchain*

*HV: Blockchain may not be the best place to place biometrics. There may be other ways. There is exploration of using biometrics and potentially storing them on the blockchain, but this is limited, naive thinking and we recommend against it.*

*[if you want to mess with blockchains, here’s how you can go about it.*

*[but you could also use these methods for other things*

*[materials things stored on the blockchain is either PUBLIC or SECRET, but something*

*in your hub is PRIVATE, off the chain*

*https://medium.com/karana/public-private-and-secret-information-f857f3931f6b*

*[it’s like a house: everyone knows your address, but they can’t see what’s inside, unless you invite them in.*

*[you can decide what the security is like on your house, and you couldn’t in a public place*

*[so where do you put it?*

*What are the most secure places to put data today? Are there other ways to make secure this data?*

* 1. Biometrics Should Be Under Your Control
     1. Be it in cloud, on device, etc
     2. (being on device is a blip)
     3. (Stuff in cloud should be treated as private property)
     4. [something about mobile devices or trust or control]

*NOTES:*  *[Next thing: talk about how this doesn’t lock us into a specific physicality.]*

*[somewhere under your control!*

*access-control*

*Be it in cloud, on device, etc*

*(being on device is a blip)*

*Adrian G - authorization (for review)*

*[different people will have different criteria for what under their control means*

*REF TWO TECHNOLOGIES: DIF/Microsoft Identity Hub, Gaia Storage used by Blockstack*

*(Stuff in cloud should be treated as private property)[Laws need to change to correyctly support this][there’s already a body of property law][need to respect the sovereignty of the data]*

*CITE: DATA SOVEREIGNTY RULES*

* 1. Biometrics Traits Should Be Reliable
     1. [they will be abused]
     2. Technology will be get better over time to assure liveness
     3. Google CAPTCHA is a good example
     4. (even though used remotely)

1. Biometric traits should be chosen for ones that are reliable
2. They should be selected for ones that are less spoofable
3. No traits are perfect

The use of biometrics (something you are) in authentication includes both measurement of physical characteristics (e.g., fingerprint, iris, facial characteristics) and behavioral characteristics (e.g., gate). Each biometric has its pros and cons and, therefore, the choice of a biometric trait for a particular application depends on a variety of issues such unversitality, permanence, performance and circumvention.

The two issues that we should highlight are the performance and circumvention.

First, based on the nature of the traits, some biometric is well known that they are more reliable than other biometrics with respect to performance such as fingerprint and iris, in comparison to voice and face [1].

Second, circumvention which refers to the ease with which the trait of an individual can be imitated using artifacts (e.g., fake fingers), in the case of physical traits, and mimicry, in the case of behavioral traits (e.g., voice). These spoof attacks is referred to presentation attacks in first ISO standard related to these attacks,ISO/IEC 30107-1:2016 [2].

Biometric systems where identifiable biometrics, such as faces or voice patterns, are more prone to attacks than other biometrics, such fingerprint. Lifting an impression of a fingerprint from a surface needs more effort and skill than Googling an image of a person’s face to use. Further, the cost of launching a face spoof attack, using a printed photo, displayed photo, or replayed video, is relatively low compared to manufacturing spoof fingers using molds or putty.

Therefore,we should highlight, no single biometric is expected to effectively meet all the requirements imposed by all applications. In other words, no biometric is ideal but a number of them are admissible . The relevance of a specific biometric to an application is established depending upon the nature and requirements of the application, and the properties of the biometric characteristic.

[but nothing is perfect]

[but it’s spoofable!]

[biometrics are not the same as passwords or tokens]

[they should only work for the person’s actual finger, attached to their live body]

[because otherwise [they will be abused]

[and they can be spoofed

[it’s particularly important for things that can go out of your control like DNA, and less so for a (random) voice-response or a gait check [can you lose control of it & have someone else duplicate it?

So how to you solve this?

Technology will be get better over time to assure liveness

Google CAPTCHA is a good example

(even though used remotely)

[CITE GARTNER REPORT

* 1. Biometrics Should Be Part of a Systematic Procedure
     1. Not Binary
     2. Balance with Risk
     3. Part of an Ensemble

[because there’s no perfect liveness measure

[we measure how hard things are to spoof

[how much time spoofing would take

[how immutable the biometric is

[what the risk the operation is

[and what the return is

[CITE 800-63 LOA, vectorized — <https://pages.nist.gov/800-63-3/sp800-63-3.html>

[CITE SOFA-B SPECS

BIOMETRICS SHOULD BE PART OF A SYSTEMATIC PROCEDURE

1. Because: Biometrics suffer from FAR and FRR
2. “Second,” provides the what
   1. Active liveness vs passive liveness
      1. Passive = Apple’s face
      2. Active = query-response, like lipreading for saying certain words
   2. Confidence
   3. Increase Security with multiple options
   4. Anomalies & Fallbacks [add in password], lost finger, bandaid
      1. You dropped phone & measurement device broke
      2. (temporary or permanent changes in sensor or person)

Choosing the most suitable system has to based on the the system accuracy and if the system has a built in approach to mitigate presentation attacks.

The basic system of measurement for the accuracy of a biometric system is based on False Rejection Rate (FRR) and False Acceptance Rate (FAR). FRR is the probability that a genuine user being rejected as an impostor. FAR is the probability of an impostor being recognized as a genuine individual.

There is a trade-off between FAR and FRR in every biometric system. In fact, both FAR and FRR are functions of the system threshold. If the threshold is decreased to make the system more tolerant to input variations and noise, then FAR increases. On the other hand, if the threshold is raised to make the system more secure, then FRR increases accordingly. The bottom line is that any biometric system can be made infinitely more secure than traditional passwords, tokens, or PINs, in a number of ways. It’s all a matter of determining the acceptable threshold for FAR and FRR and using the necessary methods and biometric modalities to meet those needs.

Second, mitigating presentation attacks can be accomplished using a combination of varied approaches. The use of multiple factors (a biometric and a password), multiple biometrics, limited number of authentication attempts, performing biometric collection under the supervision of trained operators, and finally a challenge-response approach where the users are asked to say a specific phrase, make a facial expression, or move their hand in a specific way. Most of these approaches are intrusive approaches that may affect the usability and convenience of biometric-based authentication systems, especially mobile biometric systems.

Hence, ISO standards define and highlight the needs for presentation attack detection (PAD) methods that automatically determine such attacks. The goal of PAD methods is to determine if the presented biometric data during the authentication is from the authorized person who is present at the time of capture. PAD methods are non-intrusive techniques that can be categorized into hardware-based methods or software-based methods.

Hardware-based methods are based on using special sensors to measure the life signs such as multispectral fingerprint sensors or the new iPhone X’s infrared and 3D sensors. On the other side, software-based methods analyze the captured biometric image or signal for evidences of spoofing.

Recently, there was an effort by NIST to combine these metrics (SOFA-B) [3] to have a unique metric to compare different biometric systems. SOFA-B includes variables for False Match Rate (FMR) and False Non-Match Rate (FNMR), Presentation Attack Detection Error Rate (PADER), and Effort required to perform the attack on the biometric trait.

References:

[1]Anil Jain, Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." *IEEE Transactions on circuits and systems for video technology* (2004).

[2] <https://www.iso.org/standard/53227.html>

[3] <https://pages.nist.gov/SOFA/SOFA.html>

1. Give the Solution
   1. Architecture of Safe Use of Biometrics
      1. Store in an Identity Hub
      2. Point to from a Blockchain
      3. [and example of DID-Auth too]
   2. And we explain how it meets our 5 or 6 criteria
   3. This solution will cause new problems…
2. Conclusion
   1. A Roadmap is Needed for this Future World
   2. How do we Get There?
   3. That’s the next paper

[so for each truth, why it’s axiomatic,

1. http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html [↑](#footnote-ref-0)
2. https://www.nytimes.com/2017/09/07/business/equifax-cyberattack.html [↑](#footnote-ref-1)
3. [↑](#footnote-ref-2)
4. https://www.hindustantimes.com/tech/aadhaar-breach-everything-you-need-to-know/story-VhCKHDIL8lziw6OcnhL4wO.html [↑](#footnote-ref-3)
5. https://medium.com/karana/public-private-and-secret-information-f857f3931f6b [↑](#footnote-ref-4)
6. https://medium.com/decentralized-identity/the-rising-tide-of-decentralized-identity-2e163e4ec663 [↑](#footnote-ref-5)
7. https://blockstack.org/whitepaper.pdf [↑](#footnote-ref-6)
8. Anil Jain, Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." *IEEE Transactions on circuits and systems for video technology* (2004). [↑](#footnote-ref-7)
9. https://pages.nist.gov/SOFA/SOFA.html [↑](#footnote-ref-8)